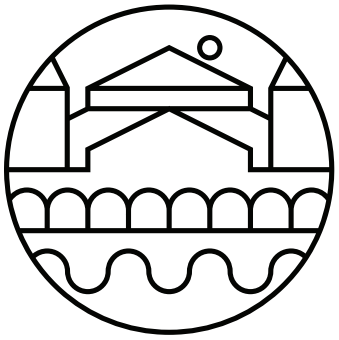
XXXV
SIMPÓSIO BRASILEIRO DE
REDES DE COMPUTADORES
E SISTEMAS DISTRIBUÍDOS

15 a 19 de maio de 2017
Belém - Pará

LIVRO TEXTO MINICURSOS



XXXV
SIMPÓSIO BRASILEIRO DE
REDES DE COMPUTADORES
E SISTEMAS DISTRIBUÍDOS

15 a 19 de maio de 2017
Belém - Pará

Livro de Minicursos

SBRC 2017

Editora
Sociedade Brasileira de Computação (SBC)

Organização
Heitor Soares Ramos (UFAL)
Stênio Flávio de Lacerda Fernandes (UFPE)
Antônio Jorge Gomes Abelém (UFPA)
Eduardo Coelho Cerqueira (UFPA)

Realização
Sociedade Brasileira de Computação (SBC)
Universidade Federal do Pará (UFPA)
Laboratório Nacional de Redes de Computadores (LARC)

Capa: Catarina Nefertari (PCT-UFPA)
Produção Editorial: Denis Lima do Rosário (UFPA)

Cópias Adicionais:
Sociedade Brasileira de Computação (SBC)
Av. Bento Gonçalves, 9500- Setor 4 - Prédio 43.412 - Sala 219
Bairro Agronomia - CEP 91.509-900 - Porto Alegre - RS
Fone: (51) 3308-6835
E-mail: sbc@sbc.org.br

XXXV Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (35: 2017: Belém, Pa).

Anais / XXXV Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos; organizado por Antônio Jorge Gomes Abelém, Eduardo Coelho Cerqueira, Heitor Soares Ramos, Stênio Flávio de Lacerda Fernandes - Porto Alegre: SBC, 2017

296 p. il. 21 cm.
Vários autores
Inclui bibliografias
ISSN: 2177-4978

1. Redes de Computadores. 2. Sistemas Distribuídos. I. Abelém, Antônio Jorge Gomes II. Cerqueira, Eduardo Coelho III. Ramos, Heitor Soares IV. Fernandes, Stênio Flávio de Lacerda V. Título.

Sociedade Brasileira da Computação

Contato
Av. Bento Gonçalves, 9500
Setor 4 - Prédio 43.412 - Sala 219
Bairro Agronomia
91.509-900 – Porto Alegre RS
CNPJ: 29.532.264/0001-78
http://www.sbrc.org.br

Laboratório Nacional de Redes de Computadores (LARC)

Contato
Universidade Federal de Pernambuco - UFPE
Centro de Informática - CIn
Av. Jornalista Anibal Fernandes, s/n
Cidade Universitária
50.740-560 - Recife - PE
http://www.larc.org.br

Organização do SBRC 2017

Coordenadores Gerais
Antônio Jorge Gomes Abelém (UFPA)
Eduardo Coelho Cerqueira (UFPA)

Coordenadores do Comitê de Programa
Edmundo Roberto Mauro Madeira (UNICAMP)
Michele Nogueira Lima (UFPR)

Coordenador de Palestras e Tutoriais
Edmundo Souza e Silva (UFRJ)

Coordenador de Painéis e Debates
Luciano Paschoal Gaspary (UFRGS)

Coordenadores de Minicursos
Heitor Soares Ramos (UFAL)
Stênio Flávio de Lacerda Fernandes (UFPE)

Coordenadora de Workshops
Ronaldo Alves Ferreira (UFMS)

Coordenador do Salão de Ferramentas
Fabio Luciano Verdi (UFSCar)

Comitê de Organização Local
Adailton Lima (UFPA)
Alessandra Natasha (CESUPA)
Davis Oliveria (SERPRO)
Denis Rosário (UFPA)
Elisangela Aguiar (SERPRO)
João Santana (UFRA)
Josivaldo Araújo (UFPA)
Marcos Seruffo (UFPA)
Paulo Henrique Bezerra (IFPA)
Rômulo Pinheiro (UNAMA)
Ronedo Ferreira (META)
Thiêgo Nunes (IFPA)
Vagner Nascimento (UNAMA)

Comitê Consultivo
Allan Edgard Silva Freitas (IFBA)
Antonio Alfredo Ferreira Loureiro (UFMG)
Christian Esteve Rothenberg (UNICAMP)
Fabíola Gonçalves Pereira Greve (UFBA)
Frank Augusto Siqueira (UFSC)
Jussara Marques de Almeida (UFMG)
Magnos Martinello (UFES)
Antonio Marinho Pilla Barcellos (UFRGS)
Moisés Renato Nunes Ribeiro (UFES)
Rossana Maria de Castro Andrade (UFC)

Mensagem dos Coordenadores Gerais

Sejam bem-vindos ao 35º Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC 2017) e à acolhedora cidade das mangueiras - Belém / Pará.

Organizar uma edição do SBRC pela segunda vez no Norte do Brasil é um desafio e um privilégio por poder contribuir com a comunidade de Redes de Computadores e Sistemas Distribuídos do Brasil e do exterior. O SBRC se destaca como um importante celeiro para discussão, troca de conhecimento e apresentação de trabalhos científicos de qualidade.

A programação do SBRC 2017 está diversificada e discute temas relevantes no cenário nacional e internacional. A contribuição da comunidade científica brasileira foi de fundamental importância para manter a qualidade técnica dos trabalhos e fortalecer a ciência, a tecnologia e a inovação no Brasil.

Após um cuidadoso processo de avaliação, foram selecionados 78 artigos completos organizados em 26 sessões técnicas e 10 ferramentas para apresentação durante o Salão de Ferramentas. Além disso, o evento contou com 3 palestras e 3 tutoriais proferidos por pesquisadores internacionalmente renomados, 3 painéis de discussões e debates, todos sobre temas super atuais, 6 minicursos envolvendo *Big Data*, sistemas de transportes inteligentes, rádios definidos por *software*, fiscalização e neutralidade da rede, mecanismos de autenticação e autorização para nuvens computacionais e comunicação por luz visível, bem como 10 workshops.

O prêmio “Destaque da SBRC” e uma série de homenagens foram prestadas para personalidades que contribuíram e contribuem com a área. O apoio incondicional da SBC, do LARC, do Comitê Consultivo da SBRC e da Comissão Especial de Redes de Computadores e Sistemas Distribuídos da SBC foram determinantes para o sucesso do evento. A realização do evento também contou com o importante apoio do Comitê Gestor da Internet no Brasil (CGI.br), do CNPq, da CAPES, do Parque de Ciência e Tecnologia Guamá, da Connecta Networking, da Dantec Telecom, da RNP e do Google. Nosso especial agradecimento à Universidade Federal do Pará (UFPA) e ao Instituto Federal do Pará (IFPA) pelo indispensável suporte à realização do evento.

Nosso agradecimento também para os competentes e incansáveis coordenadores do comitê do programa (Michele Nogueira/UFRP – Edmundo Madeira/UNICAMP), aos coordenadores dos minicursos (Stênio Fernandes/UFPE – Heitor Ramos/UFAL), ao coordenador dos workshops (Ronaldo Ferreira/UFMS), ao coordenador de painéis e debates (Luciano Gaspary/UFRGS), ao coordenador do Salão de Ferramentas (Fabio Verdi/UFSCar) e ao coordenador de palestras e tutoriais (Edmundo Souza e Silva/UFRJ). Destacamos o excelente trabalho do comitê de organização local coordenado por Denis Rosário.

Por fim, desejamos a todos uma produtiva semana em Belém.

Antônio Abelém e Eduardo Cerqueira

Coordenadores Gerais do SBRC 2017

Mensagem dos Coordenadores de Minicursos

Este livro contém os minicursos selecionados para apresentação no XXXV Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC), realizado em Belém-PA, entre os dias 15 e 19 de maio de 2017. O Livro dos Minicursos do SBRC tem sido tradicionalmente utilizado como material de estudo de alta qualidade por alunos de graduação e pós-graduação, bem como por profissionais da área. As sessões de apresentações dos minicursos são também uma importante oportunidade para atualização de conhecimentos da comunidade científica e para complementação da formação dos participantes. O principal objetivo dos Minicursos do SBRC é oferecer treinamento e atualização de curto prazo em temas normalmente não cobertos nas grades curriculares e que possuem grande interesse entre acadêmicos e profissionais.

Tivemos uma diversidade salutar no Comitê de Avaliação em termos de gênero, localização geográfica e experiência, gerando, portanto, opiniões complementares sobre as propostas submetidas. Conseguimos garantir pelo menos três revisões por proposta de minicurso submetida. Tivemos 24 propostas registradas, sendo 22 efetivamente submetidas, representando assim uma taxa de aceitação de 27%. Consideramos as propostas de altíssima qualidade, com temas, focos e escopo diversos. A quantidade de minicursos apresentados no SBRC nos últimos 10 anos tem variado entre 5 e 7. Na edição de 2017, buscamos equilibrar os benefícios trazidos aos participantes com a capacidade de alocação de espaços no evento. Desta forma, foram selecionados 6 minicursos de alta qualidade, consubstanciados em capítulos escritos pelos autores das propostas. Encorajamos a nossa comunidade a continuar enviando propostas de alto calibre para as edições vindouras do evento.

Os capítulos deste livro estão organizados como segue:

1. Big Data Analytics no Projeto de Redes Móveis: Modelos, Protocolos e Aplicações.
2. Sistemas de Transporte Inteligentes: Conceitos, Aplicações e Desafios.
3. Experimental wireless networking research using software-defined radios.
4. Fiscalização da Neutralidade da Rede: Conceitos e Técnicas.
5. Mecanismos de autenticação e autorização para nuvens computacionais: definição, classificação e análise de soluções.
6. Comunicação por luz visível: conceito, aplicações e desafios.

Os Coordenadores dos Minicursos gostariam de agradecer a todos os envolvidos na produção deste livro. Primeiramente aos coordenadores gerais do SBRC 2017, Antônio Abelém (UFPA) e Eduardo Cerqueira (UFPA) pelo convite para a coordenação deste evento, além de todo o apoio necessário para sua condução. Agradecemos também a todos os membros do comitê de avaliação pelo imenso esforço nas revisões e discussões de alta qualidade para todas as propostas submetidas. Por fim, agradecemos aos autores das propostas aceitas pela dedicação no cumprimento dos prazos para finalização deste livro. Os inscritos nos minicursos esperam ansiosamente pelas frutíferas discussões que serão geradas nesta edição.

Stênio Fernandes e Heitor Ramos

Coordenadores de Minicursos do SBRC 2017

Comitê de Avaliação

- Alex Borges Vieira (UFJF)
- André Aquino (UFAL)
- Antonio Rocha (IC/UFF)
- Artur Ziviani (LNCC)
- Carlos Ferraz (UFPE)
- Carlos Kamienski (UFABC)
- Christian Esteve Rothenberg (UNICAMP)
- Daniela Brauner (UFRGS)
- Dênio Mariz (IFPB)
- Eduardo Nakamura (UFAM)
- Fabíola Greve (UFBA)
- Flávia Delicato (UFRJ)
- José Augusto Suruagy Monteiro (UFPE)
- Leandro Villas (UNICAMP)
- Marinho Barcellos (UFRGS)
- Olga Nikolaevna Goussevskaia (UFMG)
- Patricia Endo (UPE)
- Pedro Olmo (UFMG)
- Raquel Mini (PUC-MG)
- Raquel Vigolvino Lopes (UFCG)
- Richard Pazzi (UOIT, Canadá)
- Rossana Andrade (UFC)
- Rostand Costa (UFPB)
- Thiene Johnson (INTEL, EUA)

Sumário

Capítulo 1

Big Data Analytics no Projeto de Redes Móveis: Modelos, Protocolos e Aplicações

Clayson S. F. de S. Celes (UFMG), Ivan O. Nunes (UFMG), João B. Borges Neto (UFMG), Fabrício A. Silva (UFV), Leonardo Cotta (UFMG), Pedro O. S. Vaz de Melo (UFMG), Heitor S. Ramos Filho (UFAL), Rossana M. C. Andrade (UFC) e Antonio A. F. Loureiro (UFMG)

Abstract

The popularization of smart devices with sensing capabilities has led to a huge volume of spatiotemporal data, obtained from different entities, such as people, vehicles, and objects with computing capabilities. The extraction of knowledge from such data offers unprecedented opportunities for decision making processes in several areas. In the mobile networking domain, communication protocols, infrastructure planning and service delivery are examples of applications that can benefit from mining and analysis of data that is collected from smart devices. For instance, user's historic data contain features that are important for detecting patterns and predicting both mobility and data traffic, in time and spatial domains. The goal of this chapter is to present and discuss the potential of big data analytics in the design of mobile networks. In particular, we aim at showing how 5G cellular networks, vehicular networks, and internet of mobile things (IoMT) can take advantage of the knowledge extracted from their entities' characteristics (e.g., user's mobility). In summary, this chapter presents: (i) the peculiarities and techniques of data analysis for the design of mobile networks; (ii) an overview of the recent contributions in the area; (iii) a framework that covers different aspects, ranging from handling the data to employing the knowledge obtained from such data; and (iv) research challenges and opportunities within the area.

Resumo

A popularização de dispositivos com capacidade de sensoriamento tem permitido a obtenção de um enorme volume de dados com informações espaço-temporais de diferentes entidades, tais como pessoas, veículos e objetos com capacidade de processamento.

A extração de conhecimento a partir desses dados cria oportunidades sem precedentes para processos de tomada de decisão em diversas áreas. No domínio de redes móveis, o projeto de protocolos de comunicação, o planejamento de infraestrutura e o fornecimento de serviços são exemplos de aplicações que podem se beneficiar da mineração e análise dos dados coletados a partir desses dispositivos. Por exemplo, dados históricos de usuários contêm características que são importantes para detectar padrões e realizar predições tanto da mobilidade desses usuários como do tráfego de dados no âmbito espaço-temporal. O objetivo deste capítulo é apresentar e discutir o potencial da análise de grandes volumes de dados (big data analytics) com ênfase no projeto de redes móveis. Mais especificamente, nosso objetivo é mostrar como redes celulares 5G, redes veiculares e internet das coisas móveis podem se beneficiar do conhecimento extraído a partir das características (e.g., mobilidade) de suas entidades. Em resumo, este capítulo apresenta: (i) as peculiaridades e técnicas na análise de dados para o projeto de redes móveis; (ii) uma visão geral das contribuições recentes na área; (iii) um arcabouço que aborda desde a manipulação dos dados até a aplicação do conhecimento obtido a partir desses dados; e (iv) os desafios e oportunidades de pesquisa no tema.

1.1. Introdução

Nos últimos anos, a popularização de dispositivos com capacidade de sensoriamento tem permitido a obtenção de um enorme volume de dados com informações espaço-temporais de diferentes entidades, tais como pessoas, veículos e objetos com capacidade de processamento. A extração de conhecimento a partir desses dados oferece oportunidades sem precedentes na tomada de decisão em diversas áreas incluindo planejamento urbano e sistemas de transportes [Çolak et al. 2016], mobilidade urbana [Silva et al. 2014b], epidemiologia [Tizzoni et al. 2014] e sociologia [Soto et al. 2011].

No domínio de redes móveis, sejam essas redes *ad hoc* ou infraestruturadas, o conhecimento extraído a partir de dados do sistema, dos usuários ou da interação entre eles tem fornecido resultados promissores no projeto e na concepção de soluções para tais redes. No entanto, isso exige o conhecimento multidisciplinar que demanda fundamentos distintos de diversas áreas tais como comunicações móveis, mineração de dados e estatística. Nesse sentido, este capítulo visa apresentar um arcabouço que mostra uma metodologia para aplicar análise de dados no projeto de redes móveis a partir de uma visão geral da área.

A literatura apresenta alguns trabalhos relacionados que mostram uma visão geral de esforços que tratam conjuntamente *big data* e redes de comunicação. Em [Yu et al. 2016], os autores apresentam um tutorial sobre infraestrutura e plataformas de redes para tratar o processamento de grande volume de dados. Em [Wang et al. 2016], os autores fazem uma revisão de literatura sobre a aplicação de análise de dados para entender situações de desastres (e.g., terremoto, tempestades) a fim de criar redes móveis *ad hoc*, de forma que os recursos para composição da rede sejam otimizados considerando os resultados obtidos da análise. Em [Blondel et al. 2015] e [Naboulsi et al. 2016], os autores focam exclusivamente em discutir a análise de dados provenientes das comunicações realizadas por usuários de redes celulares, destacando as perspectivas sociais, de mobilidade e de comunicação. Em [He et al. 2016], os autores aplicam *big data analytics* para melhorar o desempenho das redes celulares tanto na perspectiva dos usuários quanto das operadoras. Em [Bi et al. 2015], os autores discutem os desafios e oportunidades no

projeto de comunicações sem fio (*wireless communications*).

Este capítulo se diferencia dos trabalhos supracitados nos seguintes pontos: (i) destacamos as peculiaridades e técnicas na análise de dados a partir de fontes heterogêneas visando o projeto de redes celulares, veiculares e da *Internet* das coisas; (ii) apresentamos um arcabouço, em formato de tutorial, que aborda desde a manipulação dos dados até a aplicação do conhecimento obtido nas análises; e (iii) fazemos uma revisão de literatura evidenciando os principais exemplos de aplicação com seus desafios e oportunidades de pesquisa.

No contexto do SBRC[1], alguns minicursos apresentados nas edições anteriores podem servir de complementação por tratarem de tópicos relacionados. O minicurso de Computação Urbana [Kamienski et al. 2016] cobriu várias aplicações no contexto de cidades inteligentes que podem ser potencializadas com os conceitos e discussões apresentados neste capítulo. Os trabalhos de [Santos et al. 2016] e [Pires et al. 2015] exploram a *Internet das coisas* (IoT) desde a infraestrutura até o domínio de aplicações. Além de aprofundar os estudos sobre *Internet* das coisas móveis, este capítulo também amplia a discussão sobre como explorar os dados oriundos da IoT. Em [Teles et al. 2013] e [Silva et al. 2015d], os autores abordaram as redes sociais móveis e as redes de sensoriamento participativo, respectivamente. Essas redes, quando aplicadas a cenários de grande escala, podem se beneficiar dos conceitos e técnicas de análise de dados. Vale ressaltar que o conteúdo apresentado neste capítulo, aliado aos desafios discutidos nos minicursos citados, fornecem oportunidades interessantes de pesquisa desde a coleta dos dados até aplicações em cenários de comunicação em cidades inteligentes.

A Figura 1.1 ilustra a estrutura geral deste capítulo, que segue uma abordagem *bottom-up* em camadas, tratando desde a obtenção de dados até a sua aplicação no projeto de redes móveis. O restante deste capítulo está organizado como segue. A Seção 1.2 apresenta alguns conceitos sobre *big data*, os tipos de dados e processamento. Em seguida, a Seção 1.3 apresenta o tópico de caracterização de dados. A Seção 1.4 discute as abordagens e modelos utilizados para representar as informações obtidas dos dados brutos com a finalidade de auxiliar no projeto de redes móveis. A Seção 1.5 apresenta diversos resultados de pesquisas atuais relacionados ao projeto de redes móveis a partir de dados sensoriados. Por fim, a Seção 1.6 apresenta as considerações finais deste capítulo.

1.2. *Big data*, tipos de dados e processamento

1.2.1. *Big data*

Nos últimos anos, o termo *big data* tem se tornado bastante popular e empregado em diferentes situações no contexto de dados massivos. Apesar de existirem diversas e até contraditórias definições para o termo [Ward and Barker 2013], um entendimento comum na literatura é que quando se fala sobre *big data* estamos lidando com um grande volume de dados, gerados e processados em uma velocidade fora do comum e que possuem uma larga variedade de informação por serem obtidos de diferentes fontes [Chen et al. 2014a].

No cenário de redes móveis, infraestruturadas ou *ad hoc*, esse grande volume de dados está sendo gerado tanto no domínio do sistema como no domínio dos usuários da

[1] Simpósio Brasileiro de Redes de Computadores

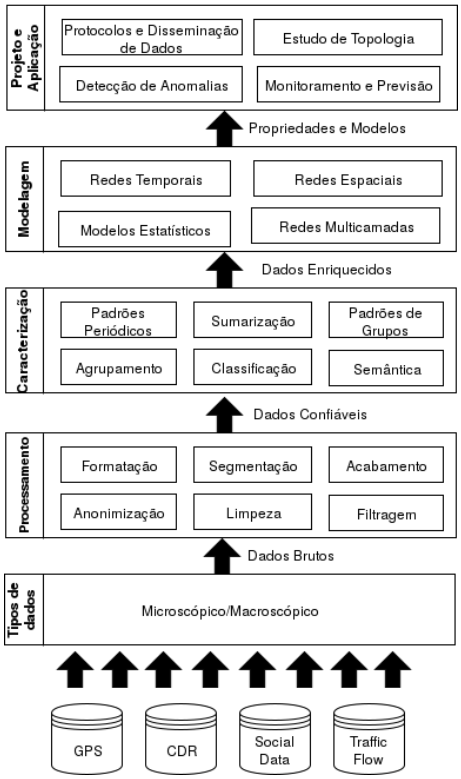


Figura 1.1. Arcabouço geral para aplicar análise de dados no projeto de redes móveis.

rede [Cheng et al. 2017]. Por exemplo, no contexto de redes celulares, do ponto de vista do sistema existem dados sendo gerados pelas interações ou medições entre as comunicações dos dispositivos móveis com as torres celulares, assim como no núcleo da rede. Enquanto no ponto de vista dos usuários, dados estão sendo gerados pela movimentações e registros de chamadas entre usuários. Uma lista bem detalhada dos tipos de dados no contexto de *big data* em redes celulares pode ser encontrada em [Imran et al. 2014].

Explorar a área de *big data* nesse caso é equivalente a agregar informações de diversas fontes a fim de auxiliar a concepção de melhores soluções no projeto de redes móveis. Essa tarefa não é simples, pois, de forma geral, envolve diversos desafios nas seguintes etapas [Wu et al. 2014b]: (i) coleta e tratamento dos dados de forma a não comprometer a segurança do sistema e usuários, além do tratamento dos dados que geralmente possuem várias imperfeições; (ii) armazenamento e manipulação utilizando plataformas computacional de alto desempenho; (iii) caracterização e análise que busca a semântica intrínseca nos dados; (iv) modelagem e aplicação de técnicas relacionadas ao domínio do problema.

1.2.2. Tipos de dados

A ideia de entidades (pessoas, veículos e objetos) como sensores surge como uma oportunidade para coletar e armazenar dados de diferentes dimensões em um cenário urbano. Nesse contexto, a aquisição dos dados pode ser passiva ou ativa [Zheng et al. 2014]. A passiva (ou implícita) consiste em explorar a infraestrutura existente para obter dados gerados pelas entidades. Por exemplo, dados do sistema de navegação de táxi e registros de

Tabela 1.1. Comparação qualitativa dos tipos de fontes de dados.

Tipos de fontes de Dados	Escala (# de entidades)	Granularidade Espacial	Granularidade Temporal	Obtenção dos dados
Registros de detalhamento da chamada (CDR)	Grande	Baixa	Baixa	Fácil
Registros de LBSN	Grande	Alta	Baixa	Fácil
Registros de Transportes	Variante	Alta	Variante	Difícil
Registros de Proximidades	Pequena	N/A	Alta	Difícil
Registros de Posicionamento (HDP)	Pequena	Alta	Alta	Difícil

chamadas nas redes celulares fornecem informações de rastreamento de veículos e pessoas. No entanto, o objetivo primário em cada um desses casos é o fornecimento de rotas e bilhetagem/faturamento para chamadas celulares, respectivamente. Na aquisição ativa (ou explícita) a entidade contribui com o dado intencionalmente para um serviço ou aplicação. Por exemplo, um usuário que compartilha sua localização no *Foursquare* (*check-in*) ou um sensor que fornece leituras de temperatura para um serviço na *Web*.

Nesta seção apresentamos cinco tipos de dados que são largamente usados no projeto de redes móveis: registros de detalhamento da chamada (*Call Detail Record* ou CDR), registros de redes sociais baseadas em localização (*Location based social network* ou LBSN), registros de transportes, registros de proximidades e registros de posicionamento de dispositivos móveis (*Handheld Device Positioning* ou HDP). A Tabela 1.1 apresenta uma comparação qualitativa desses tipos de dados. Em seguida, são apresentados os detalhes sobre cada tipo de dado, destacando suas vantagens, desvantagens, facilidade de obtenção. Além disso, são discutidas possibilidades de aplicação desses tipos de dados.

1.2.2.1. Registros de detalhamento da chamada (CDR)

No âmbito das redes celulares, diversos tipos de dados estão sendo gerados para medições e controle do funcionamento da rede [Cheng et al. 2017]. Particularmente, esta seção foca nos registros de detalhamento da chamada (CDR), que são largamente utilizados na literatura. Registros de detalhamento da chamada são dados dos usuários mantidos pelas companhias telefônicas para fins de controle da rede celular e faturamento dos serviços utilizados. Basicamente, os registros consistem de dados sobre a atividade dos usuários, incluindo realização de chamadas, envio e recebimento de mensagens de texto. Cada registro contém o identificador dos usuários envolvidos na comunicação, data e horário do registro e as estações bases intermediárias usadas na comunicação entre os usuários. Em uma rede celular, cada estação base está associada a uma região geográfica de cobertura (célula), possibilitando, assim, obter uma estimativa da localização dos usuários.

CDRs têm o benefício de cobrir grandes áreas como cidades inteiras, países e até mesmo continentes [Gonzalez et al. 2008]. Além disso, esse tipo de dado permite monitorar um grande volume de usuários, chegando, em alguns lugares, a alcançar um volume na ordem de milhões [Zhang et al. 2014]. Um das principais vantagens de CDRs

é a ausência de custos ou infraestrutura adicionais para a coleta dos dados, já que a sua coleta é realizado pelas empresas de telefonia para fins de cobrança e controle.

A localização de um usuário em CDRs consiste na posição da estação base à qual esse usuário esta associado em um determinado instante, existindo, assim, uma imprecisão espacial no posicionamento dos usuários. Além disso, uma entrada CDR é registrada apenas quando o usuário realiza uma atividade (e.g., efetuar/receber chamadas ou enviar/receber mensagens SMS). Uma vez que os usuários podem não apresentar padrões regulares para a realização dessas atividades, um segundo desafio para a utilização de CDR reside na natureza escassa e não periódica das entradas de posicionamento. Em outras palavras, os dados de CDRs apresentam algumas amostras da trajetória real de um usuário. Apesar dessas restrições, diversos trabalhos na literatura têm adotado dados de CDR para caracterizar e modelar o comportamento dos usuários e do sistema em redes celulares [Gonzalez et al. 2008, Calabrese et al. 2013, Ficek and Kencl 2010, Tanahashi et al. 2012, Wang et al. 2015, Pappalardo et al. 2015].

Geralmente, o acesso a este tipo de dado é restrito às companhias de telefonia, visto que elas o utilizam no serviço de bilhetagem e faturamento. Um desafio consiste em convencer essas companhias a disponibilizar esses dados, pois existe a preocupação com a privacidade dos clientes, além de questões relativas a segredos de negócio [Gramaglia and Fiore 2014]. No entanto, alguns CDRs estão disponíveis publicamente [Blondel et al. 2012, de Montjoye et al. 2014]. Para manter a privacidade do usuários os números de telefone geralmente passam por um processo de anonimização.

1.2.2.2. Registros de LBSN

Redes sociais baseadas em localização, tais como *Foursquare*, *Instagram* e *Twitter*, são um caso particular das redes sociais *online* nas quais os usuários podem informar suas localizações geográficas quando compartilham conteúdos (e.g., *check-in*, foto ou alerta). Com a popularização dos *smartphones*, as LBSNs tornaram-se largamente adotadas e fazem parte da rotina das pessoas que proativamente publicam conteúdos na *Web*. *Web crawlers* podem ser utilizados na extração de conteúdos que possuem dados espaço-temporais dos usuários para compor os registros de LSBN [Silva et al. 2015d].

Uma particularidade quanto à informação de localização nos registros de LBSN é que ela só é disponibilizada se o usuário decidir divulgá-la no momento de compartilhar o conteúdo. Isso introduz imprecisões tanto espaciais quanto temporais para análise de dados. Por exemplo, um usuário pode preferir realizar *check-ins* em localizações específicas ou somente em dias ou horários exclusivos. Portanto, os registros de LSBNs devem ser pre-processados e validados antes de serem submetidos a qualquer tipo de análise e caracterização. Por outro lado, registros de LBSNs têm recebido bastante atenção na literatura devido ao grande volume de check-ins sendo realizados continuamente e à facilidade de coleta desses dados.

1.2.2.3. Registros de Transportes

No contexto deste capítulo, registros de transportes compreendem todos os dados adquiridos de sistemas de transportes que podem ser explorados para obter informações sobre a mobilidade dos usuários. Esses dados vão desde sistemas de GPS (*Global Positioning System*) de veículos até bases de dados de transporte público. Nesta seção, são examinadas algumas das fontes de dados sobre o trânsito e como elas podem ser exploradas para desenvolver estudos no domínio das redes móveis.

Atualmente, uma grande quantidade de veículos estão equipados com sistema de navegação com GPS. Esse sistema fornece aos condutores informações de localização, mapas e rotas. Ao longo da rota de um veículo, o sistema de navegação é capaz de rastrear e armazenar o posicionamento do veículo ao longo do tempo. O conjunto de dados formado por rotas de veículos em um mesmo período e em uma região é tipicamente chamado de *trace*. *Traces* veiculares têm sido frequentemente utilizados para explorar o comportamento urbano e para projetar soluções para redes veiculares. Existem na literatura diversos *traces* de veículos de táxi, por exemplo: São Francisco [Piorkowski et al. 2009a, Piorkowski et al. 2009b], Roma [Amici et al. 2014, Bracciale et al. 2014], Xangai [SUVnet], Shenzhen [Chen et al. 2014b]. Por outro lado, *traces* de veículos de uso pessoal tipicamente não são disponibilizados por questões relativas à privacidade dos usuários.

Além dos *traces* GPS de veículos, com as novas tendências dos sistemas de transporte inteligentes (*Intelligent Transportation Systems* ou ITS), diversos outros tipos de sensores estão sendo instalados no ambiente urbano para a coleta de informações de mobilidade de entidades. Por exemplo, em alguns sistemas públicos de transporte, os usuários de ônibus e metrôs usam cartões inteligentes (*smart cards*) para controle de acesso e saída das estações. Tais tipos de dados podem ser interessantes para o estudo da mobilidade de fluxo de usuários entre regiões, como, por exemplo, determinar a dinâmica da cidade em termos de origem e destino de pessoas durante o horário de pico. Existem alguns conjuntos de dados de *smart cards* disponíveis em *Oyster London Database*[2] e [Zhang et al. 2014].

Os serviços de mapas *online*, tais como *Google Maps*[3], *TomTom*[4], *Here Maps*[5], permitem (por meio de APIs) o acesso a dados de fluxo de tráfego de veículos e condições de trânsito em diversas regiões do mundo quase em tempo real. Por exemplo, o *Google Maps* representa a intensidade de tráfego por uma tabela de cores, enquanto que o *Here Maps* fornece valores numéricos. Esse tipo de informação pode ser aplicada no projeto e validação de modelos e soluções no âmbito das redes veiculares e estudos de sistemas inteligentes de transporte [Tostes et al. 2013].

[2]London OpenData: http://data.london.gov.uk/dataset/oyster-card-journey-information
[3]Google Maps: http://www.google.com.br/maps/
[4]TomTom: http://www.tomtommaps.com/
[5]HereMaps: http://wego.here.com/

1.2.2.4. Registros de Proximidade

Registros de proximidade, ou *traces* de contatos, são bases dados de contatos par-a-par entre entidades. Um contato entre duas entidades consiste em uma interseção espaço-temporal nas trajetórias das duas entidades, i.e., um encontro. Em uma perspectiva de redes oportunistas, um encontro entre duas ou mais entidades é uma oportunidade para disseminação de mensagens/conteúdo entre os usuários/nós da rede. Com uma maior demanda em termos de aplicações para redes oportunistas, tais como o uso de *WiFi Direct* e *Device-to-Device* (D2D) para *offloading* em redes celulares [Rebecchi et al. 2015], os registros de proximidades têm sidos largamente aplicados para simular cenários reais, permitindo avaliação de algoritmos e protocolos.

Existem três formas principais de coleta de dados de registros de proximidades. A primeira é utilizar aplicações móveis de *smartphones* para realizar monitoramento de encontro usando comunicação *Bluetooth* [Eagle and Pentland 2005, Tsai and Chan 2015, Scott et al. 2009]. Alternativamente, uma outra forma utilizada para monitorar encontros é observar a lista de dispositivos móveis conectados aos pontos de acesso de uma rede sem fio [Barbera et al. 2013, Socievole et al. 2014, Henderson et al. 2008, Hsu and Helmy 2005]. A terceira, mais intrusiva, consiste em disponibilizar dispositivos embarcados para um conjunto de pessoas em um ambiente controlado para monitorar seus contatos [Scott et al. 2009, Leguay and Benbadis 2009, Leguay et al. 2006]. Dentre as três formas, a primeira tem se destacado devido ao avanço nas plataformas de desenvolvimento de aplicação para *smartphones*.

Quando *smartphones* são utilizados para coletar dados, primeiramente são recrutados voluntários para instalarem a aplicação de coleta em seus dispositivos. O recrutamento é uma etapa difícil pois envolve convencer os usuários a serem monitorados a partir de seus celulares. Isso se torna um desafio ainda maior em termos de escala (número de usuários), pois poucos usuários aceitam fazer a coleta. Quando se utiliza a coleta por pontos de acessos, deve-se determinar qual será a rede monitorada como, por exemplo, em um ambiente de universidade.

Devidos a esses aspectos, os registros de proximidade sofrem por não possuirem larga escala, em relação ao número de usuários, quando comparados aos registros CDR e LBSN. Por exemplo, o conjunto de dados do MIT Reality [Eagle and Pentland 2006] possui registros de proximidade *Bluetooth* de 80 usuários monitorados durante um período de um ano. O conjunto de dados Dartmouth [Henderson et al. 2008] possui registros de proximidade de 1000 usuários que se conectam aos pontos de acesso de uma universidade durante um período de três meses. Além desses dois conjuntos de dados, outros registros de proximidade foram coletados durante algumas conferências [Pietilainen and Diot 2012, Scott et al. 2009], mas eles apresentam dados de poucos dias de monitoramento.

1.2.2.5. Registros de posicionamento de dispositivos móveis (HDP)

Registros de posicionamento de dispositivos móveis (HDP – Hortonworks Data Platform) são dados obtidos a partir dos receptores GPS nos dispositivos móveis pessoais dos usuários. Normalmente, esses dados são espacialmente precisos e regulares (atualizados

periodicamente), permitindo rastrear as trajetórias dos indivíduos com alta precisão. Vale ressaltar que *traces* de proximidade podem ser obtidos a partir de dados de HDP, visto que um contato pode ser detectado a partir da interseção nas trajetórias de dois indivíduos. Por outro lado, não é possível derivar os registros de HDP a partir apenas de *traces* de contatos.

Como nos registros de proximidade, a coleta desse tipo de dados é feita por aplicações móveis (e.g., MACACO App[6] e *Device Analyzer*[7]) desenvolvidas para essa finalidade. Por isso, *traces* HDP possuem a mesma desvantagem dos *traces* de proximidade: pequena escala. Em geral, *traces* HDP possuem dezenas ou centenas de usuários. Para esse tipo de *trace* é ainda mais difícil convencer os usuários a colaborar porque (i) o monitoramento GPS aumenta consideravelmente o consumo de bateria dos dispositivos, e (ii) os usuários estão preocupados com sua privacidade, ou seja, não desejam divulgar todos os locais que visitam. Por essas razões, existem poucos *traces* desse tipo.

O *GeoLife* [Zheng et al. 2010] é um exemplo de registro HDP disponível publicamente. O conjunto de dados GeoLife possui trajetórias de 182 usuários durante um período de três anos. As entradas do *trace* são uma sequência de posições com registros de latitude, longitude e altitude com granularidade temporal de, em média, 10 segundos.

1.2.2.6. Combinando múltiplas fontes de dados

Todos os tipos de dados apresentados anteriormente possuem algum viés, o qual varia desde a baixa densidade espacial e/ou temporal até a pequena escala em termos de número de usuários. Para contornar tais limitações, uma abordagem interessante consiste em combiná-los. Nesse sentido, alguns esforços na literatura focam em combinar os diferentes tipos de dados para melhorar o entendimento e modelagem das entidades. Por exemplo, em [Silveira et al. 2016], os autores propuseram um modelo que utiliza CDR e LBSN para realizar a predição de mobilidade de pessoas capturando a popularidade de regiões, a frequência de transições entre as regiões e os contatos dos usuários.

1.2.3. Processamento

Um dos principais desafios ao se trabalhar com *big data* consiste em definir como armazenar e processar esses dados. Em geral, os dados brutos capturados provenientes de sensores ou medições possuem redundâncias, inconsistências e vários registros inúteis. Esses dados são, em maioria, não estruturados sendo portanto inadequados de serem armazenados em um banco de dado relacional tradicional [Cattell 2011]. Como alternativa, um banco de dados NoSQL é apto a lidar com tais tipos de dados.

Além disso, os dados brutos, antes de serem armazenados, precisam ser preprocessados, evitando assim o armazenamento de dados com imperfeições. Em [Chen et al. 2014a], os autores destacam as principais tarefas de preprocessando no contexto de *big data*. Limpeza é a tarefa de identificar dados incompletos e imprecisos de forma a removê-los ou modificá-los melhorando assim a qualidade dos dados [Chu et al. 2016]. Eliminação de redundâncias consiste em remover repetições ou dados excedentes, visando reduzir

[6]MACACO App: http://macaco.inria.fr/

[7]Device Analyzer: http://deviceanalyzer.cl.cam.ac.uk/

espaço de armazenamento assim como a tarefa de compressão. Anonimização objetiva ocultar informações prezando a privacidade [Cormode and Srivastava 2009]. Uma visão geral dessas tarefas e técnicas para amenizar os problemas existentes em dados brutos é apresentada em [Karkouch et al. 2016].

Outra questão importante quando se trabalha com um grande volume é o processamento. Nesse caso, o desafio consiste em distribuir as computações dos dados de maneira mais eficiente. O modelo *MapReduce* surgiu como uma excelente contribuição para o processamento paralelo de dados [Dean and Ghemawat 2008]. Atualmente, esse modelo em combinação com o HDFS (*Hadoop Distributed File System*) compõem o núcleo do Apache Hadoop. Hadoop[8] é uma plataforma para processamento distribuído de um grande volume de dados por meio de *clusters* de computadores. Similarmente, Apache Spark[9] é um plataforma para processamento de dados massivos apropriada para aplicações que utilizam múltiplas operações em paralelo, tais como algoritmos de aprendizagem de máquina e mineração de dados [Zaharia et al. 2010].

1.3. Caracterização dos Dados

A caracterização dos dados consiste em uma análise exploratória das principais características de um conjunto de dados [Han et al. 2011]. Com esse objetivo, como descrito em [Tukey 1977], a análise visa maximizar as percepções do conjunto de dados, ao descrever medidas de tendência central e medidas de variabilidade. Além disso, verifica suposições, determina relações entre variáveis, encontra *outliers* e anomalias, entre outros entendimentos implícitos nos dados.

A seguir, vários aspectos identificados na literatura na caracterização dos tipos de dados discutidos na seção anterior são apresentados. Visto que o foco de estudo é de redes móveis, os aspectos tratados aqui são relativos à mobilidade das entidades, atividade temporal, padrões de tráfego de dados e comportamento espaço-temporal de uso dos recursos da rede. Nesse sentido, a caracterização de dados fornece *insights* valiosos para entender o comportamento de usuários e do sistema. Apesar da lista de aspectos apresentada aqui não ser exaustiva, fornece a intuição sobre o propósito de cada um deles.

1.3.1. Estatísticas Básicas dos Dados

Uma necessidade comum quando se trabalha com um grande volume de dados é sumarizá-lo de forma que se possa ter uma visão geral fazendo uso de tabelas, gráficos e valores numéricos. Nesse sentido, adotam-se várias técnicas da estatística descritiva como medidas de tendência central, medidas de dispersão ou variabilidade, tabelas de frequência e gráficos (e.g., diagrama de barras, *box-plot*, histograma) [Freedman et al. 2007]. Por exemplo, em um trabalho recente sobre caracterização do tempo de vida de lugares populares, em [Lu et al. 2016], os autores utilizam dados de localização de embarque e desembarque de táxi na cidade de Nova Iorque[10]. Para sumarizar os dados de mais de um bilhão de viagens dos táxis entre os anos de 2010 e 2015, eles mostraram um conjunto

[8]Apache Hadoop: http://hadoop.apache.org/

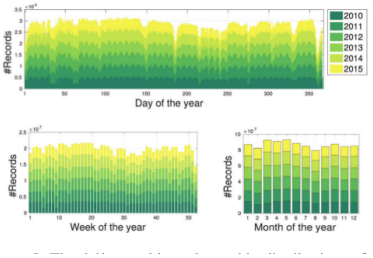
[9]Apache Spark: http://spark.apache.org/

[10]NYC taxicab data from New York City Taxi and Limousine Commission (TLC) official website: http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml

de estatísticas básicas e gráficos que representam a distribuição de viagens ao longo dos anos. A Figura 1.2 apresenta algumas informações extraídas do conjunto de dados. A Figura 1.2(a) mostra algumas médias em relação ao tempo. A Figura 1.2(b) mostra um gráfico de barras do número de registros para cada ano.

Property	Statistics
#Total records	1,039,821,410
Time Span	01/01/2010 - 12/31/2015
Avg. records / year	173,303,568
Avg. records / month	14,441,964
Avg. records / week	3,266,105
Avg. records / day	473,507

(a)



(b)

Figura 1.2. (a) Informações básicas do número de viagens de táxi. (b) Número de registros de viagens de táxi entre os anos de 2010 e 2015 em Nova Iorque apresentados semanalmente, mensalmente e anualmente.

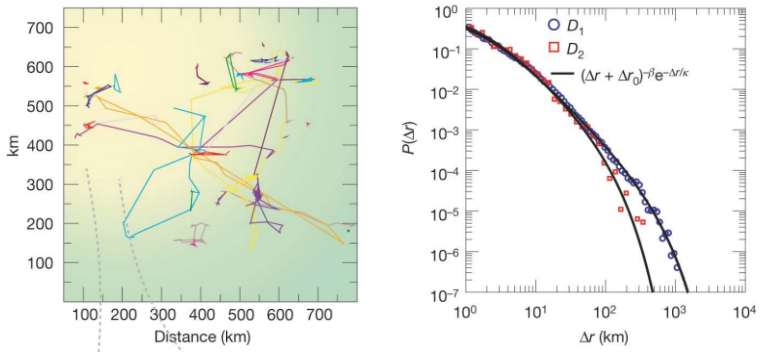
1.3.2. Métricas de Mobilidade

A mobilidade é uma característica fundamental no cotidiano das pessoas e entender os movimentos dos humanos revela muito mais que apenas suas localizações. Dessa maneira, saber como, porque e quando esses movimentos ocorrem, esclarecem vários questionamentos também no projeto de redes móveis tais como sobrecarga na rede, topologia dinâmica, etc. Nesse sentido, esta seção apresenta várias métricas de mobilidade individual e coletiva de entidades presentes no projeto de redes móveis.

Distância de viagem[11] **(*travel distance*).** É uma métrica para descrever o deslocamento de uma entidade [Brockmann et al. 2006]. Ela é computada a partir da interpolação linear entre duas localizações consecutivas em um dado período de tempo, sendo definida como: $\Delta r(u) = \sum_{i=2}^{n} |r_i - r_{i-1}|$ onde $r_u = (r_1, r_2, ..., r_n)$ é a sequência de n posições geográficas registradas no deslocamento de um usuário u durante um período de tempo e $|r_i - r_{i-1}|$ é a distância entre as localizações r_i e r_{i-1}. Em [Gonzalez et al. 2008], os autores analisaram a distância de viagem de pessoas usando dois conjuntos de dados de usuários de rede celular. Eles mediram a distância considerando as posições consecutivas de chamadas telefônicas e encontraram que a distribuição de deslocamento $P(\Delta r)$ em relação a todos os usuários da base de dados segue aproximadamente uma lei de potência truncada (*truncated power-law*). A Figura 1.3 mostra resultados interessantes obtidos nesse trabalho: a maioria dos deslocamentos dos usuários são curtos e as trajetórias humanas mostram um elevado grau de regularidade espacial. Tais resultados serviram como motivação na concepção de vários modelos de mobilidade humana como apresentado em [Hess et al. 2016] .

Raio de giro (*radius of gyration*). Quantifica a dinâmica de mobilidade de uma pessoa em relação ao centro de massa do seu movimento [Gonzalez et al. 2008]. O raio de giro é definido como: $r_g = \sqrt{1/n \sum_{i=1}^{n} (p_i - p_{center})^2}$, onde n é o número de locais visitados por um dado usuário, p_i é a i-ésima posição do usuário e p_{center} é o centro de massa do deslocamento do usuário, obtido como $p_{center} = 1/n \sum_{i=1}^{n} p_i$. O resultado de $p_i - p_{center}$ é a

[11] Também nomeado como deslocamento (*displacement*).



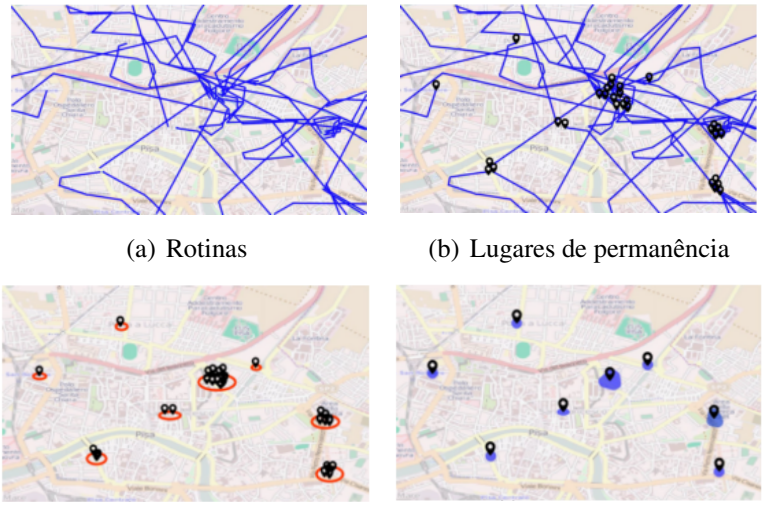
(a) (b)

Figura 1.3. (a) Trajetórias de uma semana de 40 usuários de telefones celulares indicam que a maioria das pessoas deslocam apenas em distâncias curtas, enquanto que algumas deslocam por centenas de quilômetros [Gonzalez et al. 2008]. (b) Função de densidade de probabilidade (P.D.F.) das distâncias de deslocamento obtidas para dois conjuntos de dados estudados (D1 e D2) [Gonzalez et al. 2008].

distância entre um local visitado p_i e o centro de massa da movimentação p_{center}. De uma forma geral, um raio de giro pequeno indica que o usuário se movimenta localmente com viagens curtas, enquanto que um raio de giro grande indica que o usuário se movimenta com viagens mais longas.

Pontos de Interesse (*Points of Interest – PoIs*). Os PoIs de uma cidade são as áreas ou lugares que atraem mais pessoas [Silva et al. 2014a]. A Figura 1.4 ilustra um processo comum para detectar pontos de interesse baseado em dados de mobilidade. Inicialmente, a Figura 1.4(a) mostra um cenário onde diversos indivíduos se locomovem por um certa região. Em seguida, na Figura 1.4(b), observa-se que surgem os lugares de permanência individual (*stay points*). Esses lugares são comumente obtidos pela percepção espaço-temporal entre movimentação dos indivíduos [Jiang et al. 2016, Zheng et al. 2009]. Dado os lugares de permanência, aplicando algoritmos de agrupamento (*clustering*) [Jain 2010] sobre esses lugares é possível obter *clusters* de lugares visitados os quais são potencialmente pontos de interesse, como mostrado nas Figuras 1.4(c) e 1.4(d). Existe uma vasta literatura sobre detecção e recomendação de pontos interesses que podem ser encontrados em [Zhao et al. 2016, Zheng 2015].

Perfis de mobilidade. Analisam o histórico de movimentação das entidades móveis visando estabelecer padrões de comportamentos regulares. Em [Trasarti et al. 2011], os autores exploram o conceito de perfis coletivos de mobilidade em dados de veículos (GPS) para criar um sistema de caronas. Similarmente, o trabalho proposto por [Celes et al. 2013] utiliza os perfis individuais de mobilidade para auxiliar no roteamento de mensagens em redes veiculares. Esses dois trabalhos focam em perfis de mobilidade de trajetórias como podemos ver nas Figuras 1.5 e 1.6. Em [Pappalardo et al. 2015], os autores utilizam tanto CDR como GPS para mostrar a existência de dois perfis de mobilidade baseado na frequência que as entidades visitam determinados lugares e no raio de giro. Eles nomearam de *returners* os indivíduos que vão recorrentemente a poucos lugares e de *explorers* os indivíduos cuja mobilidade não pode



(a) Rotinas (b) Lugares de permanência

(c) *Clustering* (d) Pontos de Interesse

Figura 1.4. Extração de pontos de interesse a partir de dados de mobilidade [Guidotti et al. 2014, Mamei et al. 2016]

ser reduzida a poucos lugares. No domínio de redes móveis, entender as rotinas e perfis de mobilidade das entidades traz benefícios no gerenciamento da mobilidade a partir de padrões preditivos e na melhoria da qualidade de serviços.

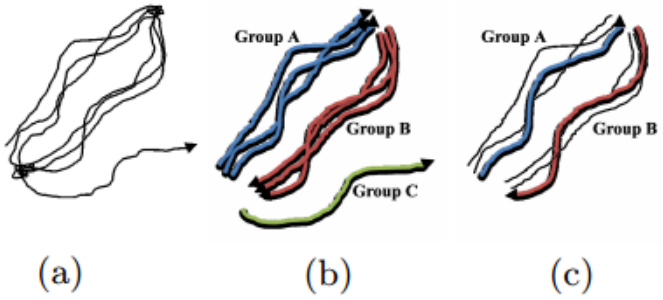


Figura 1.5. Processo para extração do perfil coletivo de mobilidade: (a) identificação das trajetórias dos veículos; (b) detecção de grupos de trajetórias; (c) seleção de perfil de mobilidade mais representativo [Trasarti et al. 2011].

Mobilidade Origem-Destino. Consiste em caracterizar espacialmente pontos de início e fim das movimentações das entidades [Calabrese et al. 2011]. Por exemplo, deslocamento entre cidades, bairros, pontos de interesse. Durante a análise é interessante definir quais as dimensões espaciais e temporais serão utilizadas. A Figura 1.7 mostra os resultados de um estudo que fez a análise de mobilidade origem-destino entre cidades do Reino Unido a partir de dados de LBSN de 2010 a 2013. A Figura 1.7(a) mostra as transições individuais de cada usuário, enquanto a Figura 1.7(b) apresenta uma matriz que mostra a intensidade de transições entre as cidades. Quanto maior a tonalidade de azul, maior é o número de transições entre as cidades. Dessa forma, pode-se perceber um elevado volume de transições entre as cidades de Londres e Edimburgo.

Analisando um outro tipo de dado, trajetórias de veículos, em [Silva et al. 2015b],

Figura 1.6. Processo para extração dos perfis individuais de mobilidade: (a) identificação das trajetórias do veículo; (b) clusterização de trajetórias; (c) sumarização e seleção de perfil de mobilidade mais representativo [Celes et al. 2013].

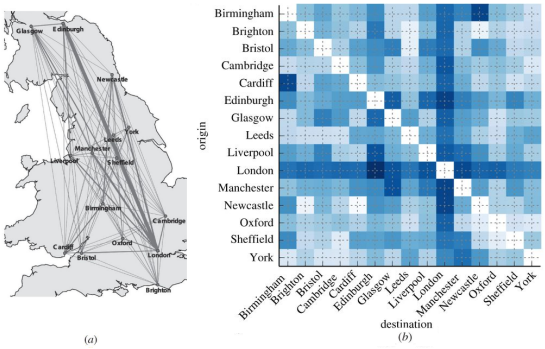


Figura 1.7. Caracterização da mobilidade origem-destino entre cidades do Reino Unido utilizando dados de LBSN [Barchiesi et al. 2015].

os autores caracterizaram a origem e destino dos veículos da cidade de Colônia na Alemanha a partir de um conjunto de dados com trajetórias de mais de 180 000 veículos. Eles particionaram a cidade em células de dimensão 1 km^2 e utilizaram um subconjunto de dados com duas horas de informações. Observando a Figura 1.8(a), percebe-se que as partidas (células de origem dos veículos) tendem a ser igualmente distribuídas por toda cidade, enquanto que observando a Figura 1.8(b), nota-se que as chegadas (células de destino dos veículos) estão concentradas na região central da cidade. Esse comportamento é justificado pelo fato dos autores terem utilizado dados do período de 6:00 às 8:00 da manhã. Nesse período, vários residentes deslocam de casa para o trabalho, ou seja, das regiões de subúrbio para o centro. O estudo realizado nessa caracterização foi aplicado no projeto de replicação de conteúdo em redes veiculares [Silva et al. 2015c].

1.3.3. Métricas de Conectividade

Métricas de conectividade caracterizam as conexões para entender quando elas ocorrem, quão frequentes elas são, quanto duram, entre outras peculiaridades que são importantes para entender a topologia da rede.

Duração de contatos e Tempo entre contatos. São duas características consideradas no domínio de redes *ad hoc*, principalmente nas redes oportunistas. A duração do contato é o intervalo de tempo no qual dois nós da rede estão aptos a se comunicarem por estarem dentro do raio de comunicação. O tempo entre contatos refere-se ao intervalo de

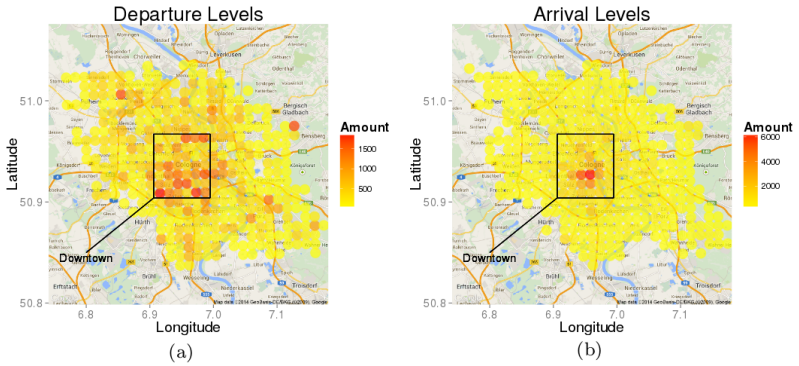


Figura 1.8. Caracterização da mobilidade origem-destino entre regiões da cidade de Colônia na Alemanha utilizando dados de trajetórias de veículos [Silva et al. 2015b].

tempo entre dois contatos. Essas duas características influenciam diretamente no projeto de redes oportunistas em termos da capacidade de dados que podem ser transmitidos em cada contato e maximizando o sucesso de transmissão de mensagens no menor tempo possível, respectivamente.

Topologia. Consiste na disposição dos elementos da rede (e.g., nós, conexões). Explorar os dados gerados pelas entidades da rede visando entender a dinâmica da topologia é extremamente útil na concepção de novos protocolos e serviços. Por exemplo, em [Naboulsi and Fiore 2016], os autores caracterizaram a topologia de uma rede veicular composta de mais de 180 000 veículos em Colônia na Alemanha. Por meio da análise de topologia feita pelos autores é possível responder se a rede formada ao longo do dia é densa ou esparsa, como a conectividade da rede varia ao longo do tempo, como a conectividade é relacionada com as regiões da cidade. A Figura 1.9 mostra como componentes[12] de uma rede veicular variam em relação ao tempo e espaço. Os autores observaram que antes das 6:00 horas, a rede é bastante esparsa e componentes pequenos de no máximo 40 veículos são formados. Entre 7:00 e 8:00 horas ocorre um impacto na topologia de surgimento de componentes gigantes formados por milhares de veículos e outros componentes médios formados por dezenas de veículos, devido ao horário de pico. Esse efeito desaparece e volta acontecer no horário de pico da tarde aproximadamente às 18:00 horas. Vale salientar que os componentes maiores surgem no centro da cidade, onde o tráfego de veículos é denso. Existem outros trabalhos que analisam a topologia de redes veiculares em termos de disponibilidade, conectividade e confiabilidade [Cunha et al. 2016b, Hou et al. 2016, Zhang et al. 2016].

1.3.4. Aspectos Sociais

É típico do comportamento humano estabelecer laços sociais que expressam afinidade e relacionamentos entre indivíduos no dia-a-dia. Esses laços sociais podem ser inferidos a partir dos dados de comunicação entre os dispositivos pessoais de cada indivíduo, já que esses dispositivos, hoje em dia, se tornaram ubíquos. Nesse sentido, observar

[12]Componente é um subgrafo dos nós da rede em que um veículo pode alcançar outro por múltiplos saltos em um intervalo definido t. O tamanho do componente é o número de veículos que pertence a ele.

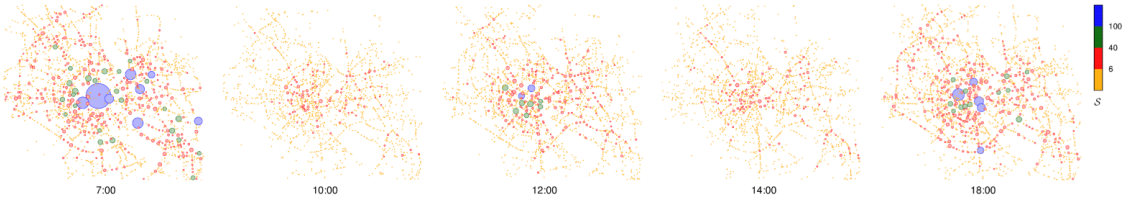
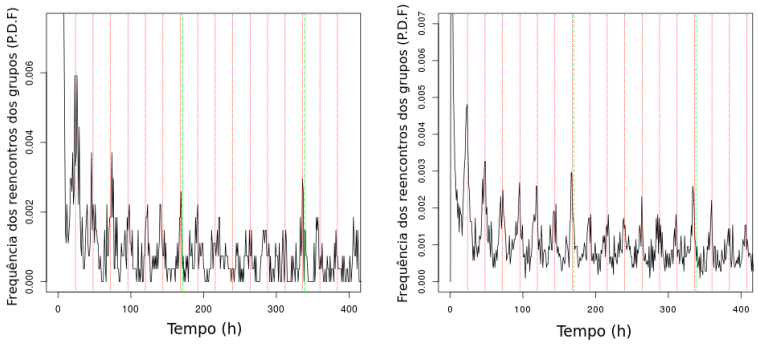


Figura 1.9. Visualização da formação dos componentes de uma rede veicular ao longo do dia em Colônia na Alemanha [Naboulsi and Fiore 2016].

aspectos sociais permite um melhor entendimento da dinâmica e do cotidiano social dos indivíduos. O contexto social, por sua vez pode ser utilizado para aprimorar o desempenho de protocolos de comunicação em redes móveis [Li et al. 2014]. Exemplos de métodos para observar o contexto social a partir de *traces* de contatos são a detecção de comunidades e a detecção de encontros de grupos.

Comunidades. Em diversos tipos de redes, sejam elas sociais, complexas ou móveis, existem nós que são mais interligados entre si formando um aglomerado de nós chamado de comunidades [Palla et al. 2005]. Por exemplo, uma rede social pode ser composta por interações entre pessoas a partir de dados de chamadas telefônicas entre usuários [Hidalgo and Rodriguez-Sickert 2008] ou encontros [Gao et al. 2009]. Os algoritmos propostos por [Palla et al. 2005] e [Gregory 2010] são os mais efetivos e conhecidos na literatura para detecção de comunidades quando tais redes são representadas como grafos estáticos. Em [Nguyen et al. 2011], os autores propuseram uma abordagem semelhante para detecção de comunidades em grafos dinâmicos, i.e., quando nós e arestas surgem e somem ao longo do tempo.

Encontros de Grupos. Em [Nunes et al. 2016b], os autores propõem uma metodologia de detecção de encontros coletivos, i.e., encontros de grupos de três ou mais pessoas. Encontros são definidos como eventos em que duas ou mais entidades (e.g., pessoas, veículos) ocupam a mesma região em um mesmo intervalo de tempo, ou seja, essas entidades encontram-se fisicamente próximas. Ao realizar uma caracterização temporal dos encontros de grupos, percebeu-se que esses encontros apresentam alta periodicidade. Como mostrado na Figura 1.10, em diferentes *traces* de contatos, temos que os encontros de grupos se repetem frequentemente, apresentando principalmente periodicidade diária e semanal. Mais especificamente, a Figura 1.10 mostra a função de densidade de probabilidade (PDF) para os reencontros de grupos de pessoas nos *traces* de contatos do MIT e de Dartmouth. Pode-se notar que as PDFs apresentam picos de maior densidade em torno de períodos de 24 horas (representados pelas linhas pontilhadas vermelhas). Além disso, são observados picos maiores em períodos de sete dias (indicados pelas linhas tracejadas verdes). Esse resultado evidencia a existência de regularidade temporal nos encontros de grupos. Na seção 1.5.1, será discutida a utilização do conhecimento sobre encontros de grupos para o projeto de um protocolo de encaminhamento oportunístico em redes D2D.



(a) *Trace* do MIT (b) *Trace* de Dartmouth

Figura 1.10. Função de densidade de probabilidade (PDF) do tempo entre reencontros de grupos de pessoas após o primeiro encontro em $t = 0$.

1.3.5. Aspectos de Uso da Rede

Com o objetivo de atender a crescente demanda do uso de recursos da rede celular, é importante entender a dinâmica do tráfego e seu impacto na alocação de recursos na rede da operadora de serviços. A partir dessa compreensão, pode-se tanto definir um melhor planejamento de recursos como auxiliar o projeto de rede, beneficiando de alguma forma os usuários de telefonia.

Distribuição espaço-temporal do tráfego. Uma investigação interessante consiste em saber como as pessoas costumam utilizar o tráfego de dados no domínio do tempo e em que região da cidade. Em [Xu et al. 2016], os autores analisaram dados da cidade de Xangai para verificar a distribuição temporal de uso de dados. A Figura 1.11 mostra a distribuição de uso de dados em diferentes escalas de tempo[13]. Na Figura 1.11(a), os dados agregados do tráfego são exibidos e percebe-se que a distribuição segue aproximadamente o comportamento da rotina humana, tráfego intenso durante o dia e tráfego reduzido à noite. Note que existem dois picos durante o dia, um em torno das 12:00 horas e outro em torno das 22:00 horas, possivelmente pelo consumo de dados após o almoço e antes de dormir. As Figuras 1.11(b) e 1.11(c) mostram o comportamento periódico existente ao longo dos dias. A primeira mostra uma semana de dados, enquanto a segunda mostra um mês. Para ambos os casos, percebe-se a regularidade quanto aos uso do dados, tendo no finais de semana uma diminuição do consumo. Além de verificar o comportamento temporal, os mesmos autores analisaram em que regiões da cidade tinham consumo (*bytes* transmitidos por hora por km^2) maior em diversos momentos do dia como mostra a Figura 1.12. Nota-se que independente da hora do dia, as torres localizadas no centro da cidade (região com coloração vermelha) apresentam uma intensidade alta de dados em relação às demais torres de outras regiões. Um trabalho semelhante é apresentado em [Nika et al. 2016], onde os autores propuseram uma metodologia para identificar *hotspots* (lugares com tráfego intenso na rede) com a finalidade de fornecer qualidade de serviço aos usuários mesmo em

[13]Em [Xu et al. 2016], os autores utilizaram dados CDR de 150 000 usuários de um mês da cidade de Xangai na China. As informações analisadas contém o identificador (ID) dos dispositivos, início e fim da conexão, estação base que o ID está associado e quantidade de dados 3G e LTE usados na conexão. Os dados contêm 2.4 petabytes registros, 77 terabytes por dia e 8 gigabytes, em média, por estação base.

situações de sobrecarga.

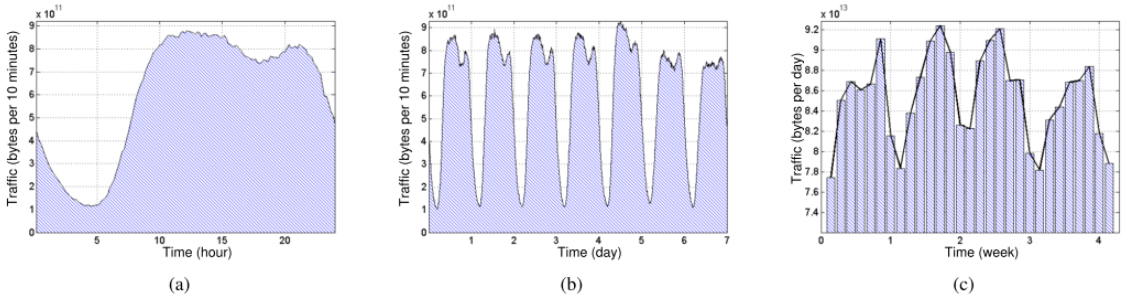


Figura 1.11. Distribuição temporal do tráfego em rede celular para diferentes escalas de tempo [Xu et al. 2016]. (a) Hora em hora. (b) Diariamente. (c) Semanalmente.

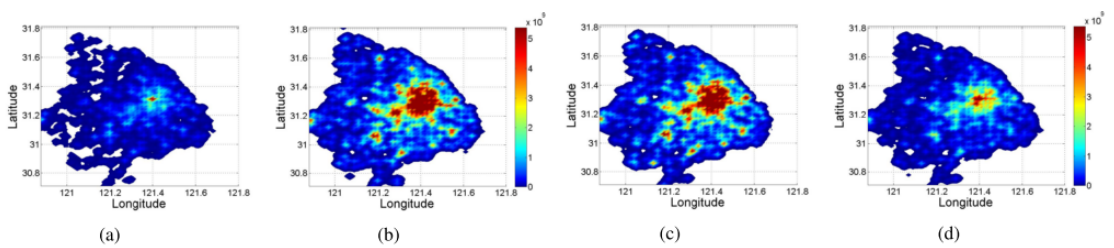


Figura 1.12. Distribuição espacial do tráfego em rede celular em diferentes momentos [Xu et al. 2016]. (a) 04:00. (b) 10:00. (c) 16:00. (d) 22:00.

Distribuição espaço-temporal de tipos de conteúdo. Um outro aspecto importante consiste em caracterizar que tipo de conteúdo os usuários da rede celular utilizam em relação ao tempo e região geográfica. Em [Shafiq et al. 2015], os autores caracterizaram a relação entre volume de tráfego e tipo de conteúdo. A Figura 1.13(a) ilustra a utilização do tráfego de quatro diferentes tipos de conteúdos (*dating*, *maps*, *social network*, *web*) em função do tempo. Pode-se observar que o volume de tráfego para *dating* e *social network* é maior durante o período da noite. Enquanto que o volume de tráfego para navegação *web* e *maps* concentra-se desde o meio-dia até o resto da tarde. A Figura 1.13(b) mostra que para determinadas regiões a intensidade de tráfego para um tipo de conteúdo predomina em relação aos demais. Em outro trabalho semelhante ao descrito anteriormente [Trestian et al. 2009], os autores investigaram a correlação dos interesses dos usuários com localizações geográficas a partir de dados da rede celular.

1.3.6. Considerações

A capacidade para identificar os aspectos descritos anteriormente está fortemente relacionada à aplicação de técnicas e algoritmos do domínio da estatística, mineração de dados e aprendizagem de máquina. Por isso, o restante desta seção apresenta uma breve introdução a essas técnicas e algoritmos.

Agrupamento. É um método não supervisionado (não requer processamento de treinamento de aprendizagem) para agrupar as amostras de dados e, principalmente, classificá-las de acordo com suas características. Os algoritmos de agrupamento buscam

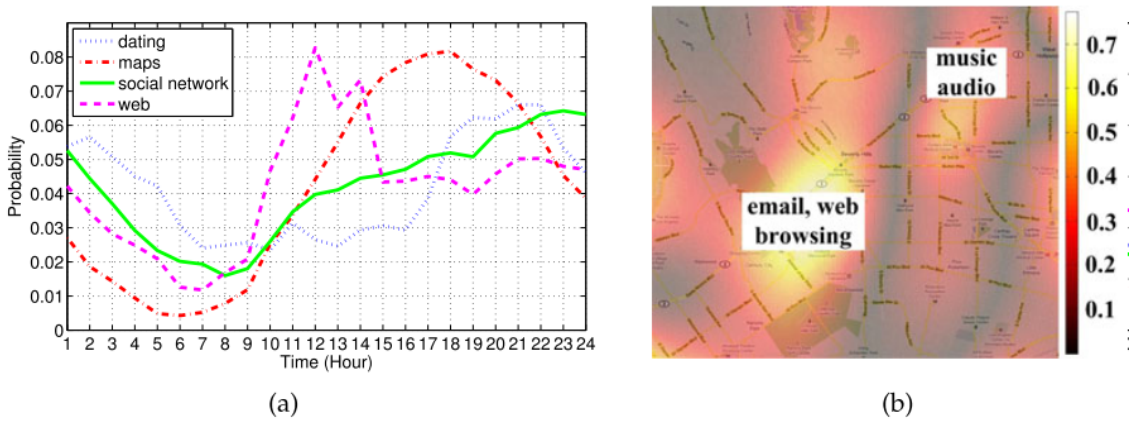


Figura 1.13. Distribuição espaço-temporal de diferentes tipos de conteúdos em redes celulares [Shafiq et al. 2015]. (a) Distribuição temporal de hora em hora. (b) Tipos de conteúdos mais consumidos em regiões distintas.

particionar as amostras em grupos (ou *clusters*) nos quais as amostras pertencentes a cada grupo possuem uma maior similaridade. Existem diferentes tipos de algoritmos de agrupamento que dependendo dos dados e aplicação podem ser mais adequados em uma dada situação tais como agrupamento baseado em densidade, agrupamento baseado em particionamento, agrupamento hierárquico e agrupamento espectral [Hand et al. 2001].

Correlação. É uma técnica para investigar a relação entre duas variáveis contínuas. As medidas comumente aplicadas para medir a correlação são as correlações de Pearson e Spearman. Por exemplo, em [Centellegher et al. 2016], os autores investigaram a correlação entre o número de mensagens SMS e chamadas telefônicas realizadas por usuário e verificaram que existe uma forte relação do uso do telefone celular para essas duas variáveis.

Regressões. A regressão linear permite explorar e estimar um valor esperado quantitativo de uma variável (Y) a partir dos valores de outras variáveis (X) [James et al. 2014]. Ela é dita linear pois existe uma relação linear entre Y e X_1, X_2, ...,X_p. Existe outro tipo de regressão conhecida como regressão logística que se diferencia da linear, principalmente pelo fato da variável resposta ser categórica.

Padrões frequentes. Consiste em detectar itens, subsequências ou estruturas que são recorrentes em um conjunto de dados [Aggarwal and Han 2014]. Encontrar padrões frequentes nos dados pode ser útil tanto para verificar associações e relacionamentos como para auxiliar nas tarefas de indexação, classificação e agrupamento dos dados.

Séries Temporais. De forma geral, trata da representação dos dados em função do tempo [Shumway and Stoffer 2010]. A análise de séries temporais compreende métodos para analisar dados a fim de extrair estatísticas significativas e outras características dos dados. Por exemplo, por meio de séries temporais é possível verificar tendências, sazonalidades e *outliers* nos dados.

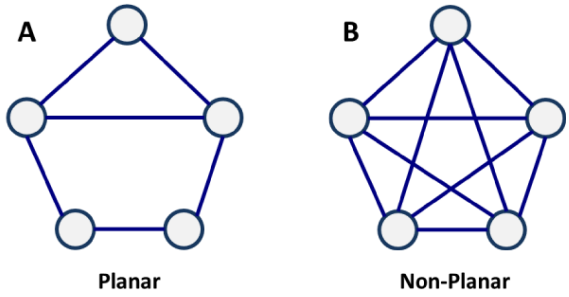


Figura 1.14. Exemplos de grafos planar (A) e não-planar (B)

1.4. Modelagem

A Seção 1.3 discutiu como caracterizar alguns exemplos de dados provenientes de redes móveis. Essa etapa é importante no entendimento de peculiaridades, comportamento macro e microscópio dos aspectos que serão explorados. Em seguida, busca-se um modelo que represente o comportamento real desses aspectos no sistema observado. Por fim, a partir de um modelo descritivo dos dados, cria-se um modelo generativo do comportamento observado.

A seguir, apresentamos como podemos modelar os dados observados usando redes espaciais, redes temporais, redes sociais, redes multicamadas e, por fim, como criar modelos generativos. Em cada seção, apresentamos os tipos de dados apropriados em cada modelo, suas vantagens e desvantagens, exemplos e estratégias de uso.

1.4.1. Redes Espaciais

Em uma modelagem tradicional de redes, ou grafos, temos um conjunto de nós N e um conjunto de arestas E conectando os nós. Usualmente, nós e arestas podem possuir atributos, caracterizando assim a geometria da rede. De maneira geral, uma rede espacial é uma rede onde os nós estão posicionados em um espaço equipado de uma métrica [Barthélemy 2011]. Para a maior parte das aplicações, esse espaço tem duas dimensões e a métrica é a distância Euclidiana. Nessa definição de rede espacial, a probabilidade de dois nós estarem conectados é inversamente proporcional à distância entre eles. Entretanto, essa definição não garante a planaridade da rede. Uma rede planar é uma rede que pode ser desenhada no plano sem que haja intersecção de arestas, como mostra a Figura 1.14. Um exemplo tipicamente encontrado na prática de uma rede espacial não planar é a de aeroportos. Nessa rede, aeroportos formam o conjunto de nós e vôos entre dois aeroportos compõem o conjunto de arestas. Nesse caso, as arestas são direcionadas no sentido do destino do vôo. Apesar da probabilidade de existir um vôo entre dois aeroportos muito distantes ser pequena, há fatores como tamanho das cidades que influenciam mais o surgimento de arestas, tornando a rede não planar. Outras redes, como a social, podem ser vistas como uma rede espacial, porém não planar. Duas pessoas localizadas longe espacialmente são menos propícias a serem amigas, entretanto, esse tipo de amizade existe por conta de diversos outros motivos, como a própria *Web*. Todavia, como veremos a seguir, a maioria das modelagens de mobilidade resultam em redes planares.

Por exemplo, se considerarmos o aspecto de mobilidade dos nós. Modelar mo-

bilidade humana com redes espaciais é bastante direto e intuitivo. Na maior parte das abordagens, um nó representa uma seção do espaço como, por exemplo, o subespaço que envolve um quarteirão ou um aeroporto de uma cidade. Uma aresta representa a possibilidade de movimentação direta entre dois espaços. Essa abordagem pode ser vista como uma maneira de discretização do espaço. Da mesma forma, a mobilidade de um entidade pode ser discretizada em uma sequência de espaços no tempo. Essa sequência, chamada de trajetória, define um caminho que a entidade percorre na rede. Os nós e arestas da rede podem receber atributos para representar a mobilidade das entidades observada nos dados. A maneira mais comum de fazer isso é atribuir às arestas pesos proporcionais ao fluxo de entidades entre os dois nós ligados pela aresta. Observe como essa modelagem permite apenas a análise estática da mobilidade. Isto é, como o componente de tempo não é considerado na rede, ela representa uma fração do tempo, no qual trajetórias são agregadas na mesma rede. Essa abordagem normalmente nos permite fazer análises mais globais sobre mobilidade. Observar eventos que dependem do tempo, como o movimento pendular em grandes cidades, se torna inviável. Todavia, podemos fazer análises macroscópicas de movimento. Em [Barthélemy 2011], usa-se o exemplo da lei de gravidade em mobilidade utilizando essa modelagem. Nesse caso, são adicionados os atributos de distância física entre nós nas arestas e de população do espaço nos nós. Os nós representam cidades ligadas por vôos existentes entre elas. É mostrado como o fluxo entre elas é diretamente proporcional às suas populações e à distância entre elas.

No contexto de redes móveis, uma análise estática, isto é, apenas espacial, de mobilidade pode ser útil em diversas aplicações. Um exemplo seria analisar quais regiões do espaço possuem menor fluxo, e portanto, deveriam receber mais atenção no envio de mensagens. Por exemplo, caso se saiba que uma entidade contém uma aresta (a,b) de peso pequeno em sua trajetória futura, pode-se encaminhar para ela o máximo de mensagens possível que tem como destino b. Em um estudo utilizando dados de registro de chamadas e SMS [Lambiotte et al. 2008], os autores investigaram a relação entre comunicações e posicionamento geográfico dos usuários. Eles modelaram os nós como as regiões de coberturas das estações bases que os usuários estão associados e as arestas como as ligações de comunicações. Essa rede possui 2,5 milhões de usuários monitorados em um período de seis meses. Um resultado interessante observado por eles foi que a duração média das chamadas aumenta com a distância até um limite de 40 km, como pode ser observado na Figura 1.15. Os autores justificam esse fato pelo fato de quando as pessoas vivem próximas umas das outras tendem a se encontrar e comunicar pessoalmente.

1.4.2. Redes Temporais

De maneira geral, uma rede temporal é uma rede, ou grafo, em que seu conjunto de arestas é uma função do tempo. Isto é, uma aresta existe ou não de acordo com uma dada janela de tempo em que observamos a rede. Existem diversas formas de modelarmos essa estrutura e a melhor escolha depende da análise que pretendemos fazer [Holme and Saramäki 2012].

A modelagem mais usual de uma rede temporal é via a análise de redes agregadas [Holme and Saramäki 2012]. Definimos uma janela de tempo e agregamos as arestas contidas nela. Assim, analisamos separadamente cada rede como uma rede estática, podendo, por exemplo, empregar algoritmos da teoria clássica de grafos. Na Figura 1.16, observamos um exemplo de redes agregadas e suas sequências. Uma prática muito comum,

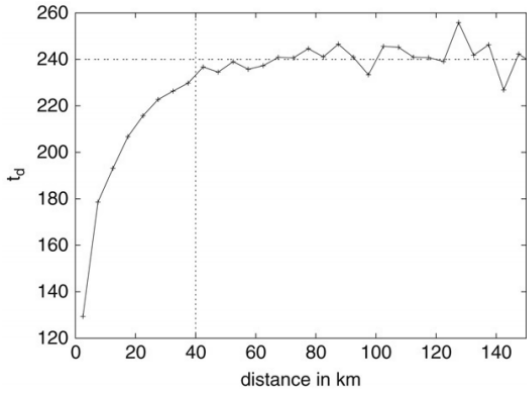


Figura 1.15. Duração média t_d de chamadas telefônicas (em segundos) em função da distância [Lambiotte et al. 2008].

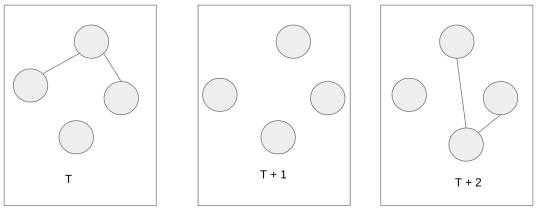


Figura 1.16. Exemplo de redes agregadas por janela de tempo

porém perigosa, é a de empregar algoritmos de caminhamento nas redes agregadas. Nada impede que o algoritmo considere um caminho impossível na prática. Considere um caminho de tamanho k na rede agregada como a sequência de arestas $a_1, a_2, ..., a_k$, o tempo de aparecimento de uma aresta a_i como $ti(a_i)$ e o tempo de desaparecimento de uma aresta a_i como $tf(a_i)$. O caminho é impossível caso exista $tf(a_{i+1}) < ti(a_i)$. Esse é um dos problemas fundamentais de redes temporais e diversos algoritmos foram propostos para lidar com ele [Holme and Saramäki 2012]. Devido ao grande uso de redes temporais para modelagem de redes de contatos e, consequentemente, encaminhamento de mensagens em redes móveis, recomendamos a leitura de tais algoritmos, suas aplicações e desafios.

No contexto de redes móveis, as redes temporais podem ser utilizadas para modelar tanto os contatos entre entidades quanto as suas mobilidades. Pelo lado da mobilidade, podemos agregar o fator tempo às redes espaciais apresentadas anteriormente, por exemplo. Criamos, então, redes espaço-temporais [Williams and Musolesi 2016], capazes de representar o fluxo de entidades no espaço condicionado ao tempo. Isto é, análises de movimento pendular podem ser feitas com essas estruturas. Em [Williams and Musolesi 2016], os autores estudam propriedades clássicas de redes, como vulnerabilidade nas redes de mobilidade propostas.

Além de estudar a mobilidade propriamente dita, podemos usar redes temporais para estudar os contatos gerados pela mobilidade das entidades. Esse é o uso mais comum e, talvez, mais importante de redes temporais em redes móveis. Um contato entre duas entidades pode ser definido como uma interseção de tempo e espaço. Um contato ocorre

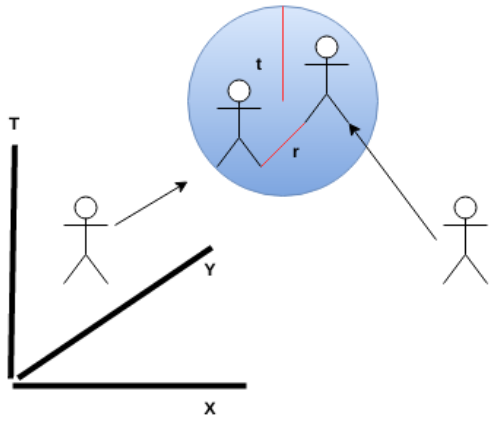


Figura 1.17. Duas entidades geram um contato na rede caso ocupem o mesmo subconjunto do espaço tempo definido por (r,t)

quando duas entidades estão a uma distância máxima r no mesmo instante de tempo. A distância r é definida de acordo com a aplicação analisada. Por exemplo, se estudamos redes D2D podemos definir r como o raio de comunicação da tecnologia Bluetooth. A duração do contato é o tempo no qual as duas entidades permanecem a uma distância de no máximo r. Caso o contato esteja sendo extraído computacionalmente de dados de trajetórias, temos de colocar uma tolerância t na intersecção de tempo. Lembrando que uma trajetória é uma amostra discreta de tempo e espaço. Assim, observamos que um contato é definido dentro de um subconjunto de tempo e espaço, como mostrado na Figura 1.17. Dessa maneira, podemos montar uma rede temporal representando os contatos entre as entidades. Nela, cada nó representa uma entidade e cada aresta representa um contato. Os intervalos de tempo de existência de cada aresta são iguais às durações dos contatos entre as entidades. Essa modelagem é bastante comum em redes móveis pois possibilita o estudo do encaminhamento de mensagens entre as entidades.

A maior parte das aplicações de redes móveis, seja de mobilidade veicular ou humana, utiliza a modelagem de redes temporais para estudar e tirar vantagem de propriedades como periodicidade, centralidade e outras métricas sociais que explicamos a seguir [Mota et al. 2014]. Um exemplo do uso de redes temporais em redes veiculares é dado em [Wu et al. 2011]. Na publicação, os autores prevêem trajetórias dos veículos podendo montar uma rede temporal futura. Nela, eles analisam o problema de roteamento de pacotes e mostram que esse é um problema NP-difícil. Os autores propõem heurísticas locais e globais para esse problema.

Em [Michel and Julien 2016], os autores modelaram um cenário de comunicação entre dispositivos em uma rede móvel *ad hoc* como uma rede temporal. A Figura 1.18 mostra uma visão geral do funcionamento da transmissão de duas mensagens em um grafo temporal. Primeiramente, na Figura 1.18(a), os autores modelaram com um tipo específico de representação de grafo temporal. Os vértices são os nós da rede e as arestas indicam os contatos, sendo que os rótulos dos contatos representam o tempo de contato entre os nós. A Figura 1.18(b) mostra os intervalos de interseção entre os contatos e a Figura 1.18(c)

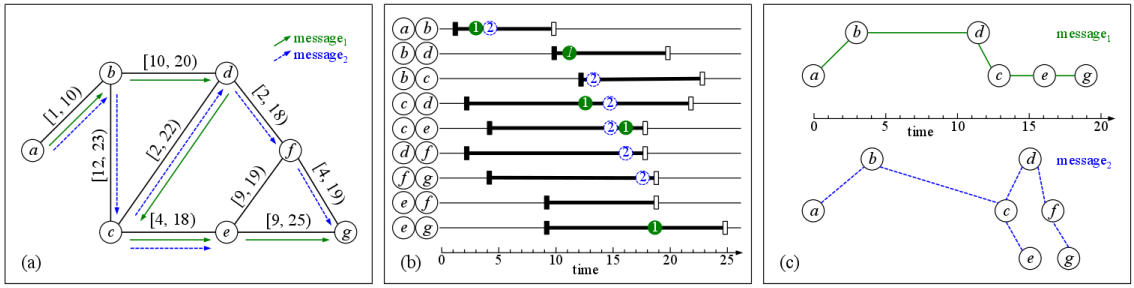


Figura 1.18. Processo de transmissão de mensagem em um grafo temporal [Michel and Julien 2016]. (a) uma representação de uma rede móvel oportunista. (b) intervalos de encontros para pares de nós. (c) identificação do tempo de recebimento de mensagens pelos nós.

exibe o tempo de recebimento das mensagens por cada nó. Um ponto interessante aqui é a possibilidade de explorar a caracterização para entender como os encontros acontecem com objetivo estimar os tempos e durações de encontros que serão utilizados como rótulos das arestas.

1.4.3. Redes Sociais

Redes Sociais Online surgiram (RSO) recentemente e já se tornaram uma vasta área de pesquisa em ciência da computação [Garton et al. 1997]. A premissa básica no estudo de RSO é a de que cada nó representa uma pessoa, ou entidade, e cada aresta representa um laço social, ou amizade, entre elas. Diversos estudos vêm abordando as estruturas formadas pelos laços sociais das pessoas e o que elas significam na prática [Leskovec et al. 2009]. Muitos outros usaram tais estruturas para projetar melhores sistemas [Benevenuto et al. 2009].

É intuitivo pensar que as redes de contatos apresentadas anteriormente podem ser vistas como redes sociais, já que o contato físico é uma forma de interação social. Todavia, muitos desafios são apresentados nessa modelagem. O primeiro é o fato de RSOs serem redes estáticas. Como vimos anteriormente, as redes de contatos são redes temporais (ou espaço-temporais). O primeiro desafio seria converter a rede temporal em uma rede estática. Em uma RSO, as pessoas declaram explicitamente que são amigas, gerando assim o laço entre elas. O problema na conversão da rede de contatos para uma RSO seria identificar quais contatos são frutos de uma amizade e quais são aleatórios. Podemos ir além e classificar o grau de socialização entre as pessoas ou entidades como, por exemplo, amigos e conhecidos. Esse desafio de pesquisa foi explorado em [de Melo et al. 2015]. Os autores utilizam a periodicidade e a estrutura social dos contatos para caracterizá-los como amigos, conhecidos ou aleatórios. Em [Hui et al. 2011], os autores consideram todos os contatos em uma janela de tempo como laços sociais e exploram as estruturas sociais, como comunidade e centralidade para criar um protocolo de roteamento. Em [Cunha et al. 2016a], os autores exploram interações sociais entre veículos, extraídas da mesma forma que em [Hui et al. 2011], para criar um novo protocolo de encaminhamento de mensagens. Por outro lado, em [Nunes et al. 2016c], os autores apresentam um novo conceito social: o encontro de grupos em redes de contatos. Com essa nova medição social, os autores projetaram um novo protocolo de encaminhamento de mensagens D2D.

A extração de propriedades sociais das redes de contatos já se mostrou muito

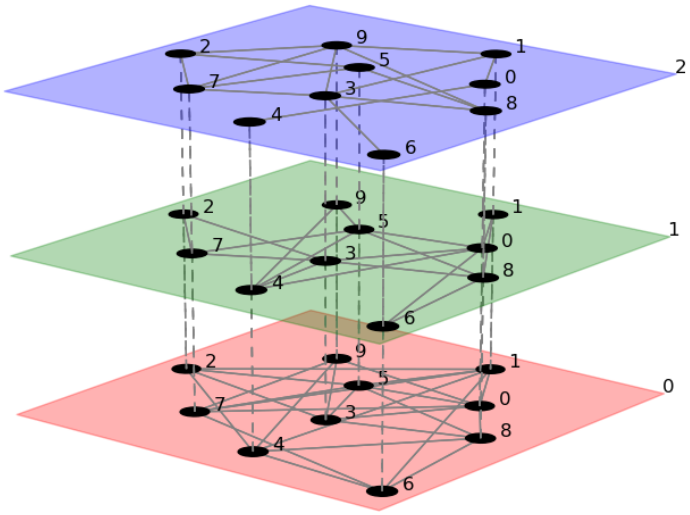


Figura 1.19. Exemplo de rede multicamada com três domínios, ou camadas, diferentes.

eficiente para aplicações de redes D2D e, mais recentemente, de redes V2V. Os principais problemas atrelados a elas são a distributividade das computações e a limitação e dependência de dados para validação. Isto é, as computações envolvidas normalmente são centralizadas, pois precisam de informações de todos os nós. Na prática, os protocolos de redes móveis são distribuídos e o acesso e computação dessas informações pelos nós se torna inviável [Hui et al. 2011]. Além disso, propriedades sociais podem ser bastante dependentes da fonte de dados que usamos. Isto é, analisar propriedades sociais de uma base de dados de um campus universitário naturalmente vai ser diferente de fazer a mesma tarefa em uma base com pessoas aleatórias de uma cidade. Na última, a rede de contatos e, consequentemente a rede social, vai ser muito esparsa. Assim, observamos que como toda solução orientada a dados, as soluções que usam propriedades sociais devem ser ponderadas pelo contexto em que foram avaliadas.

1.4.4. Redes Multicamadas

Um paradigma recente em teoria de redes é o de redes multicamadas [Kivelä et al. 2014]. Essas redes abordam o problema de entidades possuirem relações de diferentes naturezas. Isto é, para cada par de nós na rede, é definido um conjunto de domínios. Em cada domínio, ou camada, existe um conjunto de arestas, que definem as relações entre as entidades naquele domínio. A teoria de redes multicamadas propõe novos algoritmos que exploram eficientemente as estruturas definidas [Kivelä et al. 2014]. O exemplo mais intuitivo talvez seja o de relações humanas, que possuem diversas naturezas. Por exemplo, duas pessoas possuem uma relação definida no domínio do trabalho, empregado e empregador, e uma no domínio social, amigos. A Figura 1.19 ilustra uma rede com o mesmo conjunto de nós em três camadas, sendo que, para cada camada, existem diferentes relações entre os nós.

Depois de apresentarmos modelagens de redes móveis considerando domínios diferentes, a modelagem de redes multicamadas se torna bastante intuitiva nesse contexto. Como vimos, uma rede de contatos possui aspectos temporais, espaciais e sociais. Assim, podemos considerá-los como domínios de uma rede multicamada. Existem diversas outras possibilidades de domínios para redes de contatos como, por exemplo, a separação de contatos entre aparelhos móveis e veículos. Nesse último caso, podemos analisar em conjunto redes D2D e V2V, o que ainda não foi explorado na literatura.

Apesar da modelagem de redes móveis via redes multicamadas ser intuitiva, ainda não temos muitas abordagens. Isso deve a recente evolução da teoria de redes multicamadas [Kivelä et al. 2014]. Alguns trabalhos exploram como estruturar dados heterogêneos em domínios diferentes [Machado et al. 2015, Kivelä et al. 2014]. Em [Asgari et al. 2016], os autores consideram diferentes tipos de mobilidade, via diferentes meios de transporte, como os domínios de uma rede multicamadas. A partir dessa modelagem, eles projetaram um algoritmo para recuperar a trajetória completa de cada usuário, mesmo com amostras esparsas. O uso das redes multicamadas em aplicações móveis ainda é bastante restrito. Assim, uma direção importante de pesquisa em redes móveis é a modelagem de redes multicamadas para o projeto de aplicações.

1.4.5. Modelos Generativos

Até o momento, apresentamos modelos que apenas representam os dados observados. Como citado anteriormente, é de suma importância avaliar aplicações de redes móveis em larga escala. Apesar da era do *Big Data*, ainda encontramos dificuldade em trabalhar com amostras grandes de mobilidade [Mota et al. 2014]. Normalmente, os dados usados no projeto e avaliação de protocolos são limitados e, portanto, seu funcionamento pode ser questionado. Independente das tentativas de conseguir dados em maior escala, em geral, estamos limitados, o que pode dificultar o avanço da pesquisa na área. Todavia, um recurso muito usado em diversos contextos de simulações de sistemas é o uso de modelos estatísticos generativos. Com eles, podemos, a partir do comportamento dos dados observados, gerar mais amostras probabilisticamente semelhantes às observadas.

Um modelo estatístico generativo é um modelo para gerar aleatoriamente dados observáveis, tipicamente com parâmetros escondidos. Essa tarefa é fundamental nas áreas de Aprendizado de Máquina e Mineração de Dados, por exemplo. Existem diversas abordagens e, justamente por isso, existem diversos modelos generativos de mobilidade disponíveis na literatura [Mota et al. 2014]. De maneira simplificada, a tarefa é, a partir das observações disponíveis, gerar um modelo estatístico capaz de gerar amostras da mesma distribuição que gerou os dados observados. Além de agregar os dados observados, grande parte dos modelos utilizam premissas, buscando uma abordagem Bayesiana do assunto.

Há diversos modelos de mobilidade para diferentes contextos na literatura, porém nenhum modelo é considerado estado da arte. Em grande parte isso ocorre pelo fato de não existirem aspectos de mobilidade observados neles, mas observados em dados reais [Mota et al. 2014]. Os modelos mais famosos de mobilidade individual e veicular são apresentados em [Ekman et al. 2008, Lee et al. 2009, Mei and Stefa 2009, Rhee et al. 2011].

Para ilustrar essa difícil tarefa, apresentamos um modelo simples de mobilidade extraído apenas dos dados observados. Consideramos a rede espacial de mobilidade montada anteriormente. Isto é, construímos uma rede em que cada nó n_i representa um local no espaço e cada aresta direcionada (n_i, n_j) conecta dois nós alcançáveis espacialmente. As arestas possuem pesos $w(n_i, n_j)$ proporcionais ao fluxo de entidades que transitaram entre os dois espaços representados pelos nós. Podemos montar uma cadeia de Markov [Kemeny et al. 1960] usando a rede apresentada. Nela, cada estado da cadeia $X_1, ...X_n$ corresponde a um nó da rede. Cada probabilidade $P(X_j|X_i)$ é definida como $\frac{w(n_i,n_j)}{\sum_{k=1}^{n} w(n_i,n_k)}$. Assim, obtemos uma cadeia de Markov e podemos simular transições entre estados ou, na prática, entre espaços, levando a uma simulação de mobilidade. Na prática, essa simulação não é eficiente por diversas premissas tomadas. A principal é a de que apenas o espaço influencia o movimento das entidades. A probabilidade de movimentação não está condicionada ao tempo neste caso, o que não é verdade na prática. Por exemplo, uma pessoa tem mais probabilidade de se movimentar do centro para a periferia de uma cidade no final do dia do que no início dele. Outra premissa duvidosa é a de que o próximo lugar para o qual nos movimentamos depende apenas de onde estamos. A nossa trajetória por completo define de maneira muito mais precisa a nossa movimentação [Wu et al. 2011].

1.5. Projeto e Aplicações

Esta seção apresenta diversos estudos de casos de pesquisa atuais relacionados ao projeto de redes móveis a partir de dados sensoriados. Para cada um deles, as oportunidades de pesquisas e os desafios são também discutidos.

1.5.1. Redes Celulares e Comunicação Dispositivo-Dispositivo (D2D)

Atualmente, os telefones celulares *smartphones* passaram a ser o dispositivo computacional mais popular entre as pessoas em todos os lugares do mundo. Atualmente, eles produzem uma quantidade de dados imensa todos os dias. Devido a isso, as redes celulares se tornaram uma das principais fontes de dados e as operadoras se tornaram as detentoras desse grande base de dados. A partir da análise desses dados, pode-se melhorar significantemente o desempenho das redes celulares em termos de qualidade de serviço e maximização de receita, por meio de diferentes aplicações tais como balanceamento de carga, planejamento da rede e comunicação D2D [He et al. 2016].

Como discutido em [Zheng et al. 2016], a Figura 1.20 ilustra uma visão geral de como *Big data analytics* pode ser explorada no contexto de redes móveis celulares de comunicação e de acesso à Internet. Os dados podem ser obtidos de diversas fontes como mostra essa figura. São elas: equipamentos dos usuários (*User Equipment* – UE), rede de acesso (*Radio Access Network* – RAN), núcleo da rede (*Core Network* – CN) e provedores (*Internet Service Providers* – ISPs). O volume, a velocidade e variedade desse dados coletados é tão grande que surgem os primeiros desafios para manipulação e armazenamento desses dados. Depois de coletar e armazenar, o desafio seguinte consiste em processar e extrair o conhecimento dos dados, visando otimizar os recursos da rede de forma que os problemas sejam identificados e que seja possível decidir o quê fazer para contorná-los. A seguir, apresentamos uma lista (não exaustiva) de exemplos que envolvem o projeto de redes celulares e a área de análise de dados.

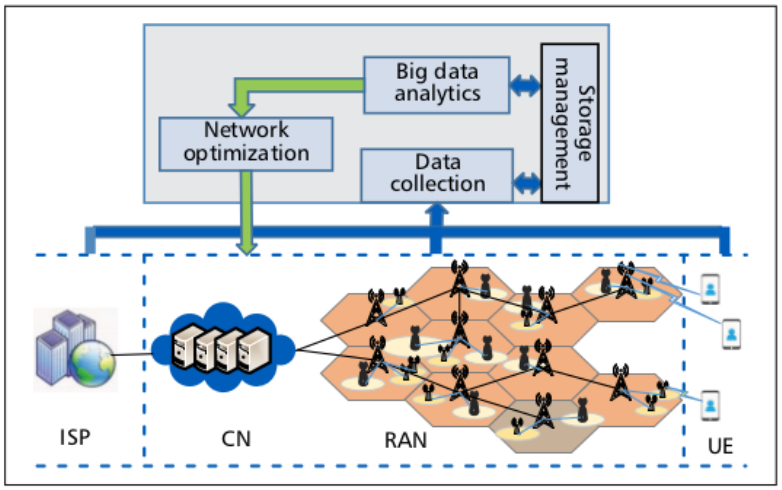


Figura 1.20. Visão geral de uma arquitetura para exploração de dados e sua aplicação em redes celulares [Zheng et al. 2016].

1.5.1.1. Alocação de Recursos e Cache de Conteúdos

Como discutido na Seção 1.3, os recursos de uma rede celular podem variar bastante em termos do domínio espaço-temporal. Além disso, anomalias como eventos sociais de larga escala (e.g., protestos, *shows*) podem ser motivos de desbalanceamento de recursos da rede causando congestionamento de tráfego na rede. Por meio da análise de dados históricos e da predição do tráfego, é possível contornar esse problema. Para tanto, técnicas relacionadas à detecção de anomalias [Chandola et al. 2009] e *Cloud-RAN* [Checko et al. 2015] podem ser combinadas e aplicadas.

Similarmente, o estudo de dados históricos pode ser aplicado a outros cenários tratando-se de redes celulares como, por exemplo, na distribuição de conteúdos. As CDNs (*Content Delivery Networks*) e ICN (*Information Centric Networks*) são redes relacionadas ao armazenamento em *cache* de conteúdos, tendo como um dos principais objetivos distribuí-los reduzindo o tempo de acesso. Uma estratégia interessante, a partir da análise de dados, é otimizar a implantação de servidores *cache* visando diminuir o tempo de resposta e reduzir custos. Por exemplo, a Figura 1.21 ilustra um caso em que servidores *cache* distribuídos podem ser implantados na infraestrutura, visando uma melhor distribuição de conteúdo na rede. Essa implantação é feita a partir da observação de padrões no uso de determinados conteúdos ao longo tempo em determinadas regiões, como foi discutido na etapa de caracterização na Figura 1.13 da Seção 1.3. Nesse sentido, conteúdos mais populares em uma dada região podem ser alocados aos servidores, sabendo que fatores associados à mobilidade, como discutidos anteriormente, também impactam essa implantação e alocação de conteúdos.

1.5.1.2. Análise de *Handovers* e Planejamento de Redes

Existem diferentes motivos para a ocorrência de *handovers* em redes celulares. No entanto, de maneira geral, a causa mais comum de *handovers* consiste da transição de uma chamada

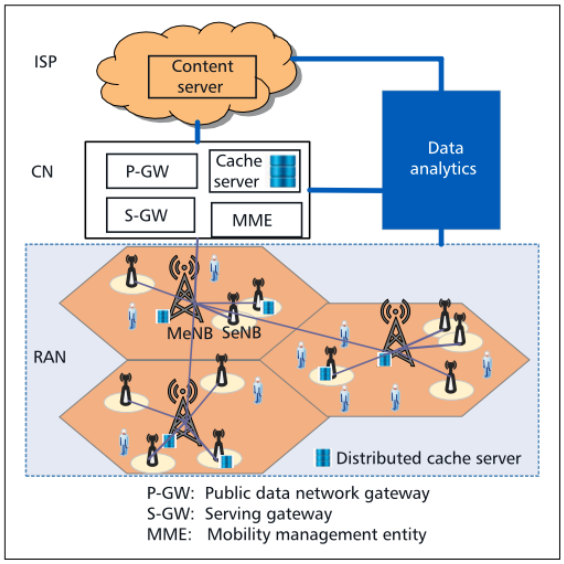


Figura 1.21. Ilustração de implantação de servidores *cache* [Zheng et al. 2016].

em curso ou sessão de dados de uma célula para outra. Em um estudo sobre mobilidade de usuários de telefonia celular com base em *handovers* [Demissie et al. 2013], os autores analisaram um conjunto de dados de Lisboa em Portugal com média de 2.5 milhões de *handovers* de um total de 7.2 milhões de chamadas.

A Figura 1.22(a) mostra o volume de chamadas em cada estação base entre 8:00 e 9:00 da manhã do dia 12 de abril de 2010. A Figura 1.22(b) mostra o fluxo de *handovers* entre as estações da rede. As setas representam as transições e a intensidade de cor representa o volume de *handovers*. Pode-se perceber que, apesar de existir uma maior quantidade de antenas na região central, a maioria das movimentações no período analisado concentra-se na região leste e parte da região norte. Do ponto de vista da operadora, é interessante observar que padrões de movimentações existem para otimizar os canais de comunicação.

A Figura 1.22(c) mostra, para cada torre, a relação de dois tipos de *handovers*: *outgoing handovers*, quando o usuário efetua uma chamada e sai da célula; e *incoming handovers*, quando o usuário entra em uma célula e finaliza a chamada. O tamanho do círculo é definido em relação ao número total de *handovers* na célula. Os autores fornecem alguns resultados que são pertinentes para o planejamento de redes celulares: torres celulares caracterizadas por alto e balanceado número de *incoming handovers* e *outgoing handovers* estão localizadas próximas às principais rodovias, visto que essas possuem um maior fluxo de pessoas. Existe uma relação entre a presença de pessoas e o número de *incoming handovers*, justificando torres próximas aos pontos de chegadas ou interesse.

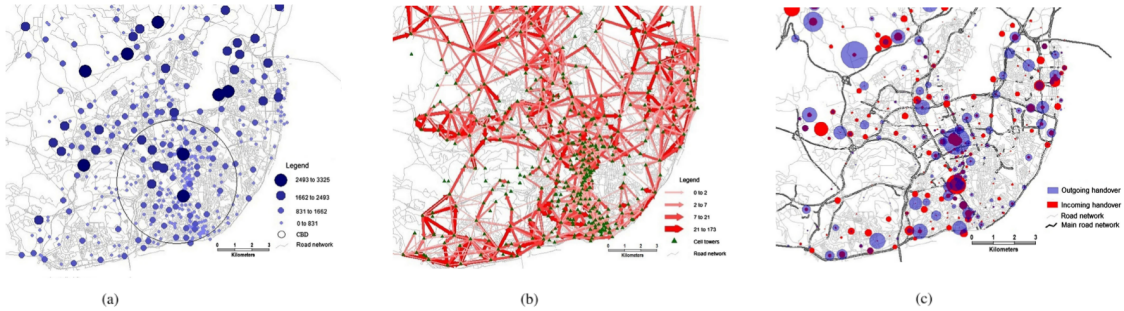


Figura 1.22. Exploração de informações de *handovers* de rede celular em um ambiente urbano. (a) Volume de chamadas por torre celular em Lisboa. (b) *Handovers* entre as torres celulares. (c) *Incoming handovers* e *Outgoing handovers* na cidade de Lisboa [Demissie et al. 2013].

1.5.1.3. Comunicação D2D

O aumento na demanda por dados nas redes celulares traz cada vez mais desafios para as operadoras de telefonia móvel. De forma a atender essa crescente demanda, as redes celulares de quinta geração (5G) propõem mudanças significativas na arquitetura de comunicação [Akyildiz et al. 2016]. Entre as novas características presentes no padrão 5G destaca-se a comunicação Dispositivo-Dispositivo ou *Device-to-Device* (D2D).

O termo D2D refere-se à transmissão direta de dados entre os dispositivos móveis da rede celular (e.g., *smartphones* e *tablets*) que estejam suficientemente próximos uns dos outros. Esse tipo de comunicação tem o potencial de descarregar a demanda das estações bases que, nas redes celulares tradicionais, são responsáveis por toda a transmissão de dados. Nas redes celulares atuais (por exemplo, 3G e 4G), quando dois dispositivos usam a Internet móvel para a compartilhar um arquivo entre si, o arquivo é primeiramente transmitido do dispositivo origem para a estação base e, posteriormente, transmitido da estação base para o dispositivo destino. Isso ocorre mesmo que os dispositivos origem e destino se encontrem lado a lado. Esse modelo, além de aumentar o atraso na comunicação (considerando transmissões D2D de apenas um salto), gera uma alta sobrecarga na estação base, que deve atender a um grande número de dispositivos ao mesmo tempo. Nesse contexto, a comunicação D2D permite o descarregamento da alta demanda por dados na estação base, possibilitando a melhoria dos serviços oferecidos aos clientes, incluindo melhores taxas de transmissão.

Como discutido em [Laya et al. 2014], o suporte à comunicação D2D modifica os planos de dados e controle das redes celulares, como explicado a seguir:

- **Plano de dados:** A comunicação D2D provê um novo caminho para os dados, que pode envolver um ou mais dispositivos móveis na rede, para enviar mensagens do dispositivo de origem para o dispositivo destino sem a necessidade dessas mensagens trafegarem pela estação base.

- **Plano de Controle:** Nas redes D2D, as mensagens de controle, que definem as políticas de encaminhamento D2D dos dados, podem ser transmitidas de forma distribuída, utilizando a própria conectividade D2D, ou de forma centralizada, com

o auxílio da estação base, que é capaz de se comunicar diretamente com todos os dispositivos móveis de sua célula. Vale ressaltar que a combinação de ambas as formas de sinalização também é possível.

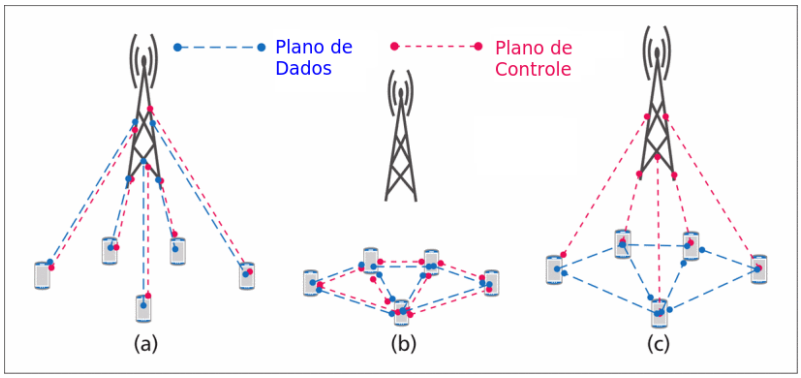


Figura 1.23. Possibilidades de arquiteturas em redes móveis e celulares: (a) Arquitetura tradicional centralizada; (b) Arquitetura *Ad Hoc* distribuída; (c) arquitetura D2D híbrida (figura adaptada de [Nunes et al. 2016c])

A Figura 1.23 mostra os tipos de arquiteturas possíveis para os planos de dados e controle. Em contraste com as redes celulares tradicionais e com as redes puramente *ad hoc* mostradas nas Figuras 1.23(a) e 1.23(b), respectivamente, a nova geração de comunicação celular oferece suporte para um plano de controle centralizado e um plano de dados distribuído, que juntos possibilitam a redução da demanda das estações bases por meio das transmissões de dados D2D (Figura 1.23(c)). A possibilidade de um plano de controle centralizado é uma diferença fundamental das redes D2D para as redes puramente *ad hoc*, em que os algoritmos de encaminhamento de dados devem ser completamente distribuídos, como mostrado na Figura 1.23(b). Vale ressaltar que as mensagens de controle são, em geral, muito menores que os dados sendo transmitidos. Por isso, mesmo que as mensagens de controle permaneçam centralizadas na estação base, a carga gerada por essas mensagens é praticamente desprezível quando comparada às taxas de transmissão requeridas por dados como vídeos, multimídia, jogos e outras aplicações.

Outro aspecto crucial nas redes D2D é a mobilidade, já que os elementos que provêm a comunicação D2D, ou seja, os próprios dispositivos móveis, são carregados por pessoas. Por sua vez, as pessoas possuem rotinas e trajetórias diárias que muitas vezes são previsíveis. Portanto, um melhor entendimento dos padrões de mobilidade das pessoas é fundamental para o projeto eficaz de protocolos para redes D2D. Nesse contexto, a análise de dados, em especial dados de mobilidade, é uma ferramenta imprescindível para extrair informações relevantes para o projeto das redes D2D, possibilitando o aprimoramento dos protocolos e serviços oferecidos aos usuários da rede.

Entre os desafios de pesquisa em redes D2D que podem se beneficiar da análise de dados de mobilidade, ressaltam-se os seguintes:

- **Alocação e gerenciamento de recursos:** Recursos em redes D2D são divididos em duas classes: recursos da rede e recursos dos dispositivos. Associados com a primeira classe estão desafios como a alocação dinâmica de canais de frequência

para a comunicação direta entre os dispositivos [Song et al. 2014, Yin et al. 2013]. Problemas de otimização como (i) alcance da transmissão versus interferência entre dispositivos; e (ii) taxas de transmissão versus largura do espectro alocado para cada par de dispositivos estão associados à primeira classe. A segunda classe trata do gerenciamento dos recursos dos dispositivos [Silva et al. 2017a, Silva et al. 2017b], como memória e energia, de forma que esses dispositivos possam colaborar com a rede possibilitando transmissões diretas de dados, mas ao mesmo tempo não sejam prejudicados pela utilização exagerada de seus recursos pela rede.

- **Comunicação D2D de múltiplos saltos e oportunística:** Nas redes D2D, a comunicação de múltiplos saltos envolve múltiplos dispositivos, formando um caminho D2D da origem de um determinado dado até o destino. Esse caminho pode ser definido oportunisticamente, utilizando os encontros entre pares de dispositivos em tempo real, de acordo com políticas de controle definidas pela estação base. Esse tipo de comunicação, em geral, envolve atrasos elevados e é adequado para alguns tipos específicos de dados, em que a entrega mais lenta dos conteúdos pode ser tolerada [Nunes et al. 2016c]. Exemplos de aplicações são a atualização de aplicações e transmissões de propagandas em vídeos. Esses tipos de conteúdos podem ser descarregados da estação base por meio da comunicação D2D oportunística de múltiplos saltos.

- **Dispositivos heterogêneos:** As redes D2D são vistas não só como uma arquitetura de comunicação entre *smartphones*, mas também como uma rede que inclui os mais diversos sistemas computacionais móveis para possibilitar uma comunicação eficiente e inteligente. Entre esses dispositivos, podemos considerar carros com capacidade de comunicação (VANETs), infraestrutura de cidades inteligentes e objetos inteligentes na Internet das Coisas.

- **Contexto social em redes D2D:** O contexto social, i.e., relações sociais entre usuários da rede podem ser utilizados, por exemplo, para prever subconjuntos de usuários da rede que têm maior probabilidade de estarem próximos uns aos outros frequentemente. Essa informação, por sua vez, pode ser utilizada para definir políticas eficientes de alocação de recursos para usuários que têm mais chance de se encontrarem e compartilharem arquivos entre si [Li et al. 2014].

No restante desta seção será apresentado um estudo de caso que exemplifica alguns dos desafios de pesquisa discutidos acima. Nesses exemplo, padrões extraídos por meio da análise de dados de mobilidade são utilizados para projetar protocolos e soluções para redes móveis D2D.

Mobilidade, Encontros e seus Papéis na Comunicação D2D

Nas redes móveis, ***encontros*** são definidos como eventos em que duas ou mais entidades (e.g., pessoas, veículos) ocupam a mesma região em um mesmo intervalo de tempo, ou seja, essas entidades encontram-se fisicamente próximas. Em outras palavras, os encontros são interseções espaço-temporais nas trajetórias móveis das entidades (veja duração de contatos e tempo entre contatos na Seção 1.3.3). Encontros são particularmente

importantes em redes D2D pois eles são as oportunidades para transmitir dados diretamente entre os dispositivos, permitindo o descarregamento da estação base. Portanto, um melhor entendimento das leis que governam os encontros entre entidades em redes móveis podem favorecer significativamente o projeto eficaz de protocolos para redes D2D.

Como discutido na Seção 1.2, um tipo de *dataset* tipicamente utilizado para esse tipo de estudo são os *traces* de contatos. Alguns exemplos de *traces* de contatos públicos são o *traces* do MIT [Eagle and Pentland 2006], de Dartmouth [Henderson et al. 2008], da USC [Hsu and Helmy 2005] e de NCCU [Tsai and Chan 2015]. Após um preprocessamento que assegure a qualidade e confiabilidade dos dados coletados, esses *traces* podem ser utilizados na análise e caracterização das propriedades dos encontros nessas redes.

Diversos trabalhos na literatura buscam caracterizar as propriedades dos encontros visando, entre outros aspectos, a aplicação dessas propriedades no projeto de redes *ad hoc* e mais recentemente de redes D2D.

Encontros de Grupos e Comunicação D2D Oportunística

A seguir, é apresentado o protocolo Groups-Net [Nunes et al. 2016d, Nunes et al. 2016a] como um estudo de caso que mostra como os padrões de mobilidade, identificados e caracterizados a partir *traces* de contatos, podem ser utilizados na modelagem e no projeto de soluções para redes D2D.

O Groups-Net é um protocolo de encaminhamento oportunístico para redes D2D de múltiplos saltos. Ele foi projetado com o objetivo de maximizar a quantidade de mensagens entregues utilizando a comunicação D2D e ao mesmo tempo minimizar as retransmissões de mensagens entre os dispositivo móveis da rede D2D. Para alcançar esse objetivo, as mensagens devem ser encaminhadas entre a origem e o destino utilizando rotas de múltiplos saltos que envolvam o menor número possível de dispositivos móveis intermediários, ou seja, rotas mais curtas. Uma alta taxa de entrega permite que mais dados sejam transmitidos usando a comunicação D2D, resultando em uma maior descarga da demanda da estação base. Por outro lado, uma grande quantidade de retransmissões impacta negativamente os dispositivos da rede, por exemplo, reduzindo a duração média de suas baterias.

O Groups-Net explora a regularidade dos encontros de grupos para realizar o roteamento das mensagens por meio de um modelo que segue as duas premissas abaixo:

- **Regularidade dos encontros de grupos:** Assume-se que grupos que se encontraram mais frequentemente no passado recente (por exemplo, cinco vezes nas últimas semanas) têm uma chance maior de se encontrarem novamente no futuro próximo do que grupos que se encontraram com uma frequência menor (por exemplo, uma vez no último mês). Para modelar esse comportamento, uma probabilidade de reencontro é assinalada a cada grupo de acordo com um processo de Poisson onde a taxa de reencontros é estimada a partir do número médio de reencontros por unidade de tempo no passado recente. A utilidade de saber quais grupos têm maior probabilidade de se reencontrarem reside no fato de que em uma reunião de um grupo, um determinada mensagem pode ser transmitida para todos os nós envolvidos.

- **Interseção nos membros de dois grupos:** Embora uma mensagem possa ser transmitida para todos os membros de um determinado grupo durante um encontro do grupo, muitas vezes a origem e o destinatário de uma determinada mensagem não pertencem a nenhum grupo em comum. Dessa forma, é necessário que a mensagem seja propaganda entre diferentes grupos. É natural que uma pessoa se encontre regularmente com mais de um grupo de pessoas (e.g., colegas de trabalho, familiares, pessoas que utilizam o mesmo ônibus no mesmo horário). Essas pessoas podem sevir para propagar as mensagens entre diferentes grupos. Com esse objetivo, a probabilidade de uma mensagem ser transmitida entre dois determinados grupos ($G1$ e $G2$) é calculada com base no número de membros que estão regularmente presentes nos encontros de ambos os grupos $G1$ e $G2$.

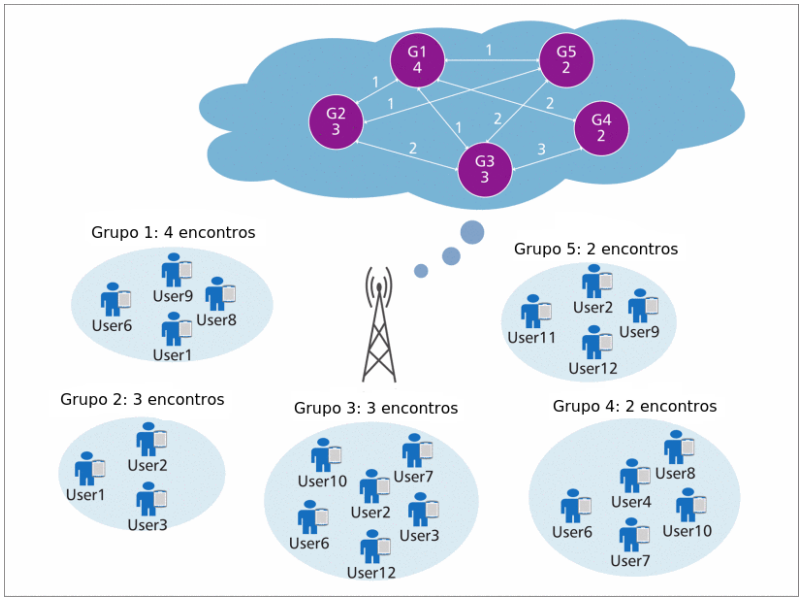


Figura 1.24. Operação do Groups-Net: A estação base mantém um grafo onde o peso dos nós representa o número de encontros de cada grupo e o peso das arestas o número de membros em comum entre dois grupos. As probabilidades de cada nó e aresta são derivadas de seus pesos e a mensagem é transmitida entre a origem e o destino pela rota grupo-a-grupo mais provável (figura adaptada de [Nunes et al. 2016c]).

Utilizando as duas probabilidades descritas acima, o Groups-Net encaminha as mensagens por meio da rota grupo-a-grupo mais provável entre a origem da mensagem e seu destinatário. Os detalhes do algoritmo de seleção da rota grupo-agrupo mais provável e a definição das fórmulas para obtenção das probabilidades descritas acima podem ser encontrados em [Nunes et al. 2016a]. Vale mencionar que a detecção dos encontros dos grupos pode ser realizada de forma distribuída, pelos próprios nós envolvidos no encontro do grupo.

Como ilustrado na Figura 1.24, o número de encontros de cada grupo é reportado (pelos próprios nós envolvidos nos encontros) para a estação base, utilizando mensagens de controle. A estação base, por sua vez, mantém um grafo em que cada nó representa um grupo. O peso das arestas nesse grafo representa o número de usuários que pertencem a

ambos os grupos interligados por uma aresta. Por exemplo, na Figura 1.24, os usuários 6, 7 e 10 pertencem a ambos os Grupos 3 e 4. Portanto a aresta interligando esses grupos tem peso 3. Por fim, o peso dos nós no grafo é definido pelo número de vezes que cada grupo se encontrou no passado recente (e.g., no último mês). Quando um dado nó origem deseja transmitir uma mensagem para um determinado destino, a estação base utiliza o grafo mostrado na Figura 1.24 para computar a rota grupo-a-grupo mais provável e envia essa rota para a origem da mensagem, que procede com o encaminhamento da mensagem de acordo com a rota recebida da estação base.

Para validar o modelo proposto em cenários reais, os autores aplicaram o Groups-Net nos *traces* de contatos do MIT e de Dartmouth. Nesses cenários, a política de encaminhamento proposta demonstrou um desempenho melhor que soluções anteriores, considerando o número de mensagens entregues com sucesso utilizando a comunicação D2D (ou seja, número de mensagens descarregadas da estação base) *versus* número de retransmissões D2D necessárias para a entrega das mensagens.

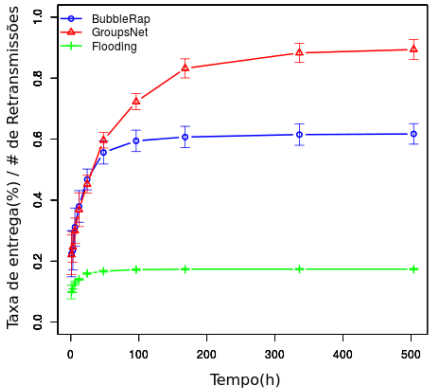


Figura 1.25. Comparação da relação benefício-custo do Groups-Net com a de protocolos de encaminhamento anteriores.

A Figura 1.25 mostra a comparação do Groups-Net com outros protocolos anteriores. Nela é apresentada a relação benefício-custo, definida como o número de mensagens entregues dividido pelo número de retransmissões necessárias para a entrega, ao longo do tempo de operação da rede. Esses resultados mostram um exemplo de como a análise e a caracterização de dados, combinadas com uma modelagem consistente do problema, podem ser aplicados para propor soluções que melhoram a eficiência de redes móveis *ad hoc* e, neste exemplo, redes D2D oportunísticas.

1.5.2. Redes Veiculares

As redes veiculares são compostas por veículos providos de capacidade de comunicação sem fio entre eles (*vehicle-to-vehicle*, ou apenas V2V), ou com estações infraestruturadas (*vehicle-to-infrastructure*, ou apenas V2I). O objetivo de uma rede veicular é prover serviços antes não viáveis para os veículos e seus usuários. As aplicações para esse tipo de rede vão desde notificações de alertas de trânsito até a entrega de conteúdo de entretenimento.

As características das redes veiculares fazem com o que o projeto da rede seja uma tarefa não-trivial. Em geral, as redes veiculares devem estar preparadas para funcionarem em larga escala, dada a enorme quantidade de veículos existentes em alguns centros urbanos. Além disso, a densidade da rede é variada, indo de topologias muito esparsas a muito densas, dependendo da região e do horário. A topologia da rede também é extremamente dinâmica devido à mobilidade dos veículos, tornando os contatos altamente intermitentes. Por fim, as cidades possuem características distintas, o que dificulta ainda mais a definição de padrões de projetos de redes.

Com todas essas dificuldades, os projetistas de redes veiculares podem se beneficiar significativamente de conhecimentos extraídos de grandes volumes de dados sobre a mobilidade urbana. No restante desta seção, são apresentados os trabalhos da literatura que exploram esses dados para obterem conhecimento útil que pode ser utilizado em diferentes aspectos das redes veiculares.

1.5.2.1. Roteamento e Disseminação de Dados

Um dos grandes desafios para os pesquisadores na área de redes veiculares está no roteamento de pacotes e disseminação de dados. As características peculiares desse tipo de rede, como larga escala, mobilidade constante, diferentes densidades ao longo do tempo e espaço, e demanda por entrega eficiente e rápida em alguns casos, fazem com que essas tarefas não sejam triviais. Portanto, nos últimos anos, muitos pesquisadores têm explorado dados de mobilidade veicular existentes para conhecer melhor a conectividade das redes veiculares e, com isso, elaborar soluções mais apropriadas. A seguir, são descritos os trabalhos que fazem uso de dados existentes para propor soluções de roteamento e disseminação em redes veiculares.

Em [Celes et al. 2013], os autores projetaram um protocolo de roteamento ciente da mobilidade diária dos veículos para realizar disseminação orientada à geo-localização (*geocast*). Para isso, primeiramente foi feita a caracterização das trajetórias de veículos de um conjunto de dados da cidade de Borlange na Suécia, com o objetivo de se conhecer os padrões de mobilidade. O protocolo realiza o *geocast* a partir das informações dos padrões de mobilidade dos veículos da rede, reduzindo o atraso do tempo de entrega e aumentando a taxa de entrega de mensagens.

Em [Monteiro et al. 2012], os autores estudaram métricas de redes complexas em redes veiculares e, com base nos resultados, sugeriram melhorias para protocolos existentes. Esse estudo avalia características dos grafos de contatos em redes veiculares urbanas e rodoviárias em termos de grau dos veículos, coeficiente de agrupamento e caminhos mínimos. Os dados utilizados são do modelo de *Automata Cellular* [Tonguz et al. 2009] e da rodovia I-80 fornecido pelo *Berkeley Highway Laboratory* [14]. Os resultados obtidos desses cenários levaram a conclusões que ajudaram a melhorar um protocolo de disseminação existente, o UV-CAST. No caso, os autores propuseram dois mecanismos: um para aumentar o alcance da rede em ambientes esparsos e outro para diminuir o impacto de mensagens *broadcast* em ambientes densos. Para isso, cada veículo avalia a quantidade de vizinhos por meio de mensagens de *beacons*. Com base neste valor, ele é capaz de inferir se está em uma região

[14] https://www.fhwa.dot.gov/publications/research/operations/06137/

com pouca densidade ou densa, e assim tomar a melhor decisão. Mais especificamente, em regiões esparsas, o veículo retransmite uma mensagem com probabilidade maior, para tentar aumentar o alcance. Por outro lado, em cenários densos, essa probabilidade diminui.

Em [Cornejo et al. 2013], os autores utilizaram a base real de táxis de São Francisco para criar um modelo de rede. Com base nesse modelo, foi proposto um algoritmo distribuído chamado *CabChat* que permite priorizar mensagens *broadcast*. O modelo e o algoritmo foram validados por meio de experimentos, que demostraram serem eficientes, e o *CabChat* foi capaz de entregar mensagens com prioridade dentro de um limite de tempo estabelecido.

O trabalho desenvolvido por [Amici et al. 2014] utilizou uma base real de mobilidade da cidade de Roma para avaliar o desempenho de um protocolo de disseminação epidêmica para redes veiculares. Os dados foram coletados de *smartphones* instalados em táxis e representam a posição de cada veículo ao longo do tempo durante um mês. A principal contribuição pode ser vista como uma nova base de dados reais que permitirá a avaliação e validação de soluções para redes veiculares.

Outro conjunto de dados de mobilidade em larga escala é utilizado por [Trullols-Cruces et al. 2015] para avaliar a disseminação epidêmica em redes veiculares. Nesse trabalho, os dados realistas da cidade de Zurique, na Suíça, são explorados para se criar um modelo de disseminação epidêmica. O modelo é então validado em outro conjunto de dados da cidade de Madri, na Espanha. Como principal conclusão, os autores observaram que mensagens maliciosas podem infectar muitos milhares de veículos em grandes áreas, em poucos minutos. Outro ponto importante está relacionado à localidade geográfica de partida da mensagem maliciosa, que pode fazer com que a infestação seja ainda mais rápida. Com base nos resultados obtidos, os autores propuseram alterações no modelo para prevenir a contaminação. A solução proposta utiliza a rede celular para atualizar veículos escolhidos de acordo com suas localizações geográficas e, assim, esses veículos seriam responsáveis por interromper a contaminação.

Em geral, a maioria dos trabalhos estudam soluções para disseminação de dados partindo da rede infraestruturada (e.g., Internet) em direção aos veículos. Por outro lado, em [Stanica et al. 2013], os autores utilizam uma base de mobilidade realista e de larga escala para propor uma solução de entrega de dados coletados pelos veículos para servidores localizados na rede infraestruturada. A solução proposta pelos autores visa escolher determinados veículos, sendo o menor número possível, com base nos resultados de contatos obtidos da análise dos dados de mobilidade. A escolha desses veículos foi modelada como o problema de *Conjunto Mínimo Dominante* em grafos, e os resultados demonstraram que é possível reduzir significativamente os dados enviados por meio de infraestrutura.

A proposta de [Liu et al. 2011] visa utilizar veículos estacionados como unidades estáticas de comunicação, com o objetivo de aumentar a conectividade da rede veicular sem a necessidade de implantação de infraestrutura que cobre 100% da área. Para validar essa ideia, os autores primeiramente avaliaram dados estatísticos sobre estacionamentos em três locais diferentes: Montreal (Canadá), Ann Harbor e Hattiesburg (EUA). Esses dados revelam que uma grande quantidade de veículos fica estacionada por um tempo significativo. Além disso, eles caracterizaram dados reais de tráfego e observaram que os

veículos parados superam significativamente os em movimento durante qualquer hora do dia. Para concluir, os autores realizaram simulações com base em dados de seis semanas da área de Chengdu, China. Os resultados mostraram que, ao se utilizar veículos estacionados como parte da rede veicular, todas as métricas de desempenho avaliadas alcançaram valores melhores.

A conectividade dos táxis de São Francisco, EUA, é explorada por [Hoque et al. 2014] para se criar um novo modelo de propagação de mensagens em redes veiculares. Para isso, os autores primeiramente definiram algoritmos para análises desses tipos de dados em larga escala. Em seguida, os algoritmos foram utilizados para analisar a conectividade dos veículos, principalmente em termos de alcance em múltiplos passos. A principal contribuição desse trabalho é a solução de análise dos dados, que permite processar grandes volumes de dados de forma eficiente em termos de espaço de armazenamento e tempo de processamento. Com isso, novas soluções que utilizam grandes volumes de dados podem ser propostas com mais facilidade.

Em [Chen et al. 2013], os autores exploram informações sobre componentes conectados de dados reais de mobilidade no projeto de algoritmos de roteamento em redes veiculares. Os dados utilizados representam as localizações de táxis das cidades de São Francisco (EUA) e Shenzhen (China). Os resultados revelaram informações importantes relativas à estabilidade e localização dos componentes maiores, que podem ser considerados os mais relevantes para a disseminação de dados. Os conhecimentos alcançados com os resultados do trabalho podem ser utilizados para a definição de algoritmos de roteamento.

1.5.2.2. Replicação e Entrega de Conteúdo

Além da disseminação de dados, outra área que vem se destacando recentemente é a entrega de conteúdo em redes veiculares [Silva et al. 2016b]. A entrega envolve, além de algoritmos para disseminar, a escolha de servidores replicadores para se formar uma rede de entrega de conteúdo (CDN) que irá atender aos clientes da maneira mais adequada. Alguns trabalhos da literatura exploram dados de mobilidade veicular existentes para a escolha de nós replicadores na rede veicular, como discutido a seguir.

Em [Gossa et al. 2008], os autores descrevem um dos primeiros trabalhos nesta área. O objetivo é propor um sistema que escolhe nós móveis replicadores visando manter um número pequeno deles sem impactar na qualidade do serviço. Para isso, os autores propuseram um modelo de predição de mobilidade que utiliza dados de mobilidade como entrada. O modelo foi avaliado em um cenário de dados reais coletados da cidade de Viena, Áustria. Os resultados revelaram que a predição da mobilidade com base em dados históricos pode melhorar a escolha de replicadores.

Em [Silva et al. 2015c], os autores exploram o conhecimento de mobilidade veicular para escolher os melhores replicadores de conteúdo em uma rede veicular. Para isso, são utilizadas informações de um modelo de mobilidade [Silva et al. 2015b] que define como os veículos se movem em termos de regiões origem-destino. Esse modelo foi desenvolvido com base em dados realistas de mobilidade veicular da cidade de Colônia, na Alemanha. Com base nesse modelo, é possível estimar os contatos dos veículos ao longo de toda a área da rede em determinados instantes de tempo e, assim, escolher bons replicadores que

sejam capazes de entregar o conteúdo a todos os veículos de maneira eficiente e com baixa redundância. Os resultados mostraram que foi possível cobrir praticamente toda a rede de maneira mais eficiente, em termos de recursos de rede, que a disseminação epidêmica.

O mesmo modelo [Silva et al. 2015b] é utilizado por [Silva et al. 2016c] para se formar uma rede de entrega de conteúdo, porém considerando conteúdos geo-referenciados, ou seja, que devem ser entregues somente dentro de uma região de interesse. Para isso, os autores propuseram um algoritmo que escolhe veículos replicadores de forma a cobrir a região de interesse sem muita redundância. Os resultados mostraram que foi possível alcançar uma boa taxa de entrega a um baixo custo de rede.

1.5.2.3. Planejamento de Infraestrutura

Uma rede veicular é composta por veículos e estações de acostamento (RSUs – *Roadside Units*). As RSUs representam pontos de acesso que os veículos podem ter com a infraestrutura e são extremamente relevantes para se manter a cobertura da rede e garantir a qualidade dos serviços. Com isso, surge uma importante questão: onde instalar as RSUs? Considerando que o custo de instalação e manutenção é alto, pode não ser viável implantar RSUs em 100% da área de cobertura da rede. Por outro lado, regiões importantes, que não sejam cobertas por RSUs, podem enfrentar momentos de desconexão, dependendo da densidade da rede. Portanto, é importante que sejam escolhidos locais estratégicos para se implantar as RSUs, de forma a equilibrar o custo e o benefício. Esse problema também vem sendo abordado por pesquisadores nos últimos anos, sendo que alguns trabalhos utilizam dados existentes sobre as redes veiculares em suas propostas, como apresentado a seguir.

Os dados de mobilidade da cidade de Colônia, Alemanha, são explorados por [Silva et al. 2015a] para a decisão sobre onde implantar RSUs. A solução propõe a divisão da rede viária em partições similares, e considera a taxa de mobilidade entre as partições para escolher aquelas que tendem a ser mais impactadas positivamente com a implantação de RSUs. Em comparação com outros estudos, o trabalho conseguiu alcançar melhores taxas de coberturas.

O mesmo conjunto de dados é explorado por [Silva et al. 2016a], também para a implantação de RSUs em uma rede veicular. Porém, o objetivo agora é que a escolha seja baseada na capacidade da rede em entregar conteúdos de diferentes tamanhos a uma maior quantidade de veículos. Em outras palavras, o tempo de conexão dos veículos com as RSUs é fundamental para que a entrega de um conteúdo seja efetivada. Para isso, os autores propuseram uma estratégia que considera o percentual de veículos cobertos por um determinado tempo. Os resultados também foram satisfatórios, indicando que o conhecimento do cenário é fundamental para as tomadas de decisões em termos de implantação de RSUs.

1.5.2.4. Fluxo de Mobilidade

O estudo do fluxo de veículos também é uma área recente que vem sendo explorada. Ao se conhecer padrões relacionados ao fluxo da mobilidade veicular, é possível tomar decisões em termos de rotas de deslocamento, roteamento de dados e planejamento de infraestrutura, por exemplo. Os trabalhos descritos abaixo utilizam dados em larga escala para tentar identificar e prever padrões de mobilidade.

Em [Wang et al. 2014], os autores modelaram a regularidade do fluxo de tráfego de veículos nas rodovias e interseções a partir da mineração de trajetórias. A partir disso, foi elaborado um método para calcular o atraso esperado de entrega de mensagens de uma origem a um destino.

Em [Zhang et al. 2013], os autores exploram os dados reais de táxis da cidade de Shanghai para criar um modelo que identifica vias importantes para o fluxo de veículos. Com base nos pontos visitados pelos táxis em cada via, os autores observaram que aproximadamente 10% das vias são responsáveis por 90% dos pontos. Enquanto isso, 33% das vias não foi visitada nenhuma vez por algum táxi monitorado. Com o conhecimento das vias relevantes, os autores avaliaram o intervalo entre contatos dos veículos par-a-par e definiram um modelo de mobilidade que leva em consideração o fluxo de veículos observado.

O conjunto de dados de Shanghai também é estudado por [Zhu et al. 2011], principalmente em termos de intervalo entre contatos. Os autores mostraram que a chance de táxis se encontrarem é alta, diferentemente de outros estudos que fazem essa análise com base em dados de contato entre pessoas. Além disso, foram descobertas áreas populares, em que táxis visitam com frequência. Com base nessas observações, foi proposto um modelo que considera essas regiões populares e o intervalo entre contatos dos veículos.

Em [Xiao et al. 2014], os autores estudaram a regularidade na mobilidade veicular com base nos dados de táxis das cidades de Shanghai e Beijing, na China. Eles dividiram as cidades em regiões, e avaliaram a frequência de visitas dos veículos em cada região. Foram encontrados padrões de visitas em termos espaciais e temporais, que podem ser previstos com uma boa precisão, conforme a avaliação apresentada.

1.5.2.5. Considerações

Nos últimos anos, vários conjuntos de dados, reais e sintéticos, referentes à mobilidade veicular, se tornaram disponíveis publicamente. Com isso, pesquisadores começaram a explorar esses dados com o intuito de conhecer melhor a mobilidade veicular e, assim, oferecerem propostas adequadas de protocolos, aplicações e serviços. Surgiram, então, propostas que utilizam esse conhecimento para projetarem soluções de roteamento e disseminação de dados, replicação e entrega de conteúdo, implantação de infraestrutura, e de alternativas relativas ao fluxo de veículos.

No entanto, ainda há muito espaço para melhorias. Primeiramente, é importante que novos conjuntos de dados atuais sejam disponibilizados por iniciativas públicas e privadas, para que possam ser explorados por pesquisadores. Em seguida, é preciso caracterizar

esses dados para que novos conhecimentos sejam obtidos. Por fim, esses conhecimentos devem ser utilizados para que novas soluções, em diferentes aspectos, sejam propostas.

1.5.3. Internet das Coisas Móveis (IoMT)

A Internet das Coisas [Atzori et al. 2010] vem recebendo uma atenção especial tanto da academia quanto da indústria nos últimos anos e tem um papel fundamental para a convergência entre os mundos físico e informacional [Gubbi et al. 2013]. Tendo como principal característica a utilização de sensores capazes de interagir com o mundo físico, por meio de monitoramento e atuação, e a sua comunicação através da Internet, estes sensores são os "olhos" e "ouvidos" para as aplicações de IoT.

A gama de aplicações baseadas nos sensores da IoT só cresce em todo o mundo, com o potencial de geração de dados das mais variadas fontes e tipos sem precedentes. Com aplicações que vão desde automação residencial [Coronado and Iglesias 2016] e de agricultura [Wu et al. 2014a] a cidades inteligentes [Zanella et al. 2014] e monitoramento climático [Greengard 2014], a quantidade de dados geradas por estes cenários é impressionante. A Internet das Coisas Móveis (IoMT, *Internet of Mobile Things*) surge como uma expansão do conceito de IoT, ampliando ainda mais o espectro de aplicações neste novo cenário [Nahrstedt et al. 2016]. Nesta seção, discutimos as principais características e desafios que essa ampliação introduz aos desafios da IoT estática, enfatizando aspectos relacionados à mobilidade dos dispositivos e o impacto nos dados gerados.

1.5.3.1. Definição e exemplos

Podemos compreender a IoMT como uma rede composta de dispositivos móveis capazes de interagir com o mundo físico na qual estão inseridos, seja através de sensores capazes de realizar monitoramento do estado de uma dada entidade física e/ou atuadores capazes de alterar o estado de uma entidade. Este conceito deriva-se dos paradigmas que a compõem: IoT e redes móveis. Desta forma, as aplicações de IoMT podem ser consideradas como um subconjunto das aplicações de IoT. Sua principal restrição diz respeito à capacidade que seus dispositivos possuem de alterar sua localização durante sua operação.

Para exemplificar esta restrição mais formalmente, seja s um dispositivo sensor da IoMT coletando dados de uma determinada entidade, para que cada uma de suas leituras s_i seja válida para uma aplicação, é necessário considerar tanto os valores monitorados quanto as suas informações espaço-temporais. Em outras palavras, $s_i = (v,t,l)$, onde i representa a i-ésima leitura do sensor, v corresponde ao valor coletado pelo sensor, no instante de tempo t e na localização demarcada por l. Diversas aplicações que seguem estas restrições compreendem o escopo da IoMT. A seguir, apresentamos alguns exemplos para ilustrar seu potencial.

Internet dos veículos. A Internet dos Veículos (IoV – *Internet of Vehicles*) surge como uma evolução do conceito de VANETs por considerar um veículo (e.g., carros, ônibus ou trens) como uma entidade inteligente conectada à IoT [Yang et al. 2014]. Neste sentido, os veículos passam a ser vistos como sensores capazes de monitorar o ambiente no qual estão inseridos (e.g., tráfego, condições climáticas, poluição) e, além disso, capazes de

prover novos serviços inteligentes por meio da integração entre humanos, veículos, coisas e o ambiente.

Para isto, além das tecnologias de comunicação em rede, já propostas pelas VANETs, agora dispomos de novas tecnologias (e.g., *big data analytics*, *deep learning*, computação cognitiva, inteligência artificial) para a operação coordenada e inteligente dos veículos. Por meio da aquisição da grande quantidade de dados gerados por todos esses veículos, será possível a criação de serviços que visam aprimorar desde a operação da própria rede dos veículos e os serviços oferecidos aos seus passageiros, até aplicações para cidades inteligentes, como estacionamentos inteligentes, redução de tráfego e poluição.

Robôs móveis. Outra possibilidade que surge a partir da inserção de entidades inteligentes à IoT, com capacidade de mobilidade no âmbito espaço-temporal, é a utilização de robôs móveis autônomos, terrestres ou aéreos (veículos aéreos não tripulados, UAV – *Unmanned Aerial Vehicles*) para auxiliar nas tarefas cotidianas dos humanos [Chen and Hu 2012, Grieco et al. 2014].

Por meio da cooperação entre os humanos, as entidades de IoT, os robôs móveis e o ambiente, será possível a criação de aplicações nas mais diversas áreas, como saúde, agricultura, militares e de resgate em locais de risco. Neste último caso, podemos destacar a utilização de UAVs para o reconhecimento e detecção de sobreviventes em ambientes de desastres naturais. Para isto, será necessária a constante coleta e processamento de dados sobre o ambiente a fim de otimizar a coordenação dos robôs como, por exemplo, o reconhecimento e identificação de vítimas, bem como a definição de prioridades e melhores rotas [Grieco et al. 2014].

Por se tratarem de entidades potencialmente dotadas de maior capacidade computacional, os robôs móveis podem operar de forma autônoma e coordenada, tanto a partir do processamento remoto e centralizado dos dados coletados quanto pelo processamento local e comunicação entre os demais robôs móveis [Stojmenovic 2014]. Esta capacidade integra um subconjunto da IoT também denominado de *Internet of Intelligent Things* (IoIT), onde as suas entidades são autônomas, móveis, capazes de sensoriar e de atuar, indo além de simples sensores coletando dados.

Dispositivos vestíveis. Do outro lado do espectro de aplicações da IoMT encontra-se a miniaturização dos dispositivos computacionais, de forma que eles possam ser "vestidos" por seus usuários. Dispositivos como relógios inteligentes, pulseiras, ou até *smartphones*, capazes de realizar o monitoramento tanto das condições físicas de seus usuários quanto do ambiente ao seu redor, são uma fonte de dados fundamentais para aplicações que vão desde cuidados médicos até comportamento social de seus indivíduos [Silva et al. 2015d].

Uma das principais características desses dispositivos é que sua mobilidade, ao contrário do caso dos robôs móveis, não é autônoma. A movimentação desses dispositivos é dependente da movimentação de seus usuários. Desta forma, vários estudos realizados sobre a mobilidade e comportamento dos usuários podem ser aplicados [Nunes et al. 2016b], de forma que modelos de mobilidade e comunicação possam auxiliar no aperfeiçoamento das aplicações neste contexto. Neste sentido, há a chamada IoT

oportunística [Guo et al. 2013], que consiste, por exemplo, na disseminação e compartilhamento de informações de forma oportunística entre comunidades, que são formadas de acordo com a mobilidade natural dos indivíduos.

1.5.3.2. Desafios e oportunidades de *Big data analytics*

Os principais desafios encontrados pelas aplicações da IoMT são uma composição dos desafios originalmente encontrados pelas aplicações de IoT e redes móveis. Essa interseção de paradigmas impacta de diferentes formas a concepção da IoMT. A seguir destacamos alguns destes desafios e oportunidades de aplicação de estratégias de *big data analytics* em sua solução.

Heterogeneidade. A grande variedade de dispositivos, tecnologias, aplicações e seus domínios, bem como a diversidade de ambientes nos quais as entidades podem estar inseridas são aspectos fundamentais que precisam ser considerados no projeto de IoMT [Sowe et al. 2014]. Por exemplo, aplicações que precisam integrar dados coletados por sensores pode ter que lidar com diferentes padrões e implementações. Isso pode gerar dados de diferentes tipos, escalas e granularidades, o que requer um maior esforço na etapa de pré-processamento [Wu et al. 2014b], de forma a abstrair tais diferenças e sincronizar os dados de forma confiável.

Escalabilidade. Da mesma forma como ocorre com a IoT, desenvolvedores de aplicações para a IoMT devem estar preparados para lidar com a impressionante quantidade de dispositivos e dados gerados pela IoMT [Ma et al. 2013]. Por exemplo, simples algoritmos de agregação de dados podem estar sujeitos a problemas de escalabilidade se não considerarem que estes números podem crescer de forma mais acentuada que um crescimento linear. Estratégias que precisem armazenar os dados gerados para mineração e processamento histórico, assim como aquelas que efetuem processamento *online* e em tempo real de *streamings* de dados, precisarão ser capazes de garantir algum nível de escalabilidade [Tsai et al. 2014].

Restrições computacionais. Apesar de não ser uma regra para todas as aplicações de IoT, e agora IoMT, a probabilidade de se deparar com dispositivos com limitadas capacidades computacionais não pode ser descartada. Tanto para sensores quanto atuadores, uma característica herdada das redes de sensores sem fio [Loureiro et al. 2003], são as limitações de processamento, memória, comunicação e energia. Isso ocorre pois, ao contrário dos cenários mais comumente encontrados no mundo informacional, os dispositivos que interagem com o mundo físico geralmente são pequenos e, consequentemente, com limitados recursos.

Conectividade. A possibilidade de mobilidade dos dispositivos, ao contrário da maioria das aplicações em IoT, traz à IoMT um novo aspecto de dinamicidade necessário de ser considerado. Muitas aplicações podem se beneficiar deste novo aspecto de mobilidade,

com o potencial de dados gerados em larga escala, tanto em quantidade quanto cobertura [Musolesi 2014]. Mas, devido à movimentação dos dispositivos, que geralmente são dotados de comunicação sem fio de curto alcance, sua conectividade está diretamente relacionada com a manutenção de um raio de distância suficiente para a comunicação com outros dispositivos. Assim, a conectividade desses dispositivos passa a ser intermitente, fazendo com que dados possam ser entregues com variações de atraso ou até mesmo perdidos. O estudo destes padrões de mobilidade e a sua aplicação na predição e otimização da utilização dos dados [Nunes et al. 2016b] pode ser de grande valia para a IoMT.

Disponibilidade. Como decorrência da intermitente conectividade dos dispositivos da IoMT, se um determinado sensor está sendo utilizado como fonte de dados para uma aplicação, não é garantido que este sensor esteja disponível durante toda a sua operação. Isto pode afetar diretamente a qualidade das aplicações baseadas nos dados da IoMT [Borges Neto et al. 2015]. Isto pode desencadear a constante necessidade de execução de processos de descoberta de novos sensores, agregando um custo maior de tempo e recursos para as suas aplicações [Perera et al. 2014].

Localidade espaço-temporal. Por se tratar de dispositivos móveis gerando dados, tão importante quanto o dado gerado são as informações de tempo e localização, que informe quando e onde o dado foi gerado. Contudo, por se tratarem de dispositivos pequenos com restrições computacionais, nem sempre estarão dotados de *hardware* específicos para detectar sua localização (e.g., GPS) [Loureiro et al. 2003]. Assim, a informação de localidade pode ter que ser obtida de forma indireta, seja pelas interações desses dispositivos com outras entidades, como pontos de acesso ou outros dispositivos que possuam tal capacidade. Assim, registros de proximidade que podem ser utilizados na inferência da localização destes dispositivos, enquanto se movimentam e interagem com outros dispositivos, pode se aplicar [Cunha et al. 2016a].

1.6. Considerações Finais

Este capítulo apresentou os principais conceitos, técnicas e aplicações relacionadas ao projeto de redes móveis, infraestruturadas ou *ad hoc*, dirigido a dados. Para tanto, foi proposto e discutido um arcabouço que engloba diferentes etapas do processamento do dado, desde a sua obtenção até a sua aplicação no projeto de redes. Esse arcabouço foi definido com base em diversos exemplos de propostas discutidas na literatura e tem como objetivo definir um fluxo que permite aplicar os conhecimentos extraídos dos dados no projeto de redes móveis. Dessa forma, este capítulo fornece tanto os primeiros direcionamentos para introduzir novos interessados na área, assim como inspirá-los em pesquisas futuras.

Nesse sentido, diversas oportunidades e desafios de pesquisa foram discutidos ao longo do texto visto que essa área tem recebido bastante atenção nas principais conferências. Por exemplo, a conferência SIGCOMM possui um *workshop* chamado BIG-DAMA[15] que trata exclusivamente do tópico sobre *big data analytics* e aprendizagem de máquina para redes de comunicação. A conferência PERCOM possui um *workshop* chamado DAMN![16]

[15]Big-DAMA Workshop: http://conferences.sigcomm.org/sigcomm/2017/workshop-big-dama.html
[16]DAMN Workshop: http://damn2017.conf.citi-lab.fr/

que trata puramente de tópicos relacionados a *data analytics* para redes móveis. Além desses *workshops*, as trilhas principais nessas e outras conferências da área têm apresentado seções relevantes no âmbito do tópico deste capítulo.

Para novas pesquisas, é importante termos novas fontes de dados que estejam disponíveis publicamente, para podermos aplicar algoritmos de aprendizagem de máquina e mineração de dados considerando as peculiaridades das redes móveis. Finalmente, questões de segurança dos dados e aspectos práticos que considerem a importância da análise de dados na concepção de redes móveis também devem ser tratadas.

Referências

[Aggarwal and Han 2014] Aggarwal, C. C. and Han, J. (2014). *Frequent pattern mining*. Springer.

[Akyildiz et al. 2016] Akyildiz, I. F., Nie, S., Lin, S.-C., and Chandrasekaran, M. (2016). 5g roadmap: 10 key enabling technologies. *Computer Networks*, 106:17–48.

[Amici et al. 2014] Amici, R., Bonola, M., Bracciale, L., Rabuffi, A., Loreti, P., and Bianchi, G. (2014). Performance assessment of an epidemic protocol in vanet using real traces. *Procedia Computer Science*, 40:92–99.

[Asgari et al. 2016] Asgari, F., Sultan, A., Xiong, H., Gauthier, V., and El-Yacoubi, M. A. (2016). Ct-mapper: mapping sparse multimodal cellular trajectories using a multilayer transportation network. *Computer Communications*, 95:69–81.

[Atzori et al. 2010] Atzori, L., Iera, A., and Morabito, G. (2010). The Internet of Things: A survey. *Computer Networks*, 54(15):2787–2805.

[Barbera et al. 2013] Barbera, M. V., Epasto, A., Mei, A., Perta, V. C., and Stefa, J. (2013). Signals from the crowd: uncovering social relationships through smartphone probes. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 265–276. ACM.

[Barchiesi et al. 2015] Barchiesi, D., Preis, T., Bishop, S., and Moat, H. S. (2015). Modelling human mobility patterns using photographic data shared online. *Royal Society open science*, 2(8):150046.

[Barthélemy 2011] Barthélemy, M. (2011). Spatial networks. *Physics Reports*, 499(1):1–101.

[Benevenuto et al. 2009] Benevenuto, F., Rodrigues, T., Cha, M., and Almeida, V. (2009). Characterizing user behavior in online social networks. In *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, pages 49–62. ACM.

[Bi et al. 2015] Bi, S., Zhang, R., Ding, Z., and Cui, S. (2015). Wireless communications in the era of big data. *IEEE Communications Magazine*, 53(10):190–199.

[Blondel et al. 2015] Blondel, V. D., Decuyper, A., and Krings, G. (2015). A survey of results on mobile phone datasets analysis. *EPJ Data Science*, 4(1):10.

[Blondel et al. 2012] Blondel, V. D., Esch, M., Chan, C., Clérot, F., Deville, P., Huens, E., Morlot, F., Smoreda, Z., and Ziemlicki, C. (2012). Data for development: the d4d challenge on mobile phone data. *arXiv preprint arXiv:1210.0137*.

[Borges Neto et al. 2015] Borges Neto, J., Silva, T., Assunção, R., Mini, R., and Loureiro, A. (2015). Sensing in the Collaborative Internet of Things. *Sensors*, 15(3):6607–6632.

[Bracciale et al. 2014] Bracciale, L., Bonola, M., Loreti, P., Bianchi, G., Amici, R., and Rabuffi, A. (2014). CRAWDAD data set roma/taxi (v. 2014-07-17). Downloaded from http://crawdad.org/roma/taxi/.

[Brockmann et al. 2006] Brockmann, D., Hufnagel, L., and Geisel, T. (2006). The scaling laws of human travel. *Nature*, 439(7075):462–465.

[Calabrese et al. 2011] Calabrese, F., Di Lorenzo, G., Liu, L., and Ratti, C. (2011). Estimating origin-destination flows using mobile phone location data. *IEEE Pervasive Computing*, 10(4):36–44.

[Calabrese et al. 2013] Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira, J., and Ratti, C. (2013). Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation research part C: emerging technologies*, 26:301–313.

[Cattell 2011] Cattell, R. (2011). Scalable sql and nosql data stores. *SIGMOD Rec.*, 39(4):12–27.

[Celes et al. 2013] Celes, C., Braga, R. B., Oliveira, C. T. D., Andrade, R. M. C., and Loureiro, A. A. F. (2013). Geospin: An approach for geocast routing based on spatial information in vanets. In *2013 IEEE 78th Vehicular Technology Conference (VTC Fall)*, pages 1–6.

[Centellegher et al. 2016] Centellegher, S., De Nadai, M., Caraviello, M., Leonardi, C., Vescovi, M., Ramadian, Y., Oliver, N., Pianesi, F., Pentland, A., Antonelli, F., et al. (2016). The mobile territorial lab: a multilayered and dynamic view on parents' daily lives. *EPJ Data Science*, 5(1):3.

[Chandola et al. 2009] Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):15.

[Checko et al. 2015] Checko, A., Christiansen, H. L., Yan, Y., Scolari, L., Kardaras, G., Berger, M. S., and Dittmann, L. (2015). Cloud ran for mobile networks—a technology overview. *IEEE Communications surveys & tutorials*, 17(1):405–426.

[Chen et al. 2014a] Chen, M., Mao, S., and Liu, Y. (2014a). Big data: A survey. *Mobile Networks and Applications*, 19(2):171–209.

[Chen and Hu 2012] Chen, Y. and Hu, H. (2012). Internet of intelligent things and robot as a service. *Simulation Modelling Practice and Theory*, 34:159–171.

[Chen et al. 2013] Chen, Y., Xu, M., Gu, Y., Li, P., and Cheng, X. (2013). Understanding topology evolving of vanets from taxi traces. *Adv. Sci. Technol. Lett*, 42(Mobile and Wireless):13–17.

[Chen et al. 2014b] Chen, Y., Xu, M., Gu, Y., Li, P., Shi, L., and Xiao, X. (2014b). Empirical study on spatial and temporal features for vehicular wireless communications. *EURASIP Journal on Wireless Communications and Networking*, 2014(1):1–12.

[Cheng et al. 2017] Cheng, X., Fang, L., Hong, X., and Yang, L. (2017). Exploiting mobile big data: Sources, features, and applications. *IEEE Network*, 31(1):72–79.

[Chu et al. 2016] Chu, X., Ilyas, I. F., Krishnan, S., and Wang, J. (2016). Data cleaning: Overview and emerging challenges. In *Proceedings of the 2016 International Conference on Management of Data*, pages 2201–2206. ACM.

[Çolak et al. 2016] Çolak, S., Lima, A., and González, M. C. (2016). Understanding congested travel in urban areas. *Nature communications*, 7.

[Cormode and Srivastava 2009] Cormode, G. and Srivastava, D. (2009). Anonymized data: generation, models, usage. In *Proceedings of the 2009 ACM SIGMOD International Conference on Management of data*, pages 1015–1018. ACM.

[Cornejo et al. 2013] Cornejo, A., Newport, C., Gollakota, S., Rao, J., and Giuli, T. J. (2013). Prioritized gossip in vehicular networks. *Ad Hoc Networks*, 11(1):397–409.

[Coronado and Iglesias 2016] Coronado, M. and Iglesias, C. A. (2016). Task automation services: Automation for the masses. *IEEE Internet Computing*, 20(1):52–58.

[Cunha et al. 2016a] Cunha, F. D., Alvarenga, D. A., Maia, G., Viana, A. C., Mini, R. A., and Loureiro, A. A. (2016a). Exploring interactions in vehicular networks. In *Proceedings of the 14th ACM International Symposium on Mobility Management and Wireless Access*, pages 131–138. ACM.

[Cunha et al. 2016b] Cunha, F. D., Silva, F. A., Celes, C., Maia, G., Ruiz, L. B., Andrade, R. M., Mini, R. A., Boukerche, A., and Loureiro, A. A. (2016b). Communication analysis of real vehicular calibrated traces. In *Communications (ICC), 2016 IEEE International Conference on*, pages 1–6. IEEE.

[de Melo et al. 2015] de Melo, P. O. V., Viana, A. C., Fiore, M., Jaffrès-Runser, K., Le Mouël, F., Loureiro, A. A., Addepalli, L., and Guangshuo, C. (2015). Recast: Telling apart social and random relationships in dynamic networks. *Performance Evaluation*, 87:19–36.

[de Montjoye et al. 2014] de Montjoye, Y.-A., Smoreda, Z., Trinquart, R., Ziemlicki, C., and Blondel, V. D. (2014). D4d-senegal: the second mobile phone data for development challenge. *arXiv preprint arXiv:1407.4885*.

[Dean and Ghemawat 2008] Dean, J. and Ghemawat, S. (2008). Mapreduce: simplified data processing on large clusters. *Communications of the ACM*, 51(1):107–113.

[Demissie et al. 2013] Demissie, M. G., de Almeida Correia, G. H., and Bento, C. (2013). Exploring cellular network handover information for urban mobility analysis. *Journal of Transport Geography*, 31:164–170.

[Eagle and Pentland 2006] Eagle, N. and Pentland, A. (2006). Reality mining: sensing complex social systems. *Personal and ubiquitous computing*, 10(4):255–268.

[Eagle and Pentland 2005] Eagle, N. and Pentland, A. S. (2005). CRAWDAD dataset mit/reality (v. 2005-07-01). Downloaded from http://crawdad.org/mit/reality/20050701.

[Ekman et al. 2008] Ekman, F., Keränen, A., Karvo, J., and Ott, J. (2008). Working day movement model. In *Proceedings of the 1st ACM SIGMOBILE workshop on Mobility models*, pages 33–40. ACM.

[Ficek and Kencl 2010] Ficek, M. and Kencl, L. (2010). Spatial extension of the reality mining dataset. In *Mobile Adhoc and Sensor Systems (MASS), 2010 IEEE 7th International Conference on*, pages 666–673. IEEE.

[Freedman et al. 2007] Freedman, D., Pisani, R., and Purves, R. (2007). *Statistics*. W. W. Norton and Co, 4th edition.

[Gao et al. 2009] Gao, W., Li, Q., Zhao, B., and Cao, G. (2009). Multicasting in delay tolerant networks: a social network perspective. In *Proceedings of the tenth ACM international symposium on Mobile ad hoc networking and computing*, pages 299–308. ACM.

[Garton et al. 1997] Garton, L., Haythornthwaite, C., and Wellman, B. (1997). Studying online social networks. *Journal of Computer-Mediated Communication*, 3(1):0–0.

[Gonzalez et al. 2008] Gonzalez, M. C., Hidalgo, C. A., and Barabasi, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, 453(7196):779–782.

[Gossa et al. 2008] Gossa, J., Janecek, A. G., Hummel, K. A., Gansterer, W. N., and Pierson, J.-M. (2008). Proactive replica placement using mobility prediction. In *Mobile Data Management Workshops, 2008. MDMW 2008. Ninth International Conference on*, pages 182–189. IEEE.

[Gramaglia and Fiore 2014] Gramaglia, M. and Fiore, M. (2014). On the anonymizability of mobile traffic datasets. *arXiv preprint arXiv:1501.00100*.

[Greengard 2014] Greengard, S. (2014). Weathering a New Era of Big Data. *Association for Computing Machinery. Communications of the ACM*, 57(9):12.

[Gregory 2010] Gregory, S. (2010). Finding overlapping communities in networks by label propagation. *New Journal of Physics*, 12(10):103018.

[Grieco et al. 2014] Grieco, L., Rizzo, A., Colucci, S., Sicari, S., Piro, G., Di Paola, D., and Boggia, G. (2014). IoT-aided robotics applications: Technological implications, target domains and open issues. *Computer Communications*, 54:32–47.

[Gubbi et al. 2013] Gubbi, J., Buyya, R., Marusic, S., and Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7):1645–1660.

[Guidotti et al. 2014] Guidotti, R., Monreale, A., Rinzivillo, S., Pedreschi, D., and Giannotti, F. (2014). Retrieving points of interest from human systematic movements. In *International Conference on Software Engineering and Formal Methods*, pages 294–308. Springer.

[Guo et al. 2013] Guo, B., Zhang, D., Wang, Z., Yu, Z., and Zhou, X. (2013). Opportunistic IoT: Exploring the harmonious interaction between human and the internet of things. *Journal of Network and Computer Applications*, 36(6):1531–1539.

[Han et al. 2011] Han, J., Kamber, M., and Pei, J. (2011). *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 3rd edition.

[Hand et al. 2001] Hand, D. J., Mannila, H., and Smyth, P. (2001). *Principles of data mining*. MIT press.

[He et al. 2016] He, Y., Yu, F. R., Zhao, N., Yin, H., Yao, H., and Qiu, R. C. (2016). Big data analytics in mobile cellular networks. *IEEE Access*, 4:1985–1996.

[Henderson et al. 2008] Henderson, T., Kotz, D., and Abyzov, I. (2008). The changing usage of a mature campus-wide wireless network. *Computer Networks*, 52(14):2690–2712.

[Hess et al. 2016] Hess, A., Hummel, K. A., Gansterer, W. N., and Haring, G. (2016). Data-driven human mobility modeling: a survey and engineering guidance for mobile networking. *ACM Computing Surveys (CSUR)*, 48(3):38.

[Hidalgo and Rodriguez-Sickert 2008] Hidalgo, C. A. and Rodriguez-Sickert, C. (2008). The dynamics of a mobile phone network. *Physica A: Statistical Mechanics and its Applications*, 387(12):3017–3024.

[Holme and Saramäki 2012] Holme, P. and Saramäki, J. (2012). Temporal networks. *Physics reports*, 519(3):97–125.

[Hoque et al. 2014] Hoque, M. A., Hong, X., and Dixon, B. (2014). Efficient multihop connectivity analysis in urban vehicular networks. *Vehicular Communications*, 1(2):78–90.

[Hou et al. 2016] Hou, X., Li, Y., Jin, D., Wu, D. O., and Chen, S. (2016). Modeling the impact of mobility on the connectivity of vehicular networks in large-scale urban environments. *IEEE Transactions on Vehicular Technology*, 65(4):2753–2758.

[Hsu and Helmy 2005] Hsu, W.-j. and Helmy, A. (2005). Impact: Investigation of mobile-user patterns across university campuses using wlan trace analysis. *arXiv preprint cs/0508009*.

[Hui et al. 2011] Hui, P., Crowcroft, J., and Yoneki, E. (2011). Bubble rap: Social-based forwarding in delay-tolerant networks. *IEEE Transactions on Mobile Computing*, 10(11):1576–1589.

[Imran et al. 2014] Imran, A., Zoha, A., and Abu-Dayya, A. (2014). Challenges in 5g: how to empower son with big data for enabling 5g. *IEEE Network*, 28(6):27–33.

[Jain 2010] Jain, A. K. (2010). Data clustering: 50 years beyond k-means. *Pattern Recogn. Lett.*, 31(8):651–666.

[James et al. 2014] James, G., Witten, D., Hastie, T., and Tibshirani, R. (2014). *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company, Incorporated.

[Jiang et al. 2016] Jiang, S., Ferreira, J., and Gonzales, M. C. (2016). Activity-based human mobility patterns inferred from mobile phone data: A case study of singapore. *IEEE Transactions on Big Data*.

[Kamienski et al. 2016] Kamienski, C., Biondi, G. O., Borelli, F. F., Heideker, A., Ratusznei, J., and Kleinschmidt, J. H. (2016). Computação urbana: Tecnologias e aplicações para cidades inteligentes. *Minicursos SBRC-Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*.

[Karkouch et al. 2016] Karkouch, A., Mousannif, H., Al Moatassime, H., and Noel, T. (2016). Data quality in internet of things: A state-of-the-art survey. *Journal of Network and Computer Applications*, 73:57–81.

[Kemeny et al. 1960] Kemeny, J. G., Snell, J. L., et al. (1960). *Finite markov chains*, volume 356. van Nostrand Princeton, NJ.

[Kivelä et al. 2014] Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., and Porter, M. A. (2014). Multilayer networks. *Journal of complex networks*, 2(3):203–271.

[Lambiotte et al. 2008] Lambiotte, R., Blondel, V. D., De Kerchove, C., Huens, E., Prieur, C., Smoreda, Z., and Van Dooren, P. (2008). Geographical dispersal of mobile communication networks. *Physica A: Statistical Mechanics and its Applications*, 387(21):5317–5325.

[Laya et al. 2014] Laya, A., Wang, K., Widaa, A. A., Alonso-Zarate, J., Markendahl, J., and Alonso, L. (2014). Device-to-device communications and small cells: enabling spectrum reuse for dense networks. *IEEE Wireless Communications*, 21(4):98–105.

[Lee et al. 2009] Lee, K., Hong, S., Kim, S. J., Rhee, I., and Chong, S. (2009). Slaw: A new mobility model for human walks. In *INFOCOM 2009, IEEE*, pages 855–863. IEEE.

[Leguay and Benbadis 2009] Leguay, J. and Benbadis, F. (2009). Crawdad data set upmc/rollernet (v. 2009-02-02).

[Leguay et al. 2006] Leguay, J., Lindgren, A., Scott, J., Friedman, T., and Crowcroft, J. (2006). Opportunistic content distribution in an urban setting. In *Proceedings of the 2006 SIGCOMM workshop on Challenged networks*, pages 205–212. ACM.

[Leskovec et al. 2009] Leskovec, J., Lang, K. J., Dasgupta, A., and Mahoney, M. W. (2009). Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics*, 6(1):29–123.

[Li et al. 2014] Li, Y., Wu, T., Hui, P., Jin, D., and Chen, S. (2014). Social-aware d2d communications: qualitative insights and quantitative analysis. *IEEE Communications Magazine*, 52(6):150–158.

[Liu et al. 2011] Liu, N., Liu, M., Lou, W., Chen, G., and Cao, J. (2011). Pva in vanets: Stopped cars are not silent. In *INFOCOM, 2011 Proceedings IEEE*, pages 431–435. IEEE.

[Loureiro et al. 2003] Loureiro, A. A. F., Nogueira, J. M. S., Ruiz, L. B., Mini, R. A. d. F., Nakamura, E. F., and Figueiredo, C. M. S. (2003). Redes de Sensores Sem Fio. *Simpósio Brasileiro de Redes de Computadores (SBRC)*, pages 179–226.

[Lu et al. 2016] Lu, X., Yu, Z., Sun, L., Liu, C., Xiong, H., and Guan, C. (2016). Characterizing the life cycle of point of interests using human mobility patterns. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 1052–1063. ACM.

[Ma et al. 2013] Ma, M., Wang, P., and Chu, C.-H. (2013). Data Management for Internet of Things: Challenges, Approaches and Opportunities. In *2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing*, pages 1144–1151. IEEE.

[Machado et al. 2015] Machado, K., Silva, T. H., de Melo, P. O. V., Cerqueira, E., and Loureiro, A. A. (2015). Urban mobility sensing analysis through a layered sensing approach. In *Mobile Services (MS), 2015 IEEE International Conference on*, pages 306–312. IEEE.

[Mamei et al. 2016] Mamei, M., Colonna, M., and Galassi, M. (2016). Automatic identification of relevant places from cellular network data. *Pervasive and Mobile Computing*, 31:147–158.

[Mei and Stefa 2009] Mei, A. and Stefa, J. (2009). Swim: A simple model to generate small mobile worlds. In *INFOCOM 2009, IEEE*, pages 2106–2113. IEEE.

[Michel and Julien 2016] Michel, J. and Julien, C. (2016). From human mobility to data mobility: Leveraging spatiotemporal history in device-to-device information diffusion. In *Mobile Data Management (MDM), 2016 17th IEEE International Conference on*, volume 1, pages 198–207. IEEE.

[Monteiro et al. 2012] Monteiro, R., Sargento, S., Viriyasitavat, W., and Tonguz, O. K. (2012). Improving vanet protocols via network science. In *Vehicular Networking Conference (VNC), 2012 IEEE*, pages 17–24. IEEE.

[Mota et al. 2014] Mota, V. F., Cunha, F. D., Macedo, D. F., Nogueira, J. M., and Loureiro, A. A. (2014). Protocols, mobility models and tools in opportunistic networks: A survey. *Computer Communications*, 48:5–19.

[Musolesi 2014] Musolesi, M. (2014). Big Mobile Data Mining: Good or Evil? *IEEE Internet Computing*, 18(1):78–81.

[Naboulsi and Fiore 2016] Naboulsi, D. and Fiore, M. (2016). Characterizing the instantaneous connectivity of large-scale urban vehicular networks. *IEEE Transactions on Mobile Computing*.

[Naboulsi et al. 2016] Naboulsi, D., Fiore, M., Ribot, S., and Stanica, R. (2016). Large-scale mobile traffic analysis: a survey. *IEEE Communications Surveys & Tutorials*, 18(1):124–161.

[Nahrstedt et al. 2016] Nahrstedt, K., Li, H., Nguyen, P., Chang, S., and Vu, L. (2016). Internet of Mobile Things: Mobility-Driven Challenges, Designs and Implementations. In *2016 IEEE First International Conference on Internet-of-Things Design and Implementation (IoTDI)*, pages 25–36. IEEE.

[Nguyen et al. 2011] Nguyen, N. P., Dinh, T. N., Tokala, S., and Thai, M. T. (2011). Overlapping communities in dynamic networks: their detection and mobile applications. In *Proceedings of the 17th annual international conference on Mobile computing and networking*, pages 85–96. ACM.

[Nika et al. 2016] Nika, A., Ismail, A., Zhao, B. Y., Gaito, S., Rossi, G. P., and Zheng, H. (2016). Understanding and predicting data hotspots in cellular networks. *Mobile Networks and Applications*, 21(3):402–413.

[Nunes et al. 2016a] Nunes, I. O., Celes, C., Vaz de Melo, P. O., and Loureiro, A. A. (2016a). Groups-net: Group meetings aware routing in multi-hop d2d networks. *arXiv preprint arXiv:1605.07692*.

[Nunes et al. 2016b] Nunes, I. O., de Melo, P. O. S. V., and Loureiro, A. A. F. (2016b). Group mobility: Detection, tracking and characterization. In *2016 IEEE International Conference on Communications (ICC)*, pages 1–6.

[Nunes et al. 2016c] Nunes, I. O., de Melo, P. O. V., and Loureiro, A. A. (2016c). Leveraging d2d multihop communication through social group meeting awareness. *IEEE Wireless Communications*, 23(4):12–19.

[Nunes et al. 2016d] Nunes, I. O., de Melo, P. O. V., and Loureiro, A. A. F. (2016d). Groups-net: Roteamento ciente de encontros de grupos em redes móveis d2d. In *Proceedings of the Brazilian Symposium on Computer Networks and Distributed Systems*, Salvador, Bahia.

[Palla et al. 2005] Palla, G., Derényi, I., Farkas, I., and Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818.

[Pappalardo et al. 2015] Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., and Barabási, A.-L. (2015). Returners and explorers dichotomy in human mobility. *Nature communications*, 6.

[Perera et al. 2014] Perera, C., Zaslavsky, A., Liu, C. H., Compton, M., Christen, P., and Georgakopoulos, D. (2014). Sensor Search Techniques for Sensing as a Service Architecture for the Internet of Things. *IEEE Sensors Journal*, 14(2):406–420.

[Pietilainen and Diot 2012] Pietilainen, A.-K. and Diot, C. (2012). CRAWDAD dataset thlab/sigcomm2009 (v. 2012-07-15). Downloaded from http://crawdad.org/thlab/sigcomm2009/20120715.

[Piorkowski et al. 2009a] Piorkowski, M., Sarafijanovic-Djukic, N., and Grossglauser, M. (2009a). CRAWDAD data set epfl/mobility (v. 2009-02-24). Downloaded from http://crawdad.org/epfl/mobility/.

[Piorkowski et al. 2009b] Piorkowski, M., Sarafijanovoc-Djukic, N., and Grossglauser, M. (2009b). A Parsimonious Model of Mobile Partitioned Networks with Clustering. In *The First International Conference on COMmunication Systems and NETworkS (COMSNETS)*.

[Pires et al. 2015] Pires, P. F., Delicato, F., Batista, T., Barros, T., Cavalcante, E., and Pitanga, M. (2015). Plataformas para a internet das coisas. *Minicursos SBRC-Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*.

[Rebecchi et al. 2015] Rebecchi, F., De Amorim, M. D., Conan, V., Passarella, A., Bruno, R., and Conti, M. (2015). Data offloading techniques in cellular networks: a survey. *IEEE Communications Surveys & Tutorials*, 17(2):580–603.

[Rhee et al. 2011] Rhee, I., Shin, M., Hong, S., Lee, K., Kim, S. J., and Chong, S. (2011). On the levy-walk nature of human mobility. *IEEE/ACM transactions on networking (TON)*, 19(3):630–643.

[Santos et al. 2016] Santos, B. P., Silva, L. A., Celes, C. S., Borges, J. B., Neto, B. S. P., Vieira, M. A. M., Vieira, L. F. M., Goussevskaia, O. N., and Loureiro, A. A. (2016). Internet das coisas: da teoria à prática. *Minicursos SBRC-Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*.

[Scott et al. 2009] Scott, J., Gass, R., Crowcroft, J., Hui, P., Diot, C., and Chaintreau, A. (2009). CRAWDAD dataset cambridge/haggle (v. 2009-05-29). Downloaded from http://crawdad.org/cambridge/haggle/20090529.

[Shafiq et al. 2015] Shafiq, M. Z., Ji, L., Liu, A. X., Pang, J., and Wang, J. (2015). Geospatial and temporal dynamics of application usage in cellular data networks. *IEEE Transactions on Mobile Computing*, 14(7):1369–1381.

[Shumway and Stoffer 2010] Shumway, R. H. and Stoffer, D. S. (2010). *Time series analysis and its applications: with R examples*. Springer Science & Business Media.

[Silva et al. 2015a] Silva, C. M., Aquino, A. L., and Meira, W. (2015a). Deployment of roadside units based on partial mobility information. *Computer Communications*, 60:28–39.

[Silva et al. 2016a] Silva, C. M., Silva, F. A., Sarubbi, J. F., Oliveira, T. R., Meira, W., and Nogueira, J. M. S. (2016a). Designing mobile content delivery networks for the internet of vehicles. *Vehicular Communications*.

[Silva et al. 2016b] Silva, F. A., Boukerche, A., Silva, T. R., Ruiz, L. B., Cerqueira, E., and Loureiro, A. A. (2016b). Vehicular networks: A new challenge for content-delivery-based applications. *ACM Computing Surveys (CSUR)*, 49(1):11.

[Silva et al. 2015b] Silva, F. A., Boukerche, A., Silva, T. R., Ruiz, L. B., and Loureiro, A. A. (2015b). A novel macroscopic mobility model for vehicular networks. *Computer Networks*, 79:188–202.

[Silva et al. 2015c] Silva, F. A., Boukerche, A., Silva, T. R. B., Benevenuto, F., Ruiz, L. B., and Loureiro, A. A. (2015c). Odcrep: Origin–destination-based content replication for vehicular networks. *IEEE Transactions on Vehicular Technology*, 64(12):5563–5574.

[Silva et al. 2016c] Silva, F. A., Boukerche, A., Silva, T. R. B., Ruiz, L. B., and Loureiro, A. A. (2016c). Geo-localized content availability in vanets. *Ad Hoc Networks*, 36:425–434.

[Silva et al. 2017a] Silva, M., de Oliveira Nunes, I., Loureiro, A. A. F., and Mini, R. (2017a). St-drop: Uma nova estratégia de gerenciamento de buffer em redes d2d oportunistas. In *SBRC 2017*.

[Silva et al. 2017b] Silva, M., Nunes, I., Mini, R. A. F., and A.F. Loureiro, A. (2017b). ST-Drop: a novel buffer management strategy for D2D opportunistic networks. In *22nd IEEE Symposium on Computers and Communication (ISCC 2017) (ISCC 2017)*, Heraklion, Greece.

[Silva et al. 2014a] Silva, T. H., De Melo, P. O. V., Almeida, J. M., and Loureiro, A. A. (2014a). Large-scale study of city dynamics and urban social behavior using participatory sensing. *IEEE Wireless Communications*, 21(1):42–51.

[Silva et al. 2015d] Silva, T. H., de Melo, P. O. V., Neto, J. B., IJT, A., Ribeiro, C. S. d. S., Mota, V. F., da Cunha, F. D., Ferreira, A. P., Machado, K. L. d. S., Mini, R. A. d. F., et al. (2015d). Redes de sensoriamento participativo: Desafios e oportunidades. *Minicursos SBRC-Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*.

[Silva et al. 2014b] Silva, T. H., Vaz de Melo, P. O., Almeida, J. M., Salles, J., and Loureiro, A. A. (2014b). Revealing the city that we cannot see. *ACM Transactions on Internet Technology (TOIT)*, 14(4):26.

[Silveira et al. 2016] Silveira, L. M., de Almeida, J. M., Marques-Neto, H. T., Sarraute, C., and Ziviani, A. (2016). Mobhet: Predicting human mobility using heterogeneous data sources. *Computer Communications*.

[Socievole et al. 2014] Socievole, A., De Rango, F., and Caputo, A. (2014). Wireless contacts, facebook friendships and interests: Analysis of a multi-layer social network in an academic environment. In *Wireless Days (WD), 2014 IFIP*, pages 1–7. IEEE.

[Song et al. 2014] Song, L., Niyato, D., Han, Z., and Hossain, E. (2014). Game-theoretic resource allocation methods for device-to-device communication. *IEEE Wireless Communications*, 21(3):136–144.

[Soto et al. 2011] Soto, V., Frias-Martinez, V., Virseda, J., and Frias-Martinez, E. (2011). Prediction of socioeconomic levels using cell phone records. In *International Conference on User Modeling, Adaptation, and Personalization*, pages 377–388. Springer.

[Sowe et al. 2014] Sowe, S. K., Kimata, T., Dong, M., and Zettsu, K. (2014). Managing Heterogeneous Sensor Data on a Big Data Platform: IoT Services for Data-Intensive Science. In *2014 IEEE 38th International Computer Software and Applications Conference Workshops*, pages 295–300. IEEE.

[Stanica et al. 2013] Stanica, R., Fiore, M., and Malandrino, F. (2013). Offloading floating car data. In *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2013 IEEE 14th International Symposium and Workshops on a*, pages 1–9. IEEE.

[Stojmenovic 2014] Stojmenovic, I. (2014). Machine-to-Machine Communications With In-Network Data Aggregation, Processing, and Actuation for Large-Scale Cyber-Physical Systems. *IEEE Internet of Things Journal*, 1(2):122–128.

[SUVnet] SUVnet. Shanghai data trace. Online (available at http://wirelesslab.sjtu.edu.cn/taxi_trace_data.html).

[Tanahashi et al. 2012] Tanahashi, Y., Rowland, J. R., North, S., and Ma, K.-L. (2012). Inferring human mobility patterns from anonymized mobile communication usage. In *Proceedings of the 10th International Conference on Advances in Mobile Computing & Multimedia*, pages 151–160. ACM.

[Teles et al. 2013] Teles, A., Pinheiro, D., Gonçalves, J., Batista, R., Almeida, V., Endler, M., and Silva, F. (2013). Redes sociais móveis: conceitos, aplicações e aspectos de segurança e privacidade. *31º Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC)*.

[Tizzoni et al. 2014] Tizzoni, M., Bajardi, P., Decuyper, A., King, G. K. K., Schneider, C. M., Blondel, V., Smoreda, Z., González, M. C., and Colizza, V. (2014). On the use of human mobility proxies for modeling epidemics. *PLoS Comput Biol*, 10(7):e1003716.

[Tonguz et al. 2009] Tonguz, O. K., Viriyasitavat, W., and Bai, F. (2009). Modeling urban traffic: a cellular automata approach. *IEEE Communications Magazine*, 47(5).

[Tostes et al. 2013] Tostes, A. I. J., de LP Duarte-Figueiredo, F., Assunção, R., Salles, J., and Loureiro, A. A. (2013). From data to knowledge: City-wide traffic flows analysis and prediction using bing maps. In *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*, page 12. ACM.

[Trasarti et al. 2011] Trasarti, R., Pinelli, F., Nanni, M., and Giannotti, F. (2011). Mining mobility user profiles for car pooling. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1190–1198. ACM.

[Trestian et al. 2009] Trestian, I., Ranjan, S., Kuzmanovic, A., and Nucci, A. (2009). Measuring serendipity: connecting people, locations and interests in a mobile 3g network. In *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, pages 267–279. ACM.

[Trullols-Cruces et al. 2015] Trullols-Cruces, O., Fiore, M., and Barcelo-Ordinas, J. M. (2015). Worm epidemics in vehicular networks. *IEEE Transactions on Mobile Computing*, 14(10):2173–2187.

[Tsai et al. 2014] Tsai, C.-W., Lai, C.-F., Chiang, M.-C., and Yang, L. T. (2014). Data Mining for Internet of Things: A Survey. *IEEE Communications Surveys & Tutorials*, 16(1):77–97.

[Tsai and Chan 2015] Tsai, T.-C. and Chan, H.-H. (2015). Nccu trace: social-network-aware mobility trace. *Communications Magazine, IEEE*, 53(10):144–149.

[Tukey 1977] Tukey, J. W. (1977). *Exploratory Data Analysis*. Behavioral Science: Quantitative Methods. Addison-Wesley, Reading, Mass.

[Wang et al. 2016] Wang, J., Wu, Y., Yen, N., Guo, S., and Cheng, Z. (2016). Big data analytics for emergency communication networks: A survey. *IEEE Communications Surveys Tutorials*, 18(3):1758–1778.

[Wang et al. 2015] Wang, W., Yuan, N., Pan, L., Jiao, P., Dai, W., Xue, G., and Liu, D. (2015). Temporal patterns of emergency calls of a metropolitan city in china. *Physica A: Statistical Mechanics and its Applications*.

[Wang et al. 2014] Wang, Y., Huang, L., Gu, T., Wei, H., Xing, K., and Zhang, J. (2014). Data-driven traffic flow analysis for vehicular communications. In *INFOCOM, 2014 Proceedings IEEE*, pages 1977–1985. IEEE.

[Ward and Barker 2013] Ward, J. S. and Barker, A. (2013). Undefined by data: a survey of big data definitions. *arXiv preprint arXiv:1309.5821*.

[Williams and Musolesi 2016] Williams, M. J. and Musolesi, M. (2016). Spatio-temporal networks: reachability, centrality and robustness. *Royal Society Open Science*, 3(6):160196.

[Wu et al. 2014a] Wu, M., Wang, Y., and Liao, Z. (2014a). A new clustering algorithm for sensor data streams in an agricultural IoT. *Proceedings - 2013 IEEE International Conference on High Performance Computing and Communications, HPCC 2013 and 2013 IEEE International Conference on Embedded and Ubiquitous Computing, EUC 2013*, pages 2373–2378.

[Wu et al. 2014b] Wu, X., Zhu, X., Wu, G.-Q., and Ding, W. (2014b). Data mining with big data. *ieee transactions on knowledge and data engineering*, 26(1):97–107.

[Wu et al. 2011] Wu, Y., Zhu, Y., and Li, B. (2011). Trajectory improves data delivery in vehicular networks. In *INFOCOM, 2011 Proceedings IEEE*, pages 2183–2191. IEEE.

[Xiao et al. 2014] Xiao, X., Li, Y., and Kui, X. (2014). Location patterns and predictability of large scale urban vehicular mobility. In *Wireless Communications and Networking Conference (WCNC), 2014 IEEE*, pages 2705–2709. IEEE.

[Xu et al. 2016] Xu, F., Li, Y., Wang, H., Zhang, P., and Jin, D. (2016). Understanding mobile traffic patterns of large scale cellular towers in urban environment. *IEEE/ACM Transactions on Networking*, PP(99):1–15.

[Yang et al. 2014] Yang, F., Wang, S., Li, J., Liu, Z., and Sun, Q. (2014). An overview of Internet of Vehicles. *China Communications*, 11(10):1–15.

[Yin et al. 2013] Yin, R., Yu, G., Zhong, C., and Zhang, Z. (2013). Distributed resource allocation for d2d communication underlaying cellular networks. In *Communications Workshops (ICC), 2013 IEEE International Conference on*, pages 138–143. IEEE.

[Yu et al. 2016] Yu, S., Liu, M., Dou, W., Liu, X., and Zhou, S. (2016). Networking for big data: A survey. *IEEE Communications Surveys & Tutorials*.

[Zaharia et al. 2010] Zaharia, M., Chowdhury, M., Franklin, M. J., Shenker, S., and Stoica, I. (2010). Spark: Cluster computing with working sets. *HotCloud*, 10(10-10):95.

[Zanella et al. 2014] Zanella, A., Bui, N., Castellani, A., Vangelista, L., and Zorzi, M. (2014). Internet of things for smart cities. *IEEE Internet of Things Journal*, 1(1):22–32.

[Zhang et al. 2013] Zhang, D., Huang, H., Zhou, J., Xia, F., and Chen, Z. (2013). Detecting hot road mobility of vehicular ad hoc networks. *Mobile Networks and Applications*, 18(6):803–813.

[Zhang et al. 2014] Zhang, D., Huang, J., Li, Y., Zhang, F., Xu, C., and He, T. (2014). Exploring human mobility with multi-source data at extremely large metropolitan scales. In *Proceedings of the 20th annual international conference on Mobile computing and networking*, pages 201–212. ACM.

[Zhang et al. 2016] Zhang, F., Jin, B., Wang, Z., Liu, H., Hu, J., and Zhang, L. (2016). On geocasting over urban bus-based networks by mining trajectories. *IEEE Transactions on Intelligent Transportation Systems*, 17(6):1734–1747.

[Zhao et al. 2016] Zhao, S., King, I., and Lyu, M. R. (2016). A survey of point-of-interest recommendation in location-based social networks. *arXiv preprint arXiv:1607.00647*.

[Zheng et al. 2016] Zheng, K., Yang, Z., Zhang, K., Chatzimisios, P., Yang, K., and Xiang, W. (2016). Big data-driven optimization for mobile networks toward 5g. *IEEE Network*, 30(1):44–51.

[Zheng 2015] Zheng, Y. (2015). Trajectory data mining: an overview. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 6(3):29.

[Zheng et al. 2014] Zheng, Y., Capra, L., Wolfson, O., and Yang, H. (2014). Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(3):38.

[Zheng et al. 2010] Zheng, Y., Xie, X., and Ma, W.-Y. (2010). Geolife: A collaborative social networking service among user, location and trajectory. *IEEE Data Eng. Bull.*, 33(2):32–39.

[Zheng et al. 2009] Zheng, Y., Zhang, L., Xie, X., and Ma, W.-Y. (2009). Mining interesting locations and travel sequences from gps trajectories. In *Proceedings of the 18th international conference on World wide web*, pages 791–800. ACM.

[Zhu et al. 2011] Zhu, H., Li, M., Fu, L., Xue, G., Zhu, Y., and Ni, L. M. (2011). Impact of traffic influxes: Revealing exponential intercontact time in urban vanets. *IEEE Transactions on Parallel and Distributed Systems*, 22(8):1258–1266.

Capítulo 2

Sistemas de Transporte Inteligentes: Conceitos, Aplicações e Desafios

Felipe D. Cunha (PUC-MG), Guilherme Maia (UFMG), Clayson S. F. S. Celes (UFMG), Bruno P. Santos (UFMG), Paulo H. L. Rettore (UFMG), André B. Campolina (UFMG), Daniel Guidoni (UFSJ), Fernanda Sumika H. Souza (UFSJ), Heitor Ramos (UFAL), Leandro Villas (UNICAMP), Raquel A. F. Mini (PUC-MG) e Antonio A. F. Loureiro (UFMG)

Abstract

Urban mobility is a current problem of modern society and large urban centers, which leads to economic and time losses, higher fuel consumption and higher CO_2 emissions. In the literature, it's possible to find works that point to Intelligent Transportation Systems (ITS) as a solution to this problem, and this research topic has received the vast attention of many researchers nowadays. In this context, vehicular networks emerge as a component of the ITS, providing cooperative communication between vehicles and infrastructure and cooperating to improve the flow of vehicles in big cities. In this mini-course, the objective is to discuss ITS, presenting an overview of the area, its challenges, and opportunities. In this way, this mini-course will introduce the main concepts involved in the ITS architecture, its implementation and integration with other computer networks, and how to evaluate its performance. We will also show the main applications in the literature that cooperate for the existence of ITS. In the end, we will discuss the challenges and opportunities found in the areas of interest of the SBRC symposium, among which we highlight: data collection and fusion, characterization, prediction, security and privacy.

Resumo

A mobilidade urbana é um problema atual da sociedade moderna e dos grandes centros urbanos, que ocasiona perdas econômicas e de tempo, maior consumo de combustível e maiores emissões de CO_2. Na literatura, é possível encontrar trabalhos que apontam os Sistemas Inteligentes de Transporte (ITS) como solução para esse problema e esse tema tem recebido destaque atualmente nos principais veículos de publicação. Neste contexto,

as redes veiculares surgem como um componente do ITS, provendo a comunicação cooperativa entre veículos e com a infraestrutura e cooperando para a melhoria do fluxo de veículos nas grandes cidades. Neste minicurso, o objetivo é discutir ITS, apresentando uma visão geral da área, seus desafios e oportunidades. Desta forma, neste minicurso serão apresentados os principais conceitos envolvidos com a arquitetura de ITS, sua implantação e integração com outras redes computacionais, e como avaliar o seu desempenho. Será apresentado também as principais aplicações existentes na literatura que cooperam para a existência do ITS. Ao final, serão discutidos os desafios e oportunidades encontrados nas áreas de interesse do simpósio SBRC, dentre elas destacam-se: coleta e fusão de dados, caracterização predição, segurança e privacidade.

2.1. Introdução

O crescimento desordenado dos grandes centros urbanos tem provocado graves problemas socioeconômicos e estruturais para a população, que contribuem para o aumento das desigualdades sociais e para um estresse significativo à estrutura das cidades. Desta forma, serviços e recursos devem ser providos de forma a lidar e minimizar esses problemas. Dentre eles, pode-se citar a má ocupação do espaço urbano que colabora para gerar diversos problemas de mobilidade. Neste contexto, os sistemas de transporte públicos são uma parte imprescindível para melhorar a mobilidade urbana e é um dos setores mais afetados. Por exemplo, em São Paulo 23% dos moradores gastam pelo menos duas horas para ir e voltar ao seu destino todos os dias [Cintra 2013, ISO 21217:2010 2010].

Com o passar dos anos, os problemas relacionados com o trânsito vêm aumentando devido ao aumento de veículos em circulação e a grande concentração de pessoas em uma mesma região. Segundo estudos realizados pela IBM, a quantidade atual de veículos automotivos no mundo atualmente ultrapassa 1 bilhão e este número pode duplicar em 2020. Com isso, as grandes cidades são as mais afetadas por esse aumento de veículos, com a presença constante de congestionamentos. Por exemplo: pesquisas recentes mostram que São Paulo têm uma perda anula de R$ 40 bilhões, e esta perda esta relacionada a 85% tempo perdido no trânsito; 13% aumento do consumo de combustível; e apenas 2% ao aumento da emissão de gases poluentes. Estes que por sua vez, contribuem também para o aumento do aquecimento nestes centros urbanos [Cintra 2013].

Algumas tentativas de solução para o problema de mobilidade são propostas como: rodízios de placas e incentivos para uso de transporte públicos. Entretanto, essas soluções não obtiveram muito sucesso. Em muitos cenários elas afetam a rotina da população e não obtém sucesso desejado. Nesse contexto, soluções com inteligência, que fazem uso de comunicação podem contribuir para um maior sucesso, melhorando o tráfego nos grandes centros urbanos. Estas soluções proveem aplicações que viabilizam o controle e gerenciamento do trafego, com serviços que vão desde um controle mais assertivo dos horários e rotas de transporte público até a sincronização inteligente de semáforos. Estes serviços compõem o arcabouço dos Sistemas de Transporte Inteligente (*Intelligent Transportation Systems* (ITS)) [Qu et al. 2010].

Sistemas de Transporte Inteligente (ITS) utilizam dados, comunicação e computação para prover serviços e aplicações que podem resolver diversos problemas de transporte nas grandes cidades atuais. Esses sistemas além de disponibilizar serviços para gerenciar e

dar maior segurança as pessoas no transito, também proveem serviços de conforto para os motoristas e passageiros como o acesso às redes sociais e serviços de *stream* de vídeo durante as viagens. Todas essas aplicações se apoiam na colaboração entre os elementos que integram o sistema como os veículos, os sensores e os demais dispositivos móveis. Cada um desses elementos exerce um papel importante, colaborando e sensoriando dados que serão avaliados pelo sistema. Toda essa colaboração de elementos é viabilizada pela comunicação entre os mesmos. Para isso, elementos como antenas e estações de controle podem intermediar essa comunicação. No contexto da comunicação direta entre os veículos, surgem as redes veiculares (*Vehicular Ad Hoc Networks*), um tipo de rede que vem exercendo grande influencia no cenário dos ITS [Karagiannis et al. 2011].

Os serviços e as aplicações providos pelos ITS possuem características e peculiaridades próprias, que deferem das demais aplicações tradicionais. São serviços que geram e consomem diferente quantidade de dados, usam diferentes tecnologias de comunicação com diferentes larguras de banda, alcance e latência. Além de possuírem diferente restrições e qualidades de serviço que diferem de acordo com a aplicação. Por esse motivo, o projeto de um serviço que faça parte destes sistemas se torna um grande desafio. Neste minicurso, o objetivo é discutir ITS e apresentar uma visão geral da área, seus desafios e oportunidades, definindo os principais conceitos envoltos com a arquitetura ITS, sua integração e cooperação com demais redes computacionais. Neste contexto, serão definidos os principais tipos de aplicações, os simuladores e demais ferramentas utilizadas para avaliar o desempenho de serviços neste cenário. Além disso, serão discutidos os desafios e oportunidades encontrados nas áreas de interesse do simpósio SBRC.

O restante deste trabalho está organizado da seguinte forma. A Seção 2.2 discute o conceito de sistemas de transporte inteligente apresentando todos as definições, arquitetura e demais integrações com outras redes. A Seção 2.3 apresenta as principais aplicações em sistemas de transporte inteligente. A Seção 2.4 discute quais são as ferramentas e simuladores existentes para avaliar o desempenho de serviços em ITS. A Seção 2.5 apresenta os desafios e as oportunidades para diversos tópicos de pesquisa atuais relacionados com sistemas de tráfego inteligente. Finalmente, a Seção 2.6 apresenta as conclusões e os trabalhos futuros.

2.2. Sistemas de Transporte Inteligentes

Os Sistemas de Transporte Inteligentes (ITS) têm como objetivo melhorar a segurança e mobilidade dos transportes, como também o aumento da produtividade das pessoas e diminuição dos efeitos nocivos do trânsito. Essa melhoria é alcançada através da integração de tecnologias de comunicação nos veículos e na infraestrutura da cidade.

ITS não é proposto apenas para melhorar as condições do tráfego de veículos, mas também tem a intenção de tornar o setor de transportes mais seguro, mais sustentável e eficiente, evitando os inconvenientes causados pelos congestionamentos dos tráfegos urbanos e efeitos dos problemas climáticos sobre o tráfego. Para isso, o foco é melhorar a gerência dos recursos das cidades e aumentar a comodidade das pessoas através do uso de serviços de informação e alerta. Por conseguinte, essa melhoria contribui para facilitar o fluxo na cidade, diminuindo o tempo gasto em congestionamentos e, consequentemente, reduzindo o consumo de combustível, emissões de CO_2 e perdas monetárias.

Nas seções seguintes serão apresentados os principais conceitos relacionados com Sistemas de Transporte Inteligentes. Assim, na Seção 2.2.1 serão definidas as principais arquiteturas de ITS e seus componentes, salientando as principais diferenças entre cada modelo proposto. Na Seção 2.2.2 será definido as Redes Veiculares, apresentado as principais características e peculiaridades deste tipo de rede. E na Seção 2.2.3 serão apresentados os principais tipos de redes de computadores que podem cooperar para o funcionamento destes sistemas.

2.2.1. Arquitetura

Com o a evolução das tecnologias de computação e comunicação, e o aumento da demanda de serviços ITS com diferentes requisitos, surge a necessidade de uma padronização de modo que se defina a maneira como os dispositivos e componente podem interagir entre eles. Dentre as arquiteturas propostas, pode-se citar a norte-americana, a europeia e a japonesa.

A arquitetura norte-americana (National ITS Achitecture) [of Transport 2016], definida pelo Departamento dos Transportes dos Estados Unidos (U.S. Department of Transportation), descreve como ocorre a comunicação entre seus elementos e subsistemas, com a defingção clara do papel de cada um deles. Essa arquitetura se divide em 4 classes (conforme é ilustrado na Figura 2.1): *Center* que define o centro de controle e gerenciamento de todo o sistema, no qual os serviços são executados; *Field* que engloba toda a parte de infraestrutura do ambiente (RSU, sensores de monitoramento, câmeras); *Vehicles* que são os veículos e sensores embarcados; e os *Traverlers* que define-se pelos dispositivos usados pelas pessoas durante a viagem.

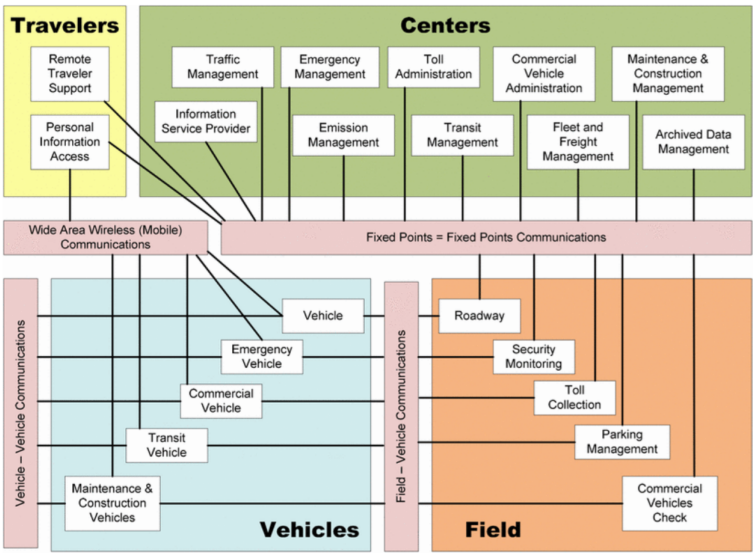


Figura 2.1: USA National ITS Architecture [of Transport 2016].

Algumas desvantagens para a utilização são apresentadas pela arquitetura norte-americana. Num primeiro momento, observa-se a dificuldade que a arquitetura tem em permitir a utilização simultânea de várias tecnologias de comunicação e também de fazer essa escolha de forma dinâmica. Outra restrição é que todos os serviços estão localizados

no *Centers*, e a comunicação entre as duas classes *Centers* e *Vehicles* ocorre por uma interface, o que limita o uso de novos paradigmas.

A arquitetura Japonesa [Lorch et al. 2006], proposta pelo projeto *Smartway* prevê a comunicação entre os veículos e entre os veículos e toda a infraestrutura inteligente das vias (sensores, RSU, Semáforos) e usa como o padrão de comunicação o DSRC, juntamente com o padrão proposto ARIB (similar ao protocolo WAVE). A arquitetura Europeia (ITS ISO CALM) possui características bem similares as demais arquiteturas como o uso de RSU e comunicação DSRC [ISO 21217:2010 2010]. Entretanto, essa arquitetura possui como maior diferença o uso do protocolo de comunicação CALM que provê uma interface de comunicação entre as tecnologias de transmissão como: 3G/4G, Wi-FI, infra-vermelho, entre outras.

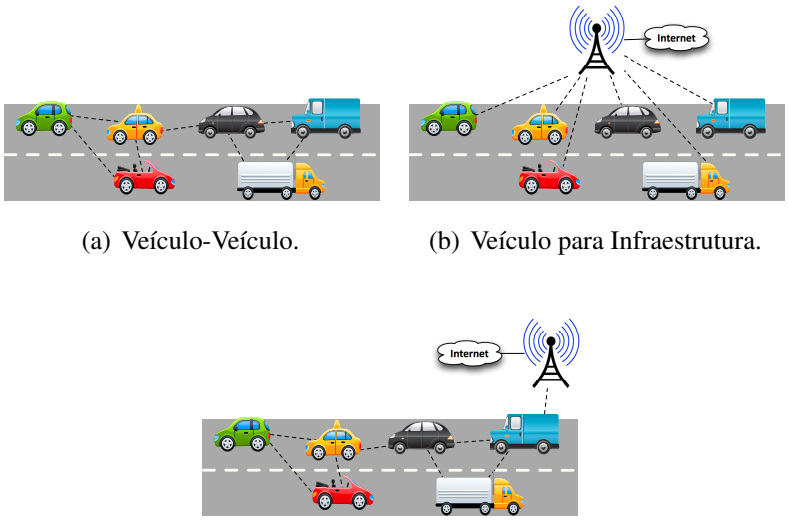
Ambas as arquiteturas Japonesa e Europeia possuem desvantagens comparadas à arquitetura norte-americana, por não possuíram flexibilidade para o uso de novas tecnologias de comunicação e novos paradigmas da computação como por exemplo a computação em nuvem e névoa. Assim, pode-se observar uma necessidade de projetar arquiteturas que permitam a fácil integração de novas tecnologias, uma vez que elas podem cooperar para o desenvolvimento e melhoria de serviços ofertados pelo ITS.

2.2.2. Redes Veiculares

Redes Veiculares são um tipo de rede emergente que tem atraído o interesse de muitos grupos de pesquisa. Estas redes são formadas por veículos com capacidade de processamento e comunicação sem fio, trafegando em ruas e rodovias, enviando e recebendo informações de outros veículos. Elas se diferenciam das redes tradicionais em muitos aspectos. O primeiro deles é a natureza dos nós que as formam, sendo automóveis, caminhões, ônibus etc., que possuem interfaces de comunicação sem fio, e por equipamentos fixados nas proximidades das vias. Além disso, esses nós possuem alta mobilidade e a trajetória deles acompanham os limites e direção definidos pelas vias públicas [Faezipour et al. 2012, Boukerche et al. 2008].

O veículo que participa da rede é equipado com um sistema *on-board* com: computador, interfaces de comunicação, sensores e interfaces para usuário. O sistema suporta uma gama de aplicações para melhorar a segurança do transporte e também proporcionar serviços aos usuários. Uma infraestrutura de rede às margens de rodovias e ruas, denominada de *Road Side Unit* (RSU), também é parte das VANETs e facilita a comunicação dos nós da rede o acesso à Internet. Adicionalmente, os dispositivos portáteis dos passageiros e o sistema do veículo podem se conectar à Internet pela infraestrutura RSU. Um sistema de gerenciamento pode ser adotado para controlar e autenticar a entrada de veículos na rede, principalmente no aspecto da segurança computacional, como distribuição de chaves criptográficas, servidores de autenticação etc. O sistema também pode fornecer serviços e gerenciar a mobilidade dos nós, durante as trocas de rede.

Por se tratarem de nós com alta mobilidade, as redes veiculares possibilitam aos mesmos a troca de informações durante a sua trajetória sem a necessidade de nenhuma infraestrutura entre eles, de forma *ad-hoc*. Assim as redes veiculares podem ser consideradas como um tipo de MANETs (*mobile ad-hoc network*). Entretanto, existe a possibilidade dos nós se comunicarem com a infraestrutura das rodovias, permitindo uma comunicação in-



(a) Veículo-Veículo. (b) Veículo para Infraestrutura.

(c) Arquitetura Híbrida.

Figura 2.2: Tipos de comunicação em Redes Veiculares.

fraestruturada [Alves et al. 2009, Hartenstein and Laberteaux 2008, Yousefi et al. 2006]. Desta forma, considerando essas características peculiares, a comunicação entre os veículos pode ser classificada em três maneiras (conforme ilustra a Figura 2.2):

- Veículo-Veículo (V2V): permite a comunicação direta de veículos sem depender de um apoio de infraestrutura fixa. Neste tipo de comunicação os veículos podem permutar dados das condições da rodovia, detectar a presença de outros veículos, e mesmo informações acerca de veículos em movimentação insegura.
- Veículo para Infraestrutura (V2I): permite que um veículo se comunique com a infraestrutura rodoviária. Desta forma o veículo pode receber da infraestrutura rodoviária informações sobre obstáculos e presença de pedestres; dados das condições da rodovia; anúncios, propagandas e também informações de segurança que auxiliarão numa condução segura.
- Arquitetura Híbrida: combina soluções V2V e V2I. Neste caso, um veículo pode se comunicar com a infraestrutura rodoviária num único salto ou múltiplos saltos de acordo com sua localização em relação ao ponto de ligação com a infraestrutura visando objetivos diferentes.

Atualmente as montadoras de veículos já colocam em circulação automóveis com computadores de bordo, dispositivos de comunicação sem fio, sensores e sistemas de navegação. Esses recursos viabilizam o estabelecimento das redes veiculares. Um exemplo de aplicação desses recursos são os veículos que dispõem de sensores para coletar as condições meteorológicas, estados do veículo, condições da rodovia e até mesmo limite de velocidade das vias. Neste cenário, os veículos podem interagir com a infraestrutura

das rodovias, obtendo informações de tráfego o que gera melhoras nas condições para o condutor tomar decisões no trânsito.

A interação entre os veículos pode evitar o acontecimento de colisões em vias públicas. Pesquisas de trânsito mostram que por ano no Brasil acontecem em média 110 mil acidentes de trânsito, em torno de 300 acidentes por dia. Além disso, 6 mil pessoas vão a óbito e outras 68 mil ficam feridas, gerando aos cofres públicos um gasto de 22 bilhões de reais [IPEA 2012]. Destes acidentes contabilizados, a principal causa apontada foi a falta de atenção dos condutores, seguidos de motoristas que não obedecem a distância de segurança e velocidade incompatível com o local [CESVI 2012]. Estudos mostram que cerca de 60% dos acidentes podem ser evitados se o condutor for avisado um segundo antes da colisão. Neste contexto o uso de redes veiculares pode proporcionar a redução destes valores, por meio da interação veículo-veículo os condutores podem ser alertados de perigos em potencias nas estradas [Yang et al. 2004].

Nas redes veiculares, normalmente as informações devem ser entregues dentro de veículos numa região de interesse considerando a posição geográfica do nó e a relevância da informação ao mesmo. Um desafio nesse contexto é como distribuir as informações aos veículos de forma eficiente, considerando a dinâmica e mobilidade dos veículos na rede e até mesmo a urgência na entrega da informação, a fim de evitar uma colisão. Para isto, uma ferramenta importante a ser estudada é o protocolo de roteamento, que deve ser eficiente, confiável, suportar uma comunicação com múltiplos saltos e intolerante a atrasos. Ainda nesse cenário é importante que o veículo receba o aviso do possível obstáculo, mesmo que os mesmos não estejam no mesmo raio de comunicação [Li and Wang 2007].

2.2.2.1. Padrões de comunicação veicular

Em 1999, a *Federal Communications Commission* (FCC) concedeu o espectro de 5.9 GHz para a comunicação *Dedicated Short Range Communication* (DSRC) com foco na iniciativa *Intelligent Transportation System*. A FCC adotou como base para a camada física e a camada *Media Access Control* (MAC), o padrão IEEE 802.11, porque é um padrão estável. Em 2004, o IEEE *task group p* assumiu a responsabilidade do desenvolvimento do padrão IEEE 802.11p para ambientes veiculares. Outro grupo, IEEE *working group 1609*, assumiu a tarefa das especificações das camadas adicionais no conjunto de quatro protocolos: IEEE 1609.1 , 1609.2, 1609.3 e 1609.4. Coletivamente, IEEE 802.11p e IEEE 1609.x é chamado de *Wireless Access in Vehicular Environments* (WAVE), seu objetivo é facilitar o provimento do acesso a rede em ambientes veiculares [Jiang et al. 2008].

A frequência de operação do WAVE tem como base a comunicação DSRC; nos Estados Unidos é definida na frequência de 5.9 GHz, 75 MHz de banda e existe a separação dos canais de controle (CCH), reservados para a transmissão de mensagens de aplicações de segurança (*safety*) e canais de serviço (SCH) para a troca de dados de mensagens para ambas aplicações de segurança e de entretimento/serviços (*non-safety*). A alteração fundamental introduzida pelo WAVE é permitir um veículo transmitir e receber dados sem a necessidade de pertencer a um *Basic Service Set* (BSS), a priori. Isso significa que os veículos podem se comunicar imediatamente um com o outro a partir de um tempo de contato, sem qualquer sobrecarga adicional, considerando que operam no mesmo

canal [Jiang and Delgrossi 2008].

O padrão WAVE é dividido em duas partes [Uzcategui and Acosta-Marum 2009]: *i) RoadSide Unit* (RSU) que podem ser instaladas em postes de iluminação, semáforos, sinais de trânsito e assim por diante; e *ii) Onboard Unit* (OBU) que são instaladas nos veículos (carro, moto, caminhão, ônibus). As partes do padrão operam de maneira independente e os veículos podem se organizar em pequenas redes chamadas *WAVE Basic Service Set* (WBSS). A WBSS pode consistir somente de OBUs ou uma mistura de OBUs e RSUs, como ilustrado na Figura 2.3. Os membros de um determinado WBSS trocam informações por meio de alguns canais de serviço (SCH) e de controle (CCH). Porém, pacotes de *Internet Protocol* (IP) são permitidos apenas no canal SCH e os veículos devem ser membros da mesma WBSS.

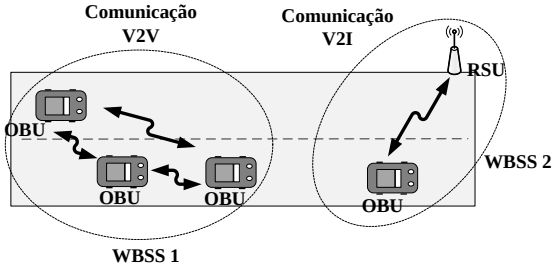


Figura 2.3: As comunicações das partes do padrão WAVE. WBSS 1: comunicação entre OBUs. WBSS 2: OBU comunicando com a RSU - adaptado de [Uzcategui and Acosta-Marum 2009]

As duas pilhas de protocolos do padrão WAVE referente aos dados, IP e *WAVE Short-Message Protocol* (WSMP), podem ser observadas na Figura 2.4. Análogo à terminologia do modelo de referência *Open Systems Interconnection* (OSI), ambas as pilhas usam a mesma camada física e camada de enlace e as camadas de sessão e apresentação. O motivo de ter duas pilhas de protocolos é para acomodar comunicações de alta prioridade e sensíveis ao tempo, bem como a tradicional comunicação por IP [Uzcategui and Acosta-Marum 2009].

O IEEE 802.11p é limitado pelo âmbito do IEEE 802.11, ou seja, unicamente o nível físico (WAVE PHY) e de acesso ao meio (WAVE MAC). Na banda de 75 MHz, são alocados multicanais de 10 MHz e taxas de dados de 3 a 27 Mb/s por canal. O problema relacionado ao gerenciamento dos multicanais da comunicação DSRC é resolvido pelas camadas superiores definidas pelos padrões IEEE 1609.x. Em particular, o padrão IEEE 1609.4 permite às camadas superiores realizarem, de maneira transparente, operações através de múltiplos canais, sem a necessidade de conhecimento dos parâmetros da camada física [Jiang et al. 2008].

A ideia é monitorar periodicamente o canal de controle (CCH), para receber mensagens de controle e de advertência e, posteriormente, ajustar para um dos canais de serviço SCH disponíveis, para a troca de dados não relacionados à segurança. O regime de coordenação divide o tempo do canal em intervalos de sincronização de 100 *ms*, que consiste em alternar a cada intervalo de 50 *ms* para CCH e para o SCH, com um tempo de guarda de 5 *ms* [Hartenstein and Laberteaux 2008].

As demais camadas do WAVE, em linhas gerais, são definidas como [Karagiannis et al. 2011]:

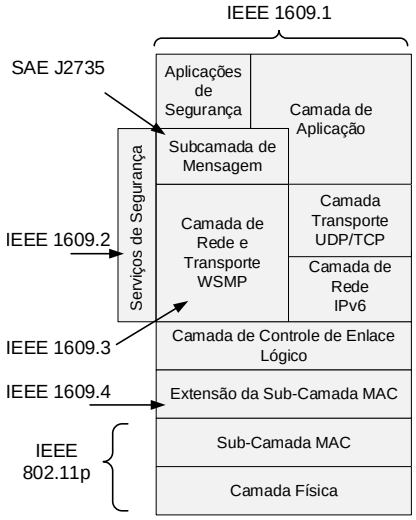


Figura 2.4: Padrão WAVE: As duas pilhas de protocolos IP e WSMP. A camada *Security* pode não se encaixar facilmente no modelo de referência OSI – adaptado de [Kenney 2011]

- *IEEE 1609.3 Network Services*: Fornece serviços de roteamento e endereçamento necessários na camada de rede WAVE; o *WAVE Short Message Protocol* (WSMP) facilita o roteamento por meio do provimento de grupos de endereços para aplicações de segurança. Além disso, utiliza ambos os canais de controle (CCH) e de serviço (SCH).

- *IEEE 1609.2 Security Services*: Especifica os conceitos de segurança do WAVE e define os formatos das mensagens e seu processamento para a comunicação segura. Adicionalmente, determina as circunstâncias para o uso da segurança na troca de mensagens.

- *IEEE 1609.1 Resource Manager*: Descreve a aplicação em uma OBU com recursos limitados que pode utilizar, remotamente, recursos de processamento de outras entidades de forma transparente.

2.2.3. Integração com Outras Redes

Com os avanços e crescente disponibilidade de tecnologias sem fio que oferecem acesso a rede em diversos padrões, tais como: IEEE 802.11, 3G/4G, LTE e Bluetooth, que podem ser usados para equipar as redes de sensores, as redes de veículos não tripulados e redes veiculares. Assim, encontramos redes de celulares (4G/LTE) fornecendo comunicação de longa distância e acesso à Internet para os veículos, e em curta distancia o padrão DSRC (*Dedicated short-range communications*) propiciando comunicação de curta distância de maneira *ad-hoc*. Neste cenário, os sistemas de transporte inteligentes devem prove serviços aos condutores e passageiros em qualquer hora e lugar. E o sucesso e disponibilidade desse serviço dependerá da integração de diferentes tecnologias e redes.

Em [Hameed Mir and Filali 2014], os autores apresentam uma análise de desempenho dos dois padrões de comunicação em redes veiculares para diferentes cenários, densidades e velocidades de veículos. Pode-se observar que o DSRC bons resultados em

cenários de redes esparsas. Mas devido as suas limitações de raio de comunicação, o seu suporte a mobilidade dos veículos é bem limitado. Já o padrão LTE apresentou um bom desempenho quanto a escalabilidade, confiabilidade e suporte a mobilidade. Entretanto, o mesmo apresenta alguns desafios para lidar com as restrições de atraso em algumas aplicações.

Quando se trata da obtenção de dados, os sistemas de transporte inteligentes devem fazer uso da integração com as redes de sensores (WSN) e as redes de veículos não tripulados (FANET). Os dados dos sensores podem ser combinados com outros dados coletados pelos veículos para, por exemplo, inferir o posicionamento de um nó da rede (veículo, RSU, dispositivo móvel do usuário), fornecer a densidade de veículos nas vias, apontar a presença de pontos de alagamentos e com obstáculos, etc. Levando em conta os veículos não tripulados, os mesmos podem ser aplicados em ocasiões especiais como acidentes ou enchentes, para ajudar na coleta e disseminação de dados. Nestes casos, ajudariam na difusão de mensagens de alerta por meio de estabelecimento de *links* de comunicação em locais onde a infraestrutura RSU foi danificada ou não estão disponíveis.

Considerando demais aspectos de tecnologia de transmissão de dados, esses padrões podem também ser usados para o estabelecimento de comunicação entre os ITS e toda a infraestrutura inteligente de tráfego. A reprogramação de semáforos, a leitura de dados de câmeras e sensores instalados nas vias públicas, comunicação com radares, etc. Todos esses dispositivos devem ser capazes de se comunicarem com as centrais de monitoramos de tráfego a fim de fornecer dados que possa colaborar com o gerenciamento de todo o tráfego.

2.3. Aplicações

Em ITS, grande parte das aplicações são projetadas para auxiliar os motoristas e passageiros durante suas viagens, visando reduzir acidentes e gerenciar o tráfego das grandes cidades. Além disso, existem outros tipos de aplicações que auxiliam e promovem serviços aos condutores, tornando a viagem mais tranquila e prazerosa. Nesta seção será apresentado uma classificação das aplicações existentes, com a discussão de cada categoria apresentada, exemplificando com trabalhos atuais encontrados na literatura.

2.3.1. Aplicações de Segurança:

Também chamados de aplicações críticas, esta classe de aplicações tem como objetivo avisar o condutor sobre a possibilidades de colisões eminentes com outro veículo ou com algum obstáculo à frente. Em alguns cenários, o condutor precisa reagir tomando uma decisão rápida a fim de evitar a colisão. Por esse motivo, esse tipo de aplicação apresentar severas restrições de *delay* e confiabilidade. Alguns tipos de aplicações existentes nesta classe são: alerta de perigos na rodovia, colisões emitentes, acidentes na pista e obras à frente. Todos os serviços devem trabalhar de forma a evitar colisões entre os veículos e acelerar o acesso ao socorro.

Na literatura pode-se encontrar alguns trabalhos que exploram aplicações neste contexto. Em [Zaldivar et al. 2011] os autores apresentam uma aplicação que fez uso de uma aplicação Android e interações com a porta OBD do veículo para detecção da ocorrência do acidente. Em caso de acidente, a aplicação é acionada e os telefones

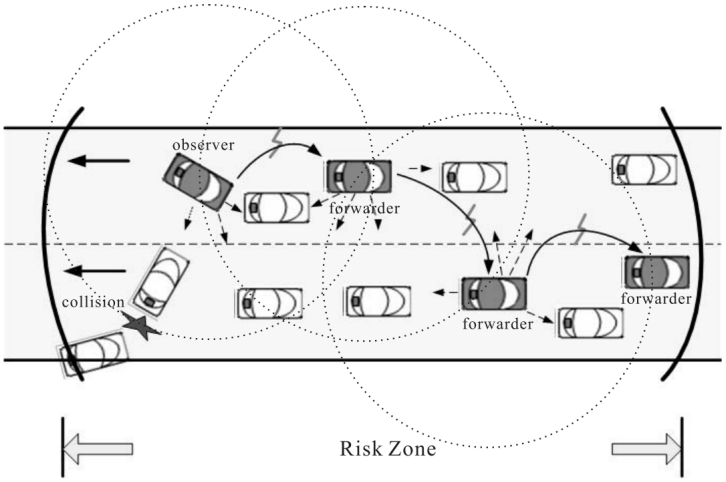


Figura 2.5: Um exemplo ilustrando um cenário de entrega de mensagem de emergência para aplicações de segurança [Tseng et al. 2010].

marcados como números de emergência serão avisados. Em experimentos, os autores mostraram que em menos de 3 segundos a aplicação reage para alertar acerca do evento ocorrido.

De forma diferente, em [Fazio et al. 2013], os autores exploram a comunicação V2X para alertar os veículos na proximidade de um acidade, alertando sobre o evento. Utilizando o protocolo WAVE, a aplicação faz uso das interações entre os veículos-veículos e veículos-infraestrutura para realizar a troca de mensagens e alertar todos os veículos no raio de influencia do acidente. De maneira semelhante, em [Chiasserini et al. 2005] os autores apresentam um estudo do desempenho da disseminação de mensagens de alertas avaliando um mecanismo de controle de acesso ao canal de forma a melhorar a eficiência da transmissão de mensagens em VANETs. Como observado na Figura 2.5, em [Tseng et al. 2010] os autores apresentam um mecanismo de encaminhamento de mensagens de alerta ciente da densidade de veículos na região interesse de interesse que foca em reduzir os longos atrasos na entrega e o *overhead* dispensado na tarefa. A aplicação trabalha a partir da origem da mensagem elegendo nós intermediários para a tarefa de retransmissor.

2.3.2. Aplicações de Eficiência de Tráfego:

O aumento exacerbado no número de veículos, em conjunto com limitações na infraestrutura rodoviária tornaram o congestionamento de veículos um dos principais problemas dos grandes centros urbanos em todo o mundo. A ineficiência no tráfego de veículos está associada à uma série de problemas, tais como, aumento no número de acidentes, efeitos negativos no desenvolvimento econômico e problemas ambientais [Bauza et al. 2010, Karagiannis et al. 2011].

De acordo com um relatório do Departamento de Trânsito dos EUA, existem três causas principais para o surgimento de congestionamentos [of Transportation 2015]. A

primeira está relacionada aos eventos capazes de influenciar o tráfego, como por exemplo, incidentes, obras nas vias, mau tempo, etc. A segunda está relacionada à demanda de tráfego, caracterizada por flutuações nas condições normais de tráfego ou eventos especiais, tais como, shows, eventos esportivos, etc. A última está relacionada às características da infraestrutura rodoviária, representada pelos dispositivos de controle de tráfego, tais como semáforos, e os gargalos físicos na infraestrutura. O relatório também afirma que os gargalos físicos são responsáveis por 40% dos congestionamentos, seguido por incidentes (25%), mau tempo (15%), obras nas vias (10%), má programação da temporização dos semáforos e eventos especiais (5%). Como controlar as condições climáticas não é uma realidade e a construção de novas vias é um processo demorado e custoso, a sociedade necessita de novas tecnologias capazes de evitar congestionamentos e seus problemas.

Os Sistemas de Gerenciamento de Tráfego (TMS) consistem em uma série de aplicações e ferramentas de gerenciamento com o objetivo de melhorar os sistemas de transporte através da integração de tecnologias da informação, comunicação e sensoriamento. Na prática, TMS coletam dados relacionados ao tráfego a partir de fontes heterogêneas, utilizam vários tipos de algoritmos para sumarizar, agregar e fundir esses dados visando a geração de informação útil, e finalmente, utilizam essa informação para conceber aplicações e serviços para os usuários com o objetivo detectar, controlar e reduzir os congestionamentos. A Figura 2.6 mostra uma arquitetura genérica de um TMS. A seguir, são apresentadas algumas soluções que visam melhorar a eficiência de tráfego.

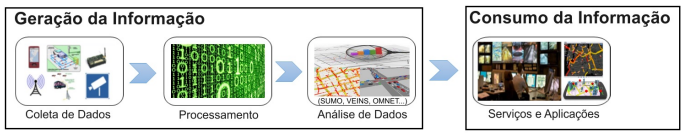


Figura 2.6: Arquitetura de um Sistema de Gerenciamento de Tráfego

O CoTEC [Bauza et al. 2010] é um sistema veicular cooperativo que utiliza comunicação V2V e lógica nebulosa para detectar pontos de congestionamento. No CoTEC, cada veículo envia mensagens para seus vizinhos com o objetivo de informá-los sobre a condição de tráfego no local em que o veículo se encontra. Ao detectar um congestionamento, cada veículo envia uma estimativa a respeito da condição de tráfego e então, de maneira colaborativa, os veículos determinam e caracterizam a condição de congestionamento. Esta solução busca apenas identificar condições de congestionamento, e não minimizá-las ou controlá-las. Em [Pan et al. 2012] é proposto um TMS centralizado para obter em tempo real a localização geográfica, velocidade e direção dos veículos com o objetivo de detectar congestionamentos. Uma vez detectado um ponto de congestionamento, os veículos que se aproximarem dessa região são re-roteados, portanto, atuando de forma a aumentar o nível de congestionamento na região.

Em [Brennand et al. 2015] é proposto um TMS que coleta informações de tráfego em tempo real na tentativa de detectar e gerenciar congestionamentos. Nesta solução, RSUs são instaladas em vários pontos de forma a garantir total cobertura em uma cidade. Cada RSU é responsável por gerenciar um conjunto de veículos e detectar congestionamentos em sua área de cobertura, conforme ilustrado na Figura 2.7. Além disso, esta solução

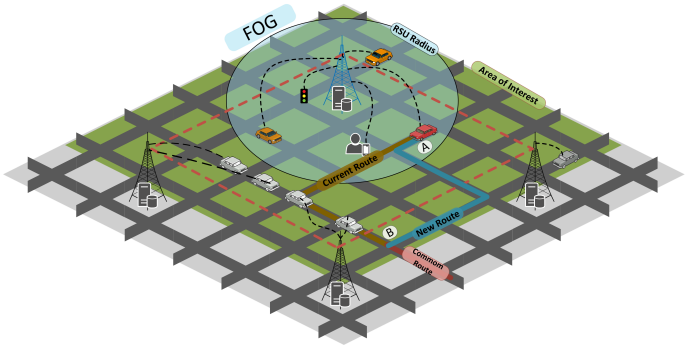


Figura 2.7: Arquitetura do Sistema de Gerenciamento de Tráfego proposto em [Brennand et al. 2015]

inclui um mecanismo de controle de congestionamento, o qual executa periodicamente o re-roteamento de todos os veículos de acordo com as informações de tráfego coletadas.

2.3.3. Aplicações de Entretenimento e Conforto:

Os objetivos das aplicações de conforto são tornar a viagem dos passageiros do veículo mais confortável reduzindo a carga de trabalho do motorista. Nas últimas décadas, várias montadoras de veículos incorporaram aplicações de conforto para ajudar o motorista em uma direção mais confortável e segura. Nesse tipo de aplicações, informações sobre tráfego nas ruas/avenidas/rodovias são sensoreados pelos elementos da rede, que podem ser os veículos, sensores instalados em rodovias ou semáforos, pontos de acesso ou por dispositivos móveis de passageiros/pedestres. Após a coleta dessas informações, as mesmas são disseminadas para os veículos.

As aplicações de conforto necessitam basicamente de duas formas de conectividade:

- Conectividade de Internet: o acesso constante à Internet se tornou um requisito primário para uma grande quantidade de aplicações de conforto, tais como: informações sobre o tempo, tráfego na rede, pontos de interesse presentes no percurso a ser realizado (ex. Postos de gasolina, restaurantes, conveniências) ou até mesmo jogos on-line.
- Conectividade peer-to-peer: para aliviar o cansaço de longas viagens, engarrafamentos ou falta de conectividade com a internet, os passageiros de um veículo podem compartilhar arquivos, músicas, imagens, vídeos, conversar ou até mesmo se divertir com jogos em rede com outros passageiros de outros carros.

Ambas as formas de conectividade devem tratar aspectos específicos de redes veiculares, como mobilidade de veículos e desconexão frequente. Soluções para ambos problemas devem ser tratadas de maneira transparente para os usuários das aplicações. A seguir, faremos uma descrição não exaustiva de várias aplicações de conforto.

Serviços de notificação: consiste em fornecer informações aos assinantes de serviços utilizando acesso à internet. Após a assinatura de um serviço específico, o usuário pode ser

notificado sobre informações relacionadas à previsão do tempo no local atual, informações relacionadas a previsão do tempo no percurso ou no destino da viagem. Além disso, o usuário de um serviço também pode receber notificações sobre condição de tráfego durante seu percurso.

Serviços ao motorista: consiste em fornecer informações sobre o mapa da cidade ou do percurso a ser realizado. Informações sobre postos de gasolina, restaurantes, farmácias, oficinas mecânicas, áreas de estacionamento, localização e horário de funcionamento de museus, shoppings ou eventos. Algumas dessas informações podem ser obtidas diretamente através da internet. Outras, podem ser obtidas utilizando a comunicação entre veículos, onde um veículo que passou por algum ponto de interesse dissemina informações relevantes para outros veículos.

Monitoramento do veículo: esse serviço permite que as montadoras de carro ou outras empresas confiáveis monitorem o funcionamento do veículo remotamente. Estatísticas como tempo de funcionamento do veículo, quilometragem rodada, consumo de combustível, nível do óleo, freios, pressão dos pneus, filtro do combustível, limpeza do ar-condicionado são coletadas e enviadas para empresas. Além disso, a própria aplicação pode notificar o motorista sobre informações do veículo. Dessa forma, não é necessário se preocupar com revisões a cada 10.000km ou uma vez por ano. A revisão será feita apenas quando necessário.

Estacionamento automático: além de obter informações sobre pontos de estacionamento, pagamento automático da tarifa ou agendamento de uma vaga a ser utilizada, o veículo pode realizar o estacionamento sem a supervisão do motorista [Paromtchik and Laugier 1996]. Algumas montadoras já introduziram esse serviço de conforto para seus usuários [BMW , Bos].

Compartilhamento de informações: servidores dedicados (utilizando a internet) ou veículos (utilizando a conectividade entre veículos) podem compartilhar informações de interesse de seus usuários como músicas, filmes, imagens ou arquivos gerais.

Serviços de jogos ou chat: também podem ser considerados aplicações de entretenimento. Proveem o serviço distribuído de jogos e chat utilizando apenas a comunicação entre veículos.

Para se conseguir atingir os objetivos funcionais das aplicações descritas, características de comunicação estão fortemente relacionadas às exigências tecnológicas e de infraestrutura de comunicação da rede, que podem variar de uma aplicação para outra. Por exemplo, aplicações de notificação de eventos, necessitam de uma conectividade com a internet e baixa largura de banda. Por outro lado, aplicações de troca de vídeos ou *streaming* entre veículos necessitam de uma comunicação entre veículos robusta e de tempo real. Aplicações de estacionamento automático também necessitam de uma infraestrutura de localização de alta precisão, que inclui sensores nos carros e utilização de GPS.

Seguem algumas informações mais detalhadas sobre soluções para aplicações de conforto. Com o objetivo de reduzir o congestionamento e o tempo para encontrar estacionamento, os autores em [Tasseron et al. 2016, Tasseron and Martens 2017] propõe um sistema para a reserva de estacionamento em ruas. Os autores discutem que os trabalhos da literatura abordam a reserva de estacionamento em locais próprios

para estacionamento. Assim, sensores instalados em carros e utilizando comunicação V2V. Os sensores monitoram espaços vazios e disseminam a informação de estacionamento para veículos próximos. Outros autores também abordam o problema de estacionamento em redes veiculares [Peng and Li 2016, Chen et al. 2017, Timpner et al. 2016, Rajabioun and Ioannou 2015, Wu et al. 2014]. Em [Lee et al. 2006] os autores propõem um sistema sistema peer-to-peer para compartilhamento de arquivos. Um sistema colaborativo para realizar *downlaod* de conteúdo é proposto em [Huang and Wang 2016]. Para se adaptar à rápida mudança na topologia de uma rede veicular, o sistema proposto divide a rede em células e o roteamento dentro de cada célula é feito de maneira peer-to-peer. Em [Lequerica et al. 2010], os autores exploram a criação de uma rede social no contexto de uma rede veicular para prover "social services".

2.3.4. Aplicações de Sensoriamento Urbano:

Veículos são a mais rica plataforma de coleta e computação no cenário de redes móveis *ad-hoc*. Um veículo moderno possui centenas de sensores que refletem tanto aspectos dos seus sistemas internos, quanto a influência do ambiente sobre o seu funcionamento. Além disso, o tráfego de veículos segue um padrão de mobilidade bem definido pela estrutura viária das cidades, o que permite a utilização de protocolos de VANETs para a comunicação com outros veículos, infraestrutura e, até mesmo, a Internet.

A quantidade de veículos circulando pelas grandes cidades também ressalta um outro benefício da sua utilização como agente de sensoriamento: a sua ubiquidade. Milhares de veículos circulam ao mesmo tempo em diversas regiões de uma cidade. Sendo assim, uma vez que seja possível observar a influência do ambiente urbano nas leituras dos sensores dos veículos, a agregação dos dados de múltiplos veículos possibilitará o sensoriamento de grandes áreas com resolução tão grande quanto o número de veículos contribuindo para o sensoriamento.

Sendo assim, aplicações de sensoriamento urbano são aquelas que buscam extrair, dos dados de sensores veiculares, informações sobre o contexto no qual as medidas foram feitas. A seguir, apresentamos alguns exemplos de aplicações que usam dados de múltiplos sensores e veículos para construir imagens de variáveis do ambiente.

Uma consequência do aumento dos números de veículos trafegando nas cidades é a degradação da qualidade do trânsito, que, por sua vez, aumenta o consumo de combustível dos mesmos. Ganti et al. [Ganti et al. 2010] coletaram dados de consumo de combustível de diversos veículos para determinar o consumo de combustível esperado nas ruas de Urbana-Champaign. De posse de um mapa de consumo de combustível da cidade, os autores desenvolveram uma aplicação que traça a melhor rota entre dois pontos, do ponto de vista do consumo de combustível, que pode ser reduzido em até 10% ao escolher rotas corretas.

A informação sobre a condição das vias de uma cidade é do interesse de múltiplas partes: motoristas, passageiros, prestadores de serviço e administradores públicos. Entretanto, devido à extensão das vias pavimentadas nas grandes cidades, monitorar a condição de vias individuais é inviável devido aos custos envolvidos na implementação de uma infraestrutura para esse fim e sua operação. Chen et al. [Chen et al. 2016] desenvolveram uma aplicação que utiliza os dados de acelerômetros instalados em táxis para monitorar a

qualidade das vias de Shenzhen, conseguindo encontrar, com 90% de precisão, os buracos nas vias trafegadas. Adicionalmente, o número de veículos traz a possibilidade de monitorar o tempo em uma cidade com precisão maior que as estações meteorológicas. Massaro et al. [Massaro et al. 2017] utilizaram um conjunto de dados de mais de 1900 viagens de carros para estimar a temperatura local com base nos sensores veiculares. Os autores mostraram que as leituras de temperatura dos veículos são condizentes com as temperaturas aferidas por estações meteorológicas, no entanto, as primeiras leituras possuem frequência e resolução superiores às últimas, mostrando que é possível monitorar o clima de uma região ou cidade microscopicamente utilizando dados de sensores veiculares.

Um interesse comum à maioria dos motoristas de grandes cidades é o estado do trânsito ao longo de suas rotas. Apesar de valiosa, essa informação é de difícil sensoriamento, uma vez que demanda uma complexa e abrangente infraestrutura para monitorar o estado das vias de uma cidade. Bauza et al. [Bauza et al. 2010] propuseram um sistema que usa dados veiculares compartilhados em uma VANET para identificar congestionamentos no trânsito, bem como sua localização, gravidade e extensão. Por sua vez, Wang et al. [Zuchao Wang et al. 2013] desenvolveram um método de identificação de congestionamentos baseado na sobreposição de trajetórias de dispositivos GPS embarcados em veículos.

2.4. Ferramentas e Simuladores

Nesta seção serão apresentadas as principais ferramentas e simuladores abertos e gratuitos que são utilizados por profissionais e pesquisadores da comunidade de Sistemas de Transportes Inteligentes. Será apresentado o serviço de mapeamento colaborativo OpenStreetMap, o qual permite exportar informações a repeito do mapeamento de regiões específicas, cidades ou até países inteiros. Tais informações podem servir como base para a criação de cenários virtuais, os quais podem ser utilizados tanto por simuladores de tráfego, tais como o SUMO, quanto por simuladores de redes de comunicação, como por exemplo, OMNeT, Veins, ns, etc.

2.4.1. OpenStreetMap

O OpenStreetMap[1] é um serviço de mapeamento construído de maneira colaborativa por usuários, profissionais e entusiastas (Figura 2.8). Dessa forma, uma comunidade de usuários é responsável por contribuir e manter informações a respeito de ruas, linhas ferroviárias, prédios, etc. Tais informações são construídas utilizando-se dados de GPS, imagens de satélite, dentre outras tecnologias de mapeamento.

Um aspecto interessante a respeito do OpenStreetMap é que todas as informações de mapeamento estão livremente disponíveis para qualquer usuário utilizá-las. Com isso, é possível, por exemplo, extrair informações a respeito do mapeamento de regiões específicas, cidades, países ou até mesmo continentes. O OpenStreetMap permite selecionar uma região específica a partir da qual o usuário deseja exportar informações de mapeamento. A Figura 2.9 mostra mostra o processo de exportação de uma região do centro da cidade de Belo Horizonte. Em seguida, é gerado um arquivo XML contendo todas as informações de mapeamento da região selecionada (Figura 2.10). Tal arquivo XML é formatado

[1] `www.openstreetmap.org`

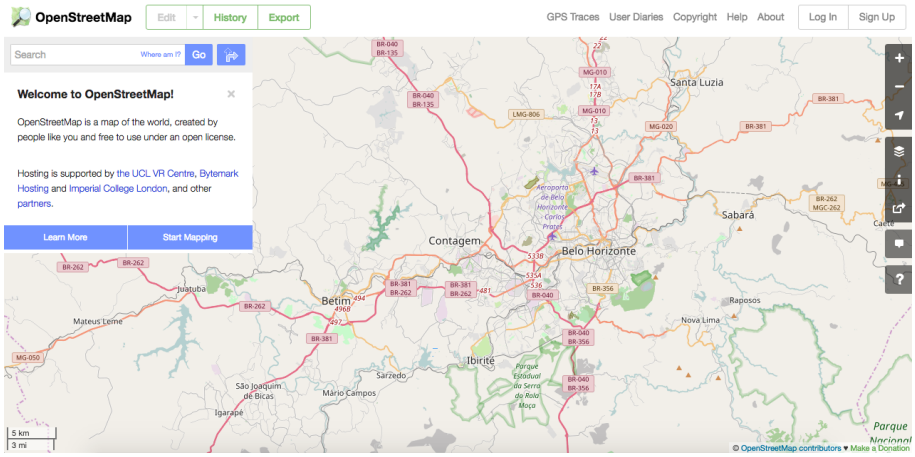


Figura 2.8: Interface do serviço de mapeamento OpenStreetMap

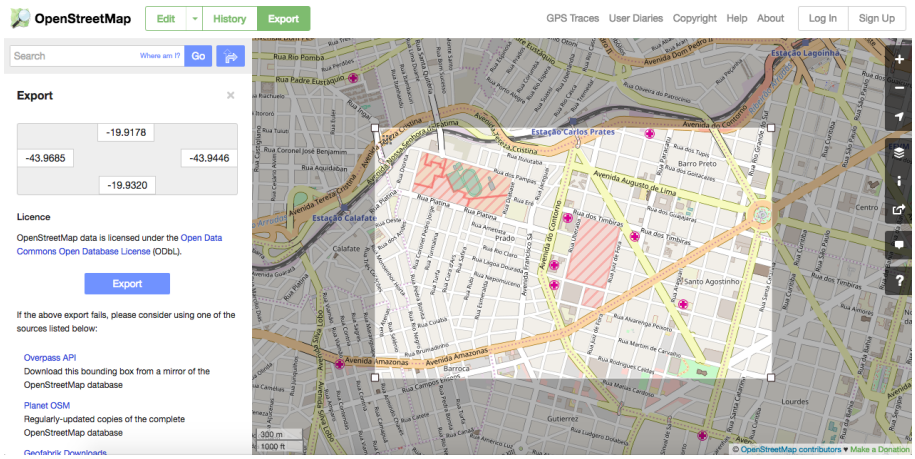


Figura 2.9: Exportando informações de mapeamento a partir do openStreetMap de uma região específica selecionada pelo usuário

obedecendo-se regras previamente estabelecidas pelo OpenStreetMap. Isso permite que tais arquivos de mapeamento possam ser processados por outras ferramentas, tais como o **netconvert**, o qual será apresentado na próxima seção.

2.4.2. Simulator of Urban MObility - SUMO

Pesquisadores e profissionais da comunidade de Sistemas de Transportes Inteligentes utilizam simuladores de tráfego como forma de estudar o impacto de algoritmos de roteamento de veículos, alterações no controle de semáforos e mudanças na infraestrutura viária antes das mesmas serem implementadas no mundo real. Um exemplo de simulador de tráfego amplamente utilizado pela comunidade é o SUMO[2]. Dentre as principais características do SUMO, pode-se citar:

[2] `sourceforge.net/projects/sumo/`

Figura 2.10: Arquivo XML contendo as informações de mapeamento de uma região específica exportada a partir do openStreetMap

- Granularidade dos elementos da simulação: no SUMO é possível modelar e controlar de maneira explícita veículos individuais, pedestres e sistemas de transporte público;

- Importar e criar cenários: o SUMO contém um conjunto de ferramentas que permitem a criação de diferentes tipos de redes rodoviárias, além de possibilitar a criação de cenários rodoviários a partir de informações de mapeamento previamente obtidas a partir de outros serviços, tais como o OpenStreetMap;

- Interação online: o SUMO possibilita a interação com os elementos da simulação de forma online. Ou seja, tal funcionalidade permite, por exemplo, alterar em tempo de simulação as rotas individuais de veículos, a temporização de semáforos, além de permitir a integração do simulador de tráfego com simuladores de redes de comunicação;

- Desempenho: no SUMO é possível realizar a simulação de grandes redes, tais como o tráfego de veículos em uma grande cidade.

Além de permitir a simulação microscópica da mobilidade de veículos e pedestres, a qual é a sua principal funcionalidade, o SUMO fornece um conjunto de ferramentas e bibliotecas que tem o objetivo de facilitar o desenvolvimento dos mais variados tipos de cenários, possibilitando o estudo de várias questões relacionadas ao tráfego de veículos e pedestres. Por exemplo, com o SUMO é possível criar diferentes tipos de infraestrutura viária, importar uma infraestrutura viária a partir de serviços de mapeamento, definir a demanda de veículos e suas rotas, estudar o consumo de combustível e emissão de gases dos veículos, etc. A seguir, serão apresentadas algumas das principais ferramentas que acompanham o SUMO. Para uma lista completa, consultar a documentação oficial do SUMO.

netconvert

Esta ferramenta permite que redes rodoviárias obtidas a partir de serviços de mapeamento, tais como o OpenStreetMap, possam ser convertidas para o formato de redes

rodoviárias compreendido pelo SUMO e suas demais ferramentas. Por exemplo, assumindo que o trecho exportado a partir do OpenStreetMap na Figura 2.9 foi salvo no arquivo *belo_horizonte.osm.xml*, cuja parte de seu conteúdo é exibido na Figura 2.10. Com isso, para converter a rede que se encontra no formato definido pelo OpenStreetMap para o formato compreendido pelo SUMO, basta executar o comando abaixo:

```
netconvert --osm belo_horizonte.osm.xml
```

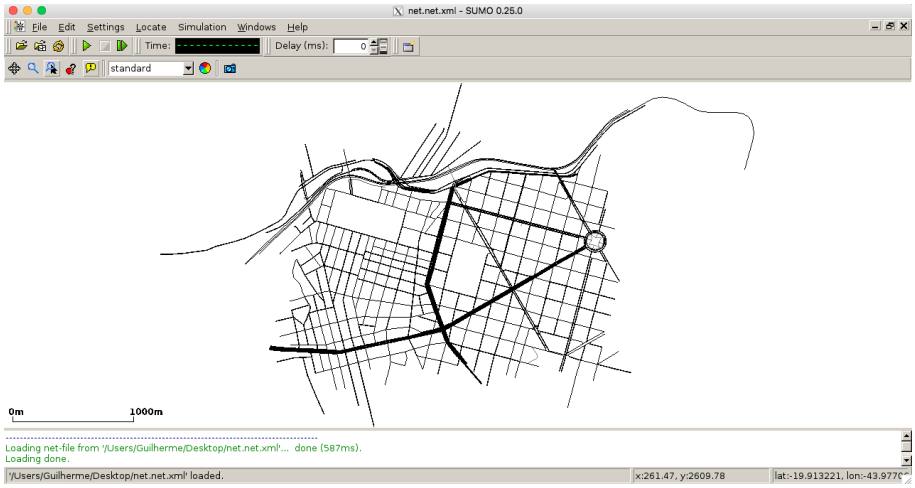


Figura 2.11: Visualização de uma rede rodoviária utilizando-se a interface gráfica do SUMO

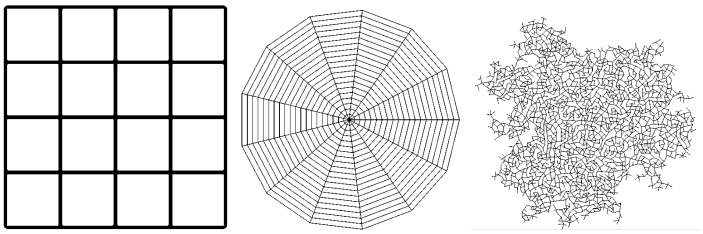
O resultado da execução desse comando é a geração do arquivo *belo_horizonte.net.xml*, o qual nada mais é do que uma descrição da rede rodoviária da região selecionada na Figura 2.9, no entanto, agora em um formato adequado para se trabalhar com o SUMO e suas ferramentas. A Figura 2.11 mostra o resultado dessa conversão, utilizando-se como ferramenta de visualização a interface gráfica do SUMO.

De posse de tal rede rodoviária, é possível, por exemplo, definir uma demanda de veículos e suas rotas utilizando ferramentas como **jtrrouter**, **duarouter** e **marouter**, e estudar o consumo de combustível e o nível de emissão de gases nessa região específica da cidade de Belo Horizonte.

netgenerate

Além de permitir a importação de redes rodoviárias a partir de serviços de mapeamento, o SUMO possibilita a criação de redes rodoviárias abstratas, tais como redes em grade ou redes aleatórias, conforme ilustrado na Figura 2.12.

Essas redes foram geradas utilizando-se os comandos abaixo, os quais especificam basicamente o tipo de rede que deverá ser gerada e o nome do arquivo em que a definição da rede deverá ser armazenada. É importante ressaltar que esta ferramenta possui uma



(a) Rede em grade (b) Rede em teia (c) Rede aleatória

Figura 2.12: Exemplos de redes rodoviárias abstratas geradas utilizando-se a ferramenta netgenerate

série de parâmetros que permitem alterar o formato da rede gerada. Para mais informações, consultar a documentação do **netgenerate**.

```
netgenerate -g -o grid.net.xml
netgenerate -s -o spider.net.xml
netgenerate -r -o random.net.xml
```

TraCI

O TraCI é uma interface de programação para o SUMO que possibilita o acesso à simulações que estão sendo executadas. O TraCI permite, por exemplo, recuperar ou modificar valores dos objetos da simulação de maneira online, ou seja, durante a execução da simulação. O TraCI utiliza uma arquitetura do tipo cliente/servidor, onde o SUMO funciona como um servidor e um programa externo (por exemplo, um *script* em Python ou um simulador de redes de comunicação) funciona como cliente, conforme ilustrado na Figura 2.13.

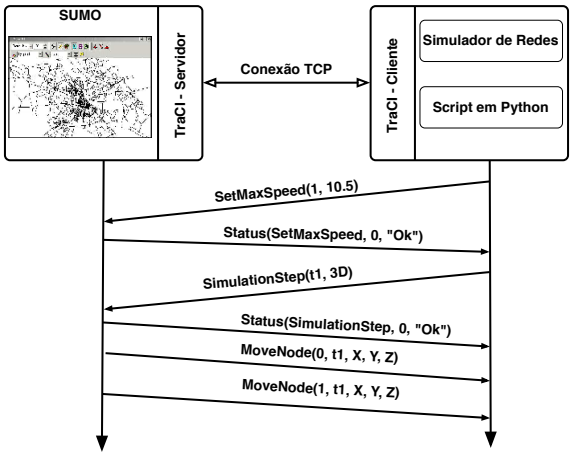


Figura 2.13: Arquitetura do TraCI

O cenário típico para a utilização do TraCI é a simulação de uma rede VANET. Neste cenário, dois simuladores trabalham em conjunto. De um lado, atuando como o servidor, está o simulador de tráfego, que no caso do TraCI sempre é o SUMO. Do outro lado, atuando como cliente, está um simulador de redes habilitado para utilizar a interface de programação do TraCI, como por exemplo, o simulador de redes Veins, que será apresentado na próxima seção. O simulador de redes é responsável por modelar todos os aspectos relacionados à comunicação de dados, como por exemplo, a troca de pacotes entre veículos, a perda de pacotes como resultado de colisões, atenuação de sinal, erros de bits, etc. Já o simulador de tráfego recebe, por exemplo, o tempo de simulação atual e fornece para o simulador de redes, utilizando a interface do TraCI, a localização atualizada dos veículos, que são determinadas obedecendo as condições de tráfego e modelos de mobilidade que estão sendo simulados no SUMO.

A grande vantagem em se utilizar o TraCI é que ele permite alterar os parâmetros dos objetos de simulação do simulador de tráfego (SUMO) a partir do simulador de redes. Neste cenário, é possível, por exemplo, que após a troca de dados entre veículos, o simulador de redes envie um comando para o SUMO alterar a rota de um veículo ou um conjunto de veículos, com o objetivo de evitar uma área congestionada. Perceba que o TraCI aumenta a flexibilidade dos cenários e aplicações que podem ser simuladas em um simulador de redes. Na abordagem tradicional, um *trace* estático de mobilidade com as posições dos veículos para todos os instantes de simulação é fornecido como entrada para um simulador de redes. Neste caso específico, os aspectos de mobilidade não podem ser alterados durante a simulação, inviabilizando o estudo de alguns tipos de aplicações de Sistemas de Transportes Inteligentes.

2.4.3. Veins

Veins[3] é um *framework* de simulação de redes de comunicação composto por um conjunto de modelos especificamente desenvolvidos para o estudo de redes veiculares. A execução destes modelos é realizada pelo simulador de eventos discretos OMNeT++[4] em conjunto com o simulador de tráfego SUMO. A Figura 2.14 mostra a estrutura geral do Veins. Por se tratar de um *framework* de simulação, o Veins serve como base para o desenvolvimento de aplicações específicas. No entanto, como o Veins é composto por vários modelos, é possível utilizá-lo apenas agrupando os modelos disponíveis e modificando alguns poucos parâmetros, o que facilita o estudo de aplicações de Sistemas de Transporte Inteligentes.

No Veins, cada simulação é realizada executando-se dois simuladores em paralelo: o OMNeT++, para a simulação da rede de comunicação, e o SUMO, para a simulação do tráfego de veículos e pedestres. Ambos os simuladores se comunicam através de um *socket* TCP e o protocolo de comunicação adotado é definido pelo TraCI, conforme apresentado na seção anterior. Isso permite a simulação em conjunto tanto de aspectos de comunicação de dados quanto de aspectos de tráfego e mobilidade. O movimento de veículos no simulador de tráfego SUMO é refletido pelo movimento de nós no simulador de redes OMNeT++. Portanto, a interação com o simulador de tráfego permite, por exemplo, simular a influência da comunicação entre veículos no trânsito. O interessante é que no

[3] `veins.car2x.org`
[4] `omnetpp.org`

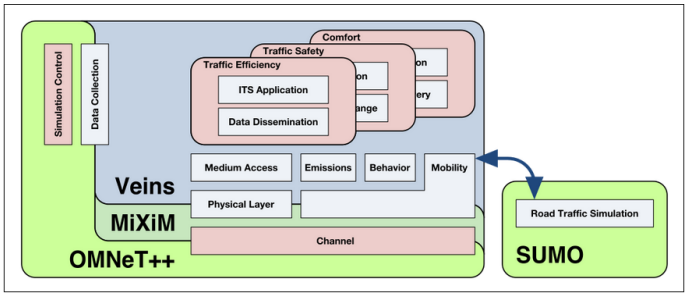


Figura 2.14: Arquitetura do simulador Veins. Fonte: [Vei]

Veins, a interação entre os simuladores de redes e tráfego é totalmente transparente para o usuário, facilitando o desenvolvimento de aplicações. Alguns dos principais modelos disponíveis no Veins são apresentados a seguir.

IEEE 802.11p e IEEE 1609.4 DSRC/WAVE

O Veins inclui um modelo para a simulação de redes sem fio 802.11 especificamente desenvolvido para ambientes de redes veiculares. Este modelo é definido obedecendo-se o padrão de comunicação IEEE 802.11p [IEE 2010]. Dentre as funcionalidades existentes neste modelo pode-se citar a existência de diferentes canais de acesso com QoS que seguem o EDCA (ou seja, 4 filas com diferentes categorias de acesso), características específicas de temporização, modulação e codificação de quadros para ambientes rodoviários, e vários modelos de canais de comunicação, conforme ilustrado na Figura 2.15. O Veins também inclui a funcionalidade de salto de canais, ou seja, a troca entre os canais de controle (CCH) e os canais de serviço (SCH), conforme definido pelo padrão DSRC/WAVE [WAV 2011]. Também estão implementados neste modelo a manipulação das *Wave Short Messages* (WSM), troca de *beacons*, *Base Safety Messages* (BSM) ou *Cooperative Awareness Messages* (CAM).

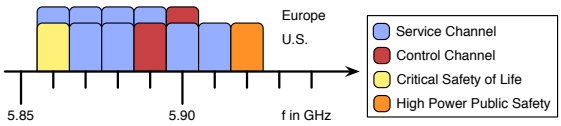


Figura 2.15: Canais disponíveis no WAVE. Fonte: [Vei]

ARIB STD-T109

O Veins também inclui um modelo do padrão de comunicação japonês para Sistemas de Transporte Inteligente ARIB T109 [Heinovski et al. 2016]. Este modelo implementa tanto características da camada física quanto da camada de acesso ao meio (MAC), a qual utiliza uma combinação de TDMA com CSMA/CA.

Modelo de Propagação de Sinal

Modelos precisos de propagação de sinal são fundamentais para o estudo de Sistemas de Transporte Inteligentes. Normalmente, assume-se que o sinal se propaga em condições livres de qualquer tipo de interferência, o que não representa um cenário realístico. Portanto, o Veins implementa o modelo de propagação de sinal Two-Ray Interference, o qual captura de maneira mais realista efeitos como reflexão de sinal [Sommer et al. 2012].

Atenuação de Sinal por Obstáculos

Transmissões de rádio são enormemente afetadas por efeitos de atenuação de sinal. Capturar de maneira precisa estes efeitos é de suma importância no estudo de aplicações de Sistemas de Transporte Inteligentes, especialmente em ambientes urbanos, onde os prédios bloqueiam a propagação dos sinais de rádio. Diante disso, o Veins inclui um modelo de atenuação de sinal causado por obstáculos que captura de maneira realista o efeito de bloqueio de sinal causado por prédios, conforme ilustrado na Figura 2.16.

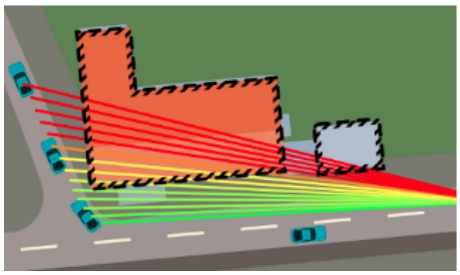


Figura 2.16: Modelo de atenuação de sinal causado por obstáculos que é utilizado no Veins. Fonte: [Vei]

2.5. Desafios e Oportunidades

Nesta seção serão apresentados os principais tópicos de pesquisa atuais relacionados com Sistemas de Transporte Inteligentes. Para cada um dos tópicos também serão elencados os principais desafios e oportunidades de trabalhos futuros.

2.5.1. Infraestrutura de ITS

O cenário dinâmico em que consiste um sistema de transporte é devido à grande mobilidade de seus componentes no ambiente urbano. Embora a mobilidade de pessoas e mercadorias exista há muitos anos, nunca havia alcançado taxas de tão grande escala como a dos tempos atuais. Sendo assim, os problemas enfrentados desde tempos remotos como acidentes, congestionamentos ou situações de perigo também se agravaram com este crescimento.

Com o avanço da tecnologia, os meios de comunicação passaram por uma grande evolução; migrando do rádio, placas de aviso e alertas dos próprios motoristas para dispositivos como computadores de bordo, sensores, telefones celulares, que podem receber notificações em tempo real através de comunicação sem fio. Novas tecnologias permitem uma comunicação mais imediata e dinâmica.

Os sistemas inteligentes de transporte possuem a flexibilidade de adotar uma arquitetura híbrida na qual é possível operar em plena conectividade à Internet, através do uso de infraestrutura, ou assumindo total autonomia do sistema, de forma *ad hoc*. Esta arquitetura possui benefícios como escalabilidade e redução de atraso, mas enfrenta diversos desafios para atuar de maneira eficiente e garantir qualidade e segurança, além de representar um custo adicional nem sempre praticável.

Como parte integrante desta arquitetura, podemos destacar os principais componentes como sensores, OBUs (*on-board units*), RSUs (*road side units*), GPS (global positioning system), semáforos inteligentes, pontos de acesso, dispositivos portáteis (celular, tablet, laptop), satélites, servidores especializados, e a própria Internet. Para garantir a comunicação entre os componentes, diversas tecnologias podem ser adotadas, tais como Wi-Fi, WiMAX, LTE, GSM, 3G, 4G, satélite, bluetooth.

Um dos maiores desafios consiste em projetar soluções de comunicação apropriadas neste conjunto heterogêneo de tecnologias disponíveis. Considerando que o sistema deve operar de maneira colaborativa, é preciso o estabelecimento de padrões que facilitem a integração dos componentes. Além disso, devido à alta mobilidade, é preciso se preocupar com uma deposição adequada de infraestrutura (por exemplo, pontos de acesso, RSUs), além de levar em consideração a tolerância à atraso e falhas, inerente a tais sistemas.

Os componentes de um sistema inteligente de transporte podem ser equipados com múltiplos tipos de transceptores sem fio, podendo se comunicar por mais de um canal de dados sem fio. O protocolo IEEE 802.11p, variante da tecnologia Wi-Fi, provê bandas alocadas para comunicação específica V2V e V2I. A comunicação pode se dar em curto alcance, possibilitando comunicação V2V e V2I, por meio de GPS e rádios DSRC – *dedicated short range communication* (criado para suportar transferência de dados em ambientes de rápida mudança de comunicação) ou longo alcance, principalmente para V2I e I2I, utilizando transceptores de dados celulares, GSM-based, GPRS, UMTS.

O trabalho [Gerla and Kleinrock 2011] destaca a importância e o papel desempenhado pela infraestrutura de Internet no contexto das redes veiculares. Por ser onipresente e prontamente disponível nos diversos ambientes urbanos, a infraestrutura de Internet cabeada pode prover suporte em diversas aplicações, seja no download de propagandas e entretenimento ou no armazenamento de dados sensoriados e enviados pelos próprios veículos. Além disso, conteúdos que já estiverem em poder de algum veículo, poderão ser também compartilhados por conexões P2P oportunísticas entre os veículos e demais dispositivos. Os autores concluem que a grande tendência para Internet do Futuro é justamente a interação entre as comunicações sem fio P2P lado a lado com uma infraestrutura de suporte para o provimento adequado de aplicações e serviços. Entre estes, destacam-se principalmente: segurança de navegação, eficiência de navegação, entretenimento, monitoramento dos veículos, sensoriamento urbano, sensoriamento participativo e emergências.

A seguir, destacamos alguns trabalhos que fazem uso de infraestrutura integrada às redes ad hoc, demonstrando como um sistema inteligente de transporte pode se tornar mais completo e eficiente com o uso de uma arquitetura híbrida, além dos desafios a serem superados.

A deposição de RSUs utilizadas na comunicação V2I através do protocolo IEEE

802.11p é estudada no trabalho de [Gozálvez et al. 2012]. O objetivo principal consiste na análise dos impactos de características urbanas, juntamente com a deposição adequada de RSUs e das configurações de comunicação para garantir que a comunicação V2I obtenha sucesso. Os resultados apresentados para um conjunto vasto de testes conduzidos na cidade de Bolonha demonstram que a qualidade da comunicação V2I através do IEEE 802.11p é fortemente afetada pelo layout das ruas, elevação do terreno, árvores e vegetações, densidade do tráfego, presença de veículos pesados sendo necessário levar tais fatores em consideração na adequada deposição de RSUs e configuração de rádio. Os autores propõem diretrizes a serem seguidas para uma deposição eficiente no projeto de redes veiculares.

Em [Jeong et al. 2010], o problema de entrega de dados I2V é investigado, e consiste em estimar com precisão a posição de destino, considerando o encontro temporal e espacial do pacote e do veículo de destino. A solução proposta, protocolo TSF (Trajectory-based Statistical Forwarding), utiliza uma distribuição de atraso de pacote e uma distribuição de atraso do veículo para selecionar um ponto alvo visando minimizar o atraso de entrega do pacote enquanto satisfaz a probabilidade de entrega de pacote requisitada pelo usuário. É considerada a instalação de RSUs como infraestrutura, veículos equipados com OBUs e comunicação DSRC, GPS presente tanto nos veículos quanto nos nós estacionários e conhecimento da trajetória pelo veículo, que é compartilhada na Internet periodicamente através de pontos de acesso.

O uso de infraestrutura no projeto de sistemas inteligentes de transporte é explorado em diversos trabalhos da literatura. O uso de RSUs pode ser encontrado em [Peng et al. 2006, Trullols et al. 2010]. A fusão de VANET e cloud computing é abordada em [Olariu et al. 2011, Hussain et al. 2012, He et al. 2014]. Mecanismos de segurança são tratados em [Plößl and Federrath 2008, Studer et al. 2009].

2.5.2. Coleta e Qualidade de Dados

Hoje em dia, os veículos modernos têm sistemas embarcados de alta tecnologia que objetivam melhorar a segurança da condução, o desempenho e o consumo de combustível. Para alcançar esses objetivos, os fabricantes têm investido tanto na quantidade quanto na qualidade dos sensores que os veículos possuem [Fleming 2001]. Atualmente, um veículo coleta informações de centenas de sensores que estão conectados à Unidade de Controle do Motor (*Engine Control Unit* – ECU) através de uma rede interna de sensores com fio [Qu et al. 2010] e os dados de saída são acessíveis por meio de uma interface *On-Board Diagnostic* (OBD).

Os sistemas de controle de direção dos veículos modernos dependem fortemente dos dados coletados dos sensores embarcados. Esses sistemas permitem controlar a sua estabilidade e contribuem para uma condução mais segura. Os dados de sensores estão disponíveis através da interface OBD, que foi introduzida para fins regulatórios e de manutenção, mas tem sido explorada para diversas outras finalidades devido às informações que disponibiliza.

Parte dos dados coletados dos sensores dos veículos não representam informações relevantes, do ponto de vista de direção, para os motoristas, uma vez que a maioria desses dados é usada pela ECU e não tem um significado claro para o motorista comum (e.g.,

sensor de oxigênio e pressão de combustível). Além disso, os sensores que indicam informações significativas para o condutor são apresentados por indicadores existentes nos veículos como, por exemplo, rotações por minuto, velocidade e temperatura do motor.

Desse modo, o desafio é extrair informações úteis dos sensores veiculares com o objetivo de correlacioná-los com variáveis internas e externas, possibilitando fornecer serviços personalizados para os motoristas e um sistema de transporte. Para melhor exemplificar o assunto tratado nessa seção, foram coletados dados a partir de adaptadores Bluetooth conectados à interface OBD e *smartphones*.

A interface OBD-II foi introduzida para padronizar o conector físico, os protocolos e o formato das mensagens com as quais eles lidam. O sistema é geralmente empregado para monitorar e regular as emissões de gás e está presente em todos os carros produzidos na Europa e nos Estados Unidos desde 1996 e, no Brasil, desde 2010. A interface OBD também auxilia os serviços de manutenção, ao rastrear a origem de problemas mecânicos [Lin et al. 2009]. Ao possibilitar o armazenamento dos códigos de falha do motor, essas informações fornecem aos mecânicos um histórico de problemas do veículo e possíveis fontes associadas. A Figura 2.17 ilustra o processo de coleta: os dados adquiridos, dos sensores, por meio da interface OBD são transferidos para um *smartphone* com o sistema operacional Android, onde são processados e registrados.

Figura 2.17: Esquema de coleta de dados usando a interface OBD e o *smartphone*

Tabela 2.1: Protocolos utilizados com a interface OBD

Protocolos	Taxa de Transferência
SAE J1850 PWM	41.6 kbit/s
SAE J1850 VPW	10.4 kbit/s
ISO 9141-2	10.4 kbit/s
ISO 14230 KWP 2000	10.4 kbit/s
ISO 15765 CAN	250 or 500 kbit/s

A Tabela 2.1 apresenta os cinco protocolos permitidos com a interface OBD. Todos esses protocolos usam o mesmo conector OBD, porém os pinos têm funções diferentes exceto os que fornecem alimentação da bateria. Os dados coletados dos sensores do veículo estão disponíveis através dos PIDs do OBD. A Tabela 2.2 mostra algumas das informações disponibilizadas via *smartphone*, veículo e também dados fornecidos por sensores virtuais (cujos valores são gerados a partir de dados de sensores físicos e processamento matemático e fusão de dados). Existem também outras centenas de sensores que podem ser acessados através dos PIDs, alguns dos quais são definidos pelos padrões OBD e outros pelos fabricantes dos veículos.

É importante notar que dados provenientes de sensores físicos estão inerentemente sujeito a erros causados por diversos motivos, entre eles a precisão do próprio sensor, o registro dos valores lidos em arquivos e até mesmo falhas no funcionamento tanto do veículo como do sensor [Rettore et al. 2016]. Sendo assim, a primeira etapa do processamento e análise dos dados de sensores virtuais é a sua verificação para garantir que estejam de acordo com os eventos que foram medidos. Entre os fenômenos observados nessa etapa estão dados discrepantes ou *outliers*, informações conflitantes de dois ou mais sensores,

Tabela 2.2: Amostra de dados coletados da ECU e do *smartphone*

Dados Coletados				
Smartphone		*Veículo*		*Sensor Virtual*
Data/Hora	Distância da Viagem	Torque	Rotações Por Minuto	Aceleração
GPS	Nível de Combustível Restate	Fluxo de Combustível	Velocidade	Tempo de Reação
Velocidade (GPS)	Temperatura do Ambiente	Temperatura do Motor	Média de CO_2	Força de Atrito do Ar
GPS HDOP	Custo do km no Inst (R$)	Voltagem	CO_2 Instantâneo	Marcha
Bússola	Custo da Viagem (R$)	Nível de Combustível	Posição do Pedal	
Giroscópio	Barômetro	Temperatura de Entrada do Ar	Média de KPL	
Altitude		Média de KPL da Viagem	KPL Instantâneo	

dados incompletos, ambíguos e correlatos. Feita a verificação dos dados, é possível aplicar a fusão de dados, que têm como objetivo obter novos valores com um significado mais importante que os dados individuais, sem que os resultados obtidos sejam prejudicados pela fonte de informação.

2.5.3. Caracterização de Dados

A informação sobre o contexto dos veículos é fundamental para melhor compreender os padrões de tráfego, o comportamento dos condutores e os padrões de mobilidade de uma cidade. Um exemplo de informação contextual gerada pelos dados coletados dos sensores dos carros é apresentado por Ganti at.al [Ganti et al. 2010], onde o consumo de combustível em toda a cidade foi inferido a partir das leituras de alguns carros. Para determinar quais sensores - individualmente ou combinados - representam melhor o contexto onde o veículo se insere, é necessário, primeiramente, caracterizar os dados das leituras em contextos previamente conhecidos. Para isso, é fundamental que os conjuntos de dados sejam devidamente anotados.

Atualmente, não existem conjuntos de dados publicamente disponíveis contendo um número significativo de leituras de sensores veiculares, desse modo, para demonstrar o comportamento destes dados, um adaptador OBD Bluetooth foi instalado em um veículo para coletar as leituras de seus sensores. Para caracterizar os dados desses sensores, foram selecionados uma amostra de trajeto que compreende uma viagem entre duas cidades – Belo Horizonte e Pedro Leopoldo, MG - Brasil – a 40 km de distância e em condições normais de tráfego.

No processo de coleta, um passo importante é identificar os dados que fornecem informações valiosas sobre o veículo. No estudo de caso, 25 variáveis foram monitoradas, mas apenas 16 destas foram analisadas. Sendo, algumas leituras diretas dos sensores do veículo, outras baseadas em processamentos dos dados coletados (sensores virtuais) e outras utilizam os sensores do *smartphone*. Estas variáveis representam linhas e colunas da matriz de correlação par-a-par da Figura 2.18.

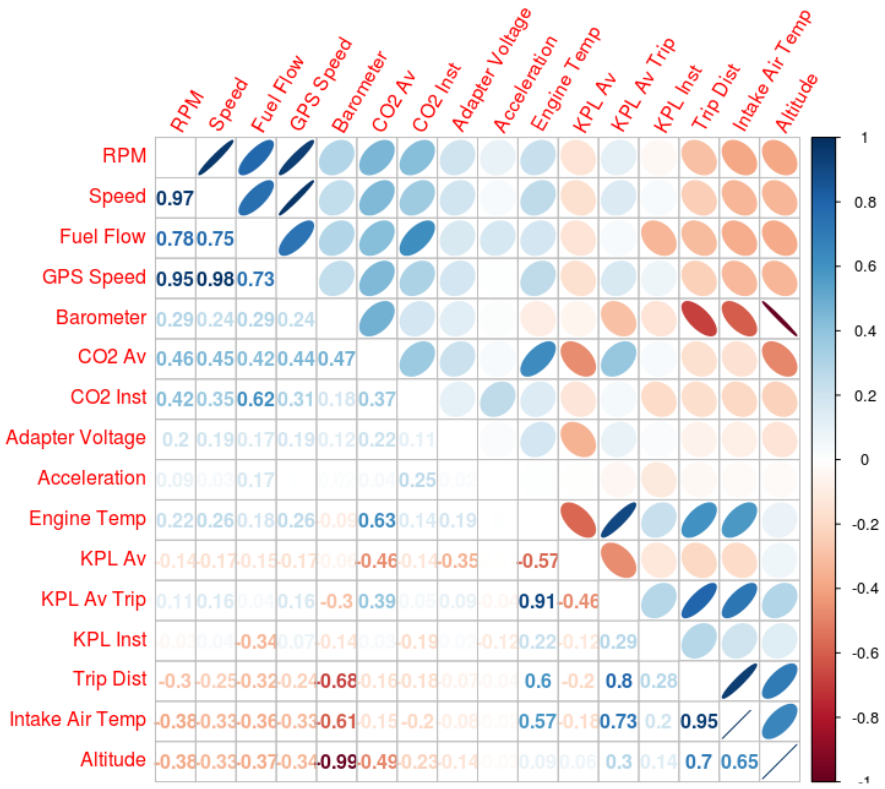


Figura 2.18: Correlação entre sensores

As variáveis que são diretamente coletadas do veículo por meio do scanner OBD são:

1. *Intake Air Temp*: temperatura do ar utilizado na mistura de ar e combustível.
2. *Engine Temp*: temperatura atual do líquido de arrefecimento do motor.
3. *Adapter Voltage*: tensão no módulo de controle.
4. *CO2 Inst*: emissão instantânea de CO_2 do motor.
5. *Fuel Flow*: fluxo de combustível usado pelo motor em um instante.
6. *Speed*: velocidade indicada pelo odômetro.
7. *RPM*: número de rotações do motor por minuto.

As variáveis obtidas a partir de processamentos matemáticos, conhecidas como variáveis virtuais são:

1. *Trip Dist*: distancia percorrida da viagem.
2. *KPL Av Trip*: consumo médio de combustível em quilometro por litro da viagem.

3. *KPL Av*: consumo médio de combustível em quilometro por litro.
4. *Acceleration*: variação da velocidade entre duas observações.
5. *KPL Inst*: consumo instantâneo de combustível em quilometro por litro.
6. *CO2 Av*: média de emissões de CO_2 do motor.

Finalmente, as variáveis obtidas dos sensores embarcados no *smartphone* são:

1. *Altitude*: altitude instantânea do veículo.
2. *Barometer*: pressão atmosférica instantânea.
3. *GPS Speed*: velocidade medida pelo sensor GPS.

A Figura 2.18 também mostra a correlação baseada no *Pearson Product Moment Correlation* (PPMC), entre todos esses sensores em uma viagem. Como a matriz de correlação é simétrica, um lado mostra os valores explícitos da correlação e o outro lado, o mesmo valor é visualmente apresentado como uma elipse, pois corresponde a uma distribuição bivariada com o mesmo valor de correlação. Assim, visualmente, elipses próximas a linhas retas representam dois sensores estritamente correlatos, que podem ser diretas ou inversamente correlacionados, dependendo da direção da linha. Por outro lado, sensores pouco relacionados são representados por um círculo quase invisível, devido à escala de cores e o grau de correlação. Foram considerados altos valores de correlação entre 0.5 a 1.0 ou -0.5 a -1.0, correlação média entre 0.3 a 0.5 ou -0.3 a -0.5, baixa correlação entre 0.1 a 0.3 ou -0.1 a -0.3 e não correlacionados quando igual a 0.

Em uma observação mais detalhada da matriz de correlação, são apontados na Figura 2.19 os quatro diferentes tipos de correlação em seus respectivos graus. Por exemplo, a Figura 2.19(A) representa uma alta correlação entre a velocidade medida pelo GPS e velocidade medida pelo sensor do veículo, destacando uma relação linear. Contudo, alguns pontos não estão alinhados com o relacionamento, isso acontece devido a erros e diferenças nas leituras dos sensores. Outro exemplo de alta correlação está na Figura 2.19(B), que mostra a relação entre pressão atmosférica (identificada como "Barômetro") e altitude. É conhecido que a pressão atmosférica é inversamente proporcional à altitude, assim a relação é quase linear $-0{,}99$.

A Figura 2.19(C) mostra baixa correlação entre -0.1 a -0.3. No entanto, a curiosidade é que o gráfico de dispersão apresenta algo semelhante a uma distribuição exponencial. Esta figura mostra que quanto menos litros são consumidos por quilômetro, mais gases são emitidos. Outro ponto é que as emissões de dióxido de carbono mais baixas ocorrem com o menor consumo de combustível (mais quilômetros por litro) e pode caracterizar momentos em que o motorista para de acelerar.

Finalmente, no extremo da matriz de correlação, é apresentado na Figura 2.19(D) um par de variáveis sem correlação, representado por um coeficiente de correlação de Pearson de -0.08. A relação entre a tensão da bateria e a temperatura do ar de admissão não representa informação relevante. Uma vez que, a tensão da bateria tem seu comportamento

afetado pela aceleração do veículo. Em outras palavras, o alternador funciona com o movimento do veículo e é usado para carregar a bateria e para alimentar o sistema elétrico do veículo. Ao mesmo tempo, o sensor de temperatura do ar de admissão não é afetado pela tensão da bateria.

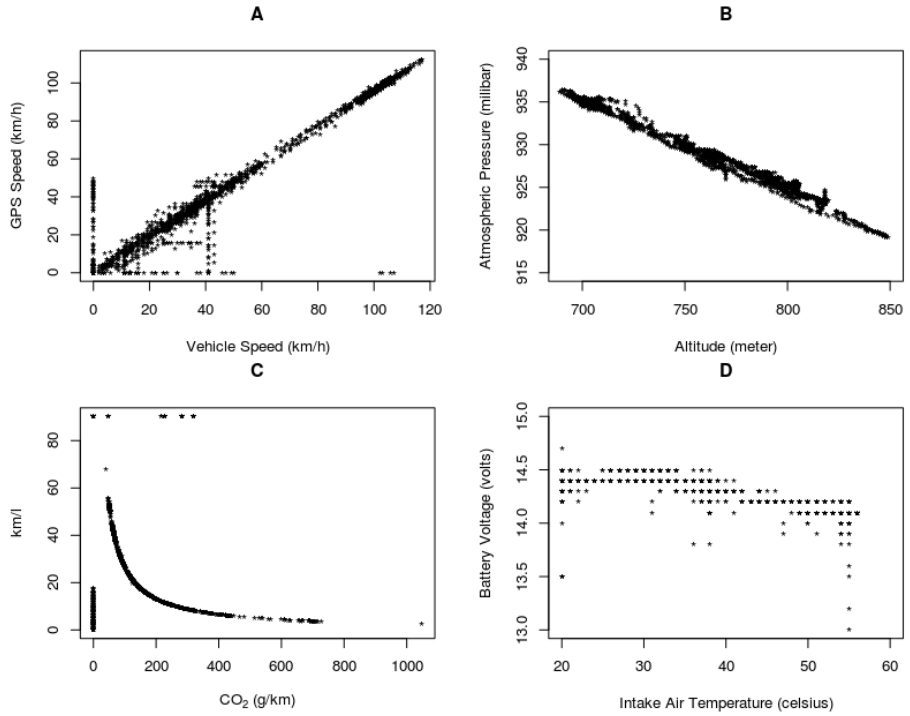


Figura 2.19: Exemplos de correlações entre pares de sensores

Durante o tempo de coleta, foram capturados uma variedade de situações de tráfego: ambientes urbanos com vários níveis de tráfego, rodovias, greves e estradas bloqueadas. Como exemplo das observações realizadas, algumas das leituras dos sensores de sua viagem são ilustradas na Figura 2.20 e representam o estado do veículo. Foi considerado como o estado do veículo, a percepção do contexto em que se localiza através de suas leituras de sensores. No gráfico, as cores das colunas dividem a linha do tempo em cenários: tráfego urbano na cidade de origem, tráfego rodoviário, rotas de acesso à cidade de destino - chamada "Transição" e tráfego urbano na cidade de destino.

O ambiente urbano é caracterizado pelo comportamento da velocidade do veículo, que não sobe acima de 60 km/h, devido à legislação e densidade de tráfego. Esta densidade também é visível no final da linha do tempo, quando o tráfego da cidade de destino é mais intenso e, assim, os carros se movem em um movimento conhecido como parada e arrancada (*stop-and-go*), ou seja, paradas em semáforos ou cruzamentos, movendo nas oportunidades, até que novamente paradas ocorram. Esse tipo de comportamento reflete-se nas linhas horizontais em 0 km/h nos ambientes urbanos, seguidos por pequenos picos de velocidade. A aceleração, que é a variação da velocidade ao longo do tempo, também se comporta diferente nessas situações. Devido à constante aceleração e desaceleração do carro, a variação de velocidade é maior em tais situações.

Por outro lado, a parte rodoviária da viagem mostra um comportamento diferente.

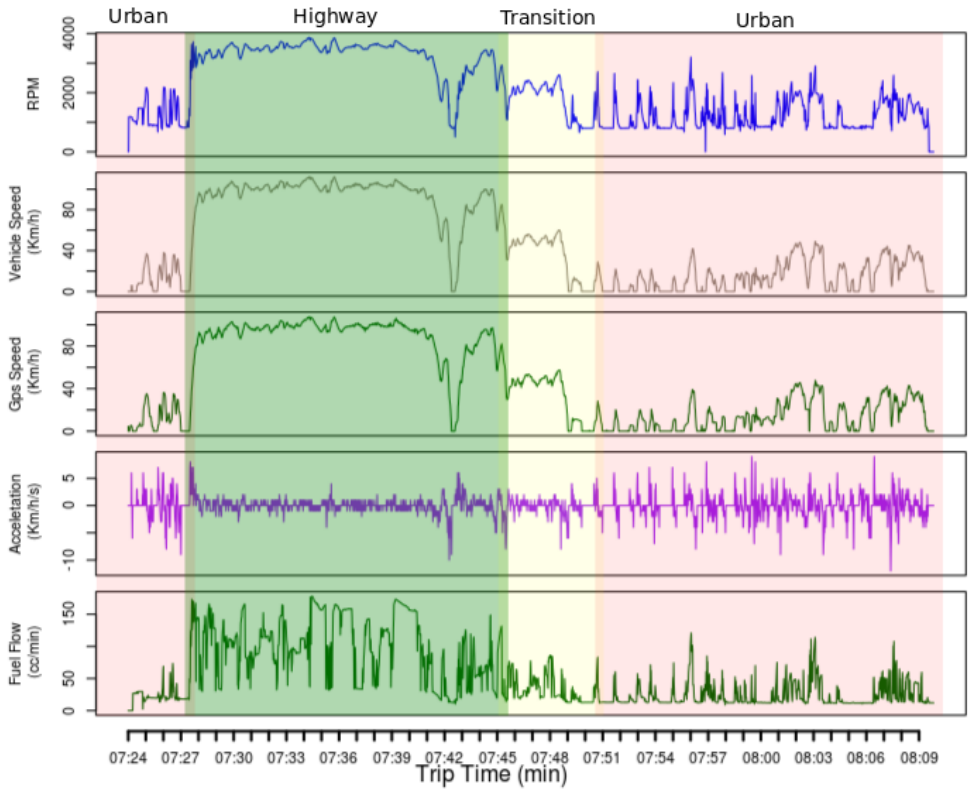


Figura 2.20: Comportamento dos sensores (RPM - Velocidade - Aceleração - Fluxo de Combustível) ao longo do trajeto

A velocidade é constantemente alta e não há grandes intervalos entre aceleração ou desaceleração e a velocidade raramente reduz abaixo de 60 km/h. Para manter o veículo em movimento a velocidades altas, o motor também deve trabalhar mais, traduzido em RPMs mais altas, que também apresentam valores diferentes dos cenários urbanos. Mesmo assim, existem alguns pontos no trânsito urbano onde o motor gira a mais de 3000 vezes por minuto, estas ocorrências são raras e não duram tanto quanto na rodovia, que por aproximadamente 15 minutos as rotações não foram muito inferiores a este valor. Um aspecto único, do cenário de rodovia, detectado nestes dados corresponde ao fluxo de combustível, que é significativamente mais alto, mas não constante quanto o RPM ou a velocidade. Esse comportamento pode refletir a condição da estrada, variações de altitude e pressão atmosférica, torque exigido a cada instante ou até mesmo qualidade do combustível e questões mecânicas do automóvel.

Em resumo, o que se pode notar é que uma caracterização mais detalhada pode ser aplicada com o objetivo de melhor entender a dinâmica desses dados. Por exemplo, a identificação e distinção entre engarrafamentos, greves, bloqueios de estradas e acidentes em uma área urbana e rodovia são questões em destaque e exigem ampla investigação.

2.5.4. Mobilidade e Trânsito

Problemas relacionados à urbanização, principalmente quanto a mobilidade humana e ao trânsito, são uns dos principais desafios de pesquisa relacionados com a qualidade de

vida das pessoas e ao meio ambiente nas cidades. Nesse sentido, diversos esforços têm sido feitos para reduzir congestionamentos, proporcionar meios seguros de locomoção, reduzir poluição ambiental, reduzir poluição sonora, entre outros objetivos. Os sistemas inteligentes de transporte podem desempenhar um papel fundamental no provimento de soluções tecnológicas para se alcançar tais objetivos.

Um desafio é entender a dinâmica das cidades. Graças a popularização de dispositivos com a capacidade de sensoriamento e a evolução dos ITS, um enorme volume de dados tem sido gerado e disponibilizado para análise do comportamento das entidades (e.g., veículos, pessoas) nas cidades, facilitando assim o entendimento da mobilidade humana e o comportamento do trânsito ao longo dos dias. Por exemplo, o Portal Data.rio[5] fornece vários conjuntos de dados abertos que podem ser livremente utilizados para o estudo da mobilidade na cidade do Rio de Janeiro, como foi realizado em [da Cruz et al.].

Outras fontes de dados como redes sociais e aplicativos (*Waze*[6] e *Bing Maps*[7]) são uma poderosa forma de coleta de dados para o estudo da mobilidade e trânsito. Por exemplo, [Silva et al. 2014] analisaram dados de redes sociais (*Instagram*[8] e *Foursquare*[9]) para estudar a dinâmica das cidades, destacando as regiões mais visitadas pelos usuários em diferentes horários do dia. [Tostes et al. 2013] utilizaram dados do *Bing Maps* para analisar e predizer pontos de congestionamentos em Chicago nos Estados Unidos. Esses trabalhos mostram como a disciplina de análise de dados pode ser interessante para facilitar o entendimento da dinâmica das cidades. Por exemplo, identificar quais as principais vias utilizadas pela população, coletar informações sobre a demanda de veículos privados e públicos, descobrir quais as causas dos congestionamentos, entre outros questionamentos. Além disso, existem diversas oportunidades relacionadas com a utilização de fontes heterogêneas de dados, manipulação e processamento de grande volume de dados e técnicas para sumarizar e entender esses dados.

Além da análise para entender a mobilidade da população nas cidades, uma outra perspectiva importante é o oferecimento de serviços que permitam otimizar recursos e utilizar eficientemente os meios de transportes, considerando as particularidades de cada cidade tais como dimensão territorial e populacional, relevo, cultura, entre outros aspectos. Nesse sentido, o restante dessa seção concentra-se em expôs soluções existentes no domínio da mobilidade e trânsito, destacando as principais oportunidades e desafios associados a elas.

Mobilidade compartilhada. Nesse caso, novas soluções de transportes permitem os usuários utilizarem sistemas de meios de transporte compartilhados por um certo tempo tais como carros e bicicletas. Geralmente, nesses sistemas, os veículos estão disponíveis em estações e os usuários podem utilizá-los pagando uma taxa. Nesse contexto, diversos desafios de pesquisa estão relacionados. [Yang et al. 2016] propuseram um método preditivo para balanceamento de bicicletas nas estações com base no estudo de dados de mobilidade em Hangzhou na China. Em [Chen et al. 2015], os autores utilizaram diversas

[5]Data.rio - http://data.rio/
[6]http://www.waze.com/
[7]http://www.bing.com/maps/
[8]http://www.instagram.com/
[9]http://www.foursquare.com/

fontes de dados para explorar o problema de alocação de estações. Além dessas questões, outras oportunidades existem no sentido de investigar a demanda de veículos, monitorar em tempo-real e aumentar a segurança dos condutores, visando a redução de congestionamentos e mitigar poluição sonora e ambiental. Similarmente, alguns esforços têm concentrado em investigar o compartilhamento de veículos [Nair et al. 2013] [Boldrini et al. 2016].

Sistemas de caronas. Uma ocorrência comum em várias cidades é a presença condutores oferecendo caronas para diminuírem custo de viagem, considerando as suas rotinas de mobilidade. Os meios de comunicações digitais potencializaram esse comportamento, pois as pessoas passaram a se organizar nas redes sociais e grupos de mensagens para planejarem as caronas, por exemplo o serviço fornecido pelo aplicativo Blablacar[10]. Nesse sentido, um dos principais desafios para este tipo de sistema consiste na criação de serviços de recomendação que explorem a infraestrutura de ITS como VANET e dados gerados por veículos e pessoas. Por exemplo, [Elbery et al. 2013] propuseram um sistema de planejamento de rotas e recomendação de caronas baseado em informações de redes sociais. [Monteiro de Lira et al. 2016] desenvolveram um aplicativo móvel para sugerir caronas com base na reputação de condutores e satisfação dos usuários.

Sistemas integrados e transporte multimodal. Refere-se em integrar os vários modos de transporte a fim de proporcionar o deslocamento de pessoas. Por exemplo, um sistema integrado entre linhas de ônibus, metrô, bicicletas ou carros compartilhados. Para tanto, diversos desafios devem ser considerados na concepção de sistemas de transportes multimodais, tais como manipulação de informação em tempo-real, análise multicritério, fazer recomendações de rotas, considerar preferências dos usuários. Em [Campigotto et al. 2017], os autores desenvolveram um sistema de recomendação de transporte multimodal que considera, além da distância mais curta e menor tempo de viagem, as preferências dos usuários.

Tecnologias de suporte. A popularização de *smartphones* potencializou o desenvolvimento de aplicativos móveis que fornecem serviços tanto para deslocamento (e.g., Uber [11] e o Lifty [12]) como para obter informações de tráfego (e.g., Waze [13]). Nesse sentido, novas iniciativas que explorem tecnologias de suporte (e.g., computação móvel e ubíqua, *internet* das coisas, sistemas baseados em localização) aos ITS são altamente recomendadas no cenário atual.

Controle de tráfego. Monitorar e controlar fluxo de trânsito de veículos (tráfego) é um importante tópico em sistemas inteligentes de transporte. [Tian et al. 2017] fizeram uma revisão de literatura de trabalhos que utilizam câmeras para monitorar e auxiliar no gerenciamento do tráfego urbano. Eles propuseram uma taxonomia de métodos para detecção, rastreamento e reconhecimento de veículos. Um outro tópico relacionado ao problema de tráfego de veículos é o controle de cruzamentos e interseções, principalmente em horários de pico, para melhorar a fluidez e segurança dos condutores e pedestres. Nesse caso, o desafio consiste em gerenciar semáforos e cruzamentos visando o sincronismo de tráfego entre vias como discutido em [Ye and Xu 2017] e [Shirazi and Morris 2017].

[10]http://www.blablacar.com.br/
[11]https://www.uber.com
[12]https://www.lyft.com/
[13]https://www.waze.com

Detecção e gerenciamento de incidentes de trânsito. Detecção e mitigação de incidentes de trânsito é uma das principais oportunidades de pesquisa no contexto de ITS, visto que pode-se explorar o grande volume de dados gerados pelos veículos ou disponibilizado por usuários por meio de aplicativos móveis e redes sociais. [Pan et al. 2013] propuseram um sistema para detecção de incidentes (e.g, acidentes, eventos esportivos) e sugestão de rotas utilizando dados de localização do veículo e informações compartilhadas por redes sociais. No entanto, existem alguns desafios em aberto como determinar espacialmente o impacto de um incidente, tempo de duração e semântica. Uma possibilidade para investigar soluções para tais desafios é utilizar diferentes fontes de dados e aplicar técnicas de fusão de dados [Castanedo 2013].

Em resumo, soluções tecnológicas em mobilidade e trânsito buscam que as pessoas gastem menos tempo no trânsito utilizando com segurança os diversos tipos de transporte, priorizando o consumo consciente de recursos energéticos e diminuindo o impacto ambiental.

2.5.5. Segurança e Privacidade

Nesta seção será apresentada uma visão geral de segurança aplicada ao contexto de ITS. Para tanto, o conteúdo apresentado a seguir é baseado na RFC 3552 [Rescorla and Korver 2003], a qual descreve boas práticas de segurança na Internet, do relatório técnico [Levy-Bencheton and Darra 2015] que reporta recomendações sobre *cyber* segurança e resiliência em sistemas de transporte público inteligente e, finalmente, baseando-se no nas análises de segurança de informações para ITS apresentadas em [Biesecker et al. 1997].

Para que o ITS seja seguro é preciso estabelecer quais são os componentes de segurança. Neste minicurso são considerados apenas 3 componentes de segurança, porém sem perda de generalidade, o leitor pode encontrar em [Levy-Bencheton and Darra 2015, Biesecker et al. 1997] visões mais extensas sobre o assunto. Os componentes aqui considerados são: (i) *Objetivos* para manter o sistema seguro; (ii) *Ameaças* que podem causar prejuízos aos objetivos e; (iii) *Serviços de segurança* que visam combater as ameaças ao passo que põe o sistema em direção aos objetivos.

2.5.6. Objetivos de segurança para ITS

Neste minicurso serão considerados três objetivos desejáveis no contexto segurança para ITS[14], sendo eles: *confidencialidade, disponibilidade e integridade*.

- **Confidencialidade:** visa assegurar que os dados e o sistema não estejam acessíveis à entidades, processo ou sistemas não autorizados. Por exemplo, não deve ser permitido que usuários comuns tenham acesso aos dados de um sistema de controle de semáforos.
- **Disponibilidade:** tem por intuito permitir o acesso aos dados e sistema à entidades autorizadas, bem como outros processos e até mesmo sistemas. Por exemplo, se uma aplicação de proteção do motorista está ativa, então os dados do veículo devem estar disponíveis e acessíveis ao provedor do serviço.

[14]Note que esses não são os únicos objetivos de um sistema de segurança para ITS. Deste modo, foram apresentados os objetivos fundamentais segundo a visão dos autores.

- **Integridade:** visa assegurar que os dados do ITS mantenham seu significado, completude e consistência. Por exemplo, os dados não podem ser alterados enquanto estão sendo roteados na rede, ou serem deliberadamente removidos ou inseridos ao sistema objetivando desvirtuar o seu funcionamento correto.

2.5.7. Ameaças de segurança ao ITS

De modo geral, uma ameaça é tudo que potencialmente pode causar algum problema ao sistema. As ameaças que um sistema pode enfrentar surgem de três possíveis classes de origem: *desastre natural, acidentais e intencionais*. Em ITS, as ameaças também estão presentes, deste modo, a seguir são listadas as principais ameaças, sua classe de origem e um exemplo no contexto de ITS.

Negação de serviço: a negação de serviço (*Denial of Service (DoS)*) acontece quando ações são tomadas para bloquear acessos ou interromper o funcionamento apropriado de um sistema. A negação de serviço pode ser gerada por eventos intencionais, acidentais ou naturais. Ameaças naturais como inundações, terremotos e outros eventos naturais podem causar a DoS. Geralmente, entretanto, a causa do DoS é devido a introdução de códigos maliciosos ou a execução ações não autorizadas que tornam o sistema indisponível. No contexto de ITS, a negação de serviço pode ser crítica, por exemplo, se um sistema de detecção direção segura se torna indisponível acidentes podem ser gerados.

Exposição de informações *(Disclosure)*: a exposição de informações (*Disclosure*) nada mais é que a interceptação de dados sensíveis (pessoais, financeiros, etc) por entidades não autorizadas. Podendo ser classificados como exposição de informação acidental ou intencional. No contexto de ITS, a exposição pode acontecer, por exemplo, quando veículos trocam informações entre si ou entre a infraestrutura, o que cria vulnerabilidades no ITS que podem ser exploradas.

Alteração de informações: esta ameaça está relacionada com a adição, modificação ou remoção de informações do sistema para produzir efeitos não autorizados ao sistema. Essa ameaça pode ser causada por eventos de cunho natural, acidental ou intencional. Um exemplo do potencial negativo desta ameaça em ITS é a alteração de informações exibidas em rodovias por conteúdo incorreto ou inapropriado (ex: a troca de limites de velocidades das placas eletrônicas de sinalização).

Acesso não autorizado (*Masquerading*): esta ameaça tem relação ao acesso não autorizado de um usuário ou processo ao sistema, de modo que o acesso seja percebido como um acesso autêntico. Caso o usuário não autorizado ganhe acesso ao sistema ele pode obter informações confidenciais, bem como permissões exclusivas para alterar o sistema. *Masquerading* pode ser gerado por eventos acidentais ou intencionais. No que tange *Masquerading* e ITS, o usuário não autorizado poderá, por exemplo, alterar dados de rodovias, enviar informações errôneas para usuários do sistema e até mesmo interromper o funcionamento apropriado do sistema. Tudo isto tem potencial para causar sérios

problemas para o ITS e todas as entidades (usuários, processos e outros sistemas) que se relaciona com o sistema afetado.

Retransmissão (*Replay*): a retransmissão é a repetição de informações válidas sob circunstâncias inválidas para alcançar efeitos não autorizados ao sistema. Este tipo de ameaça pode ter impactos na integridade do sistema, principalmente no que diz respeito ao significado e consistência das informações dentro do sistema. *Replays* são gerados acidentalmente ou de modo intencional por parte de um atacante. No contexto de ITS, essa ameaça pode ser usada para retransmitir dados de identidade e crédito de uns usuários válidos (capturados de forma ilegal) para beneficiar quem obteve, ilegalmente, as informações válidas.

Repúdio: o repúdio é a negação de uma ação. O repúdio viabiliza que o transmissor ou receptor bloqueia a execução de uma ação. Em geral, esta ameaça atinge a integridade do sistema e é causada por eventos acidentais ou intencionais. No âmbito do ITS, o repúdio pode ocorrer geralmente em transações eletrônicas, por exemplo, suponha que o pagamento de um pedágio é automatizado, neste cenário, pode ocorrer, mesmo que acidentalmente, a não autorização do pagamento do pedágio o que acarreta na negação do prosseguimento (ação) da viagem do usuário.

Serviços de Segurança: desenvolver serviços de segurança é o passo natural que ocorre após o levantamento dos objetivos e ameaças de segurança. Serviços de segurança são proteções comumente empregadas para alcançar confidencialidade, disponibilidade e integridade ao sistema. Embora, as ameaças à segurança não possam ser eliminadas por uma única ferramenta ou serviço, elas podem ser prevenidas ou mitigadas através da aplicação de serviços de segurança.

Alguns serviços de segurança são listados e descritos a seguir:

- **Serviço de autenticação:** este serviço é um meio de verificar a identidade das entidades que se relacionam com o sistema. Tipicamente a própria identidade se identifica para o sistema.
- **Serviço de integridade:** este serviço dá suporte às análises sobre a integridade da informação que flui através do sistema e visa minimizar a manipulação de informações. Exemplos de serviços de integridade são a detecção e correção de erros.
- **Serviço de controle de acesso:** visa prover permissões distintas para cada entidade (usuário, processos, gerentes, etc.) dada a função que exerce no sistema. Geralmente o serviço de controle de acesso é executado após a autenticação do usuário e assim são aplicadas as regras que limitam o acesso às informações do sistema. Este serviço visa reduzir a exposição de informações e recursos (*disclosure*), alterações de informações, e negação de serviços.
- **Serviço de auditoria:** é utilizado para rastrear as atividades dos usuários do sistema. Exemplos de auditoria são registros de entrada e saída do sistema, acesso a recursos,

reconfiguração. Geralmente é realizado através de um sistema de julgamento sobre o registro de ações realizadas. Esse serviço deve está livre de modificações e acessos não autorizadas devido a sua importância. Esse serviço também pode indicar a responsabilidade (*accountability*) de uma ação ofensiva contra o sistema, a responsabilidade de uma ação pode também ser considerada como um objetivo de segurança.

2.5.8. Cidades Inteligentes e Sustentáveis

Com a evolução das tecnologias, das redes veiculares e também da comunicação de redes metropolitanas, torna-se possível criar ambientes nos quais os veículos interajam com o eles e também são influenciados. Neste contexto, é possível monitorar toda a trajetória dos veículos, a densidade em cada região da cidade e também a evolução do tráfego ao longo do dia, reagindo de acordo com a demanda e eventos da cidade. Por exemplo, de acordo com a densidade de veículos em um lugar, os semáforos podem ser sincronizados de forma a evitar o engarrafamento e possíveis interrupções nas nessas vias [Barba et al. 2012].

As cidades inteligentes podem incluir serviços para coordenar os semáforos, o estacionamento, os serviços de localização do local, os serviços meteorológicos, os serviços turísticos e os serviços de emergência. Todos os serviços deve ser integrados para melhorar a precisão das informações entregues ao driver [Nam et al. 2011]. Um grande desafio neste cenário é a integração de todos os serviços, para que a cidade se torne inteligente. Para isso, é importante padronizar os protocolos de comunicação entre veículo-veículo e veículo-a-infraestrutura, de forma a garantir a conectividade entre eles independente de marca ou modelo. Além disso, é essencial a cooperação entre as redes veiculares, outras redes, e dispositivos computacionais na tarefa para coletar dados de ambiente e para melhorar os serviços prestados aos cidadãos.

Cidades inteligentes também podem trabalhar para fornecer ruas com sensores que rastreia e alertam o condutor de perigos à frente. Além disso, veículos autônomos podem fazer uso destes sensores para se guiarem e conduzirem os passageiros até o seu destino. Em [Kumar et al. 2012], os autores apresentam um *framework* que permitir a comunicação entre os sensores e veículos e viabiliza o deslocamento do carro entre a origem e o destino com segurança.

Um ponto que ganha destaque nas cidades inteligentes é a preocupação com o clima do planeta e as ações relacionadas que os seres humanos têm feito para reduzir o impacto nele. Muitos serviços e aplicações de rotas são providos à população para favorecer a locomoção de pessoas, reduzindo o consumo de combustível e a emissão de gás carbônico pelos veículos. Além disso, serviços de compartilhamento de rotas e caronas também tem recebido maior adesão [Zhu et al. 2013], pois reduzem a quantidade de veículos em circulação e por consequência a emissão de gases na atmosfera.

Em outra direção, surgem os veículos elétricos, que garante o deslocamento das pessoas além de reduzir a poluição do ar. Normalmente, os veículos elétricos são pequenos e possuem uma bateria que fornece energia para todo o veículo. No entanto, esta bateria é limitada e a sua recarga deve ser planeada com o objetivo a não comprometer a trajetória do automóvel. Neste cenário surgem a demanda de serviços que promovam uma rota mais curta, com poucos engarrafamentos e com pontos de recarga ao longo do caminho, quando

necessário [Qin and Zhang 2011].

2.6. Conclusões

Neste minicurso foi apresentado os principais conceitos relacionados a sistemas de transporte inteligentes. Questões relacionadas as arquiteturas existentes, padrões de comunicação em redes veiculares e integração dos sistemas com diferentes tipos de comunicação foram apontadas e discutidas, mostrando a necessidade da padronização e integração desses sistemas.

Além disso, discutimos os principais tipos de aplicações existentes em ITS, de forma a mostrar os trabalhos encontrados na literatura que já empregam esses conceitos e deixar algumas direções de novos trabalhos. Em seguida, de forma a clarear e ajudar os pesquisadores que estão iniciando na área, apresentamos uma seção que discute as principais ferramentas e simuladores usados para avaliar e projetar soluções em ITS.

Ao final, apresentamos uma discussão dos principais tópicos de pesquisa atual e também dos desafios que são encontrados em ITS com o objetivo de nortear futuras pesquisas na área. Acreditamos que existem novos desafios podem surgir na medida que esses sistemas evoluem e com a adesão de novos usuários.

Referências

[BMW] Bmw intelligent parking. `http://www.bmw.com/com/en/insights/technology/connecteddrive/2013/driver_assistance/intelligent_parking.html`. Acessado em: 21/03/2017.

[Bos] Bosch automatic park assist. `http://www.bosch.com/en/com/boschglobal/automated_driving/technology_for_greater_safety/pagination_1.html`. Acessado em: 21/03/2017.

[Vei] Veins ivc simulator. `veins.car2x.org`. Acessado em: 21/03/2017.

[IEE 2010] (2010). Ieee standard for information technology– local and metropolitan area networks– specific requirements– part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications amendment 6: Wireless access in vehicular environments. *IEEE Std 802.11p-2010*, pages 1–51.

[WAV 2011] (2011). Ieee standard for wireless access in vehicular environments (wave)–multi-channel operation. *IEEE Std 1609.4-2010 (Revision of IEEE Std 1609.4-2006)*, pages 1–89.

[Alves et al. 2009] Alves, R. S., do V. Campbell, I., de S. Couto, R., Campista, M. E. M., Moraes, I. M., Rubinstein, M. G., Costa, L. H. M. K., Duarte, O. C. M. B., and Abdalla, M. (2009). Redes veiculares: Princípios, aplicações e desafios. In *Minicursos do Simpósio Brasileiro de Redes de Computadores*, pages 199–254, Recife-PE.

[Barba et al. 2012] Barba, C. T., Mateos, M. A., Soto, P. R., Mezher, A. M., and Igartua, M. A. (2012). Smart city for vanets using warning messages, traffic statistics and intelligent traffic lights. In *2012 IEEE Intelligent Vehicles Symposium*, pages 902–907.

[Bauza et al. 2010] Bauza, R., Gozalvez, J., and Sanchez-Soriano, J. (2010). Road traffic congestion detection through cooperative vehicle-to-vehicle communications. In *IEEE Local Computer Network Conference*, pages 606–612. IEEE.

[Biesecker et al. 1997] Biesecker, K., Foreman, E., Jones, K., and Staples, B. (1997). Intelligent transportation systems (its) information security analysis. Technical report, U.S. Department of Transportation, Federal Highway Administration.

[Boldrini et al. 2016] Boldrini, C., Bruno, R., and Conti, M. (2016). Characterising demand and usage patterns in a large station-based car sharing system. In *Computer Communications Workshops (INFOCOM WKSHPS), 2016 IEEE Conference on*, pages 572–577. IEEE.

[Boukerche et al. 2008] Boukerche, A., Oliveira, H. A. B. F., Nakamura, E. F., and Loureiro, A. A. F. (2008). Vehicular ad hoc networks: A new challenge for localization-based systems. *Computer Communications*, 31(12):2838–2849.

[Brennand et al. 2015] Brennand, C. A., de Souza, A. M., Maia, G., Boukerche, A., Ramos, H., Loureiro, A. A., and Villas, L. A. (2015). An intelligent transportation system for detection and control of congested roads in urban centers. In *2015 IEEE Symposium on Computers and Communication (ISCC)*, pages 663–668. IEEE.

[Campigotto et al. 2017] Campigotto, P., Rudloff, C., Leodolter, M., and Bauer, D. (2017). Personalized and situation-aware multimodal route recommendations: the favour algorithm. *IEEE Transactions on Intelligent Transportation Systems*, 18(1):92–102.

[Castanedo 2013] Castanedo, F. (2013). A review of data fusion techniques. *The Scientific World Journal*, 2013.

[CESVI 2012] CESVI (2012). Centro de experimentação e segurança viária.

[Chen et al. 2017] Chen, J., Li, Z., Jiang, H., Zhu, S., and Wang, W. (2017). Simulating the impacts of on-street vehicle parking on traffic operations on urban streets using cellular automation. *Physica A: Statistical Mechanics and its Applications*, 468:880 – 891.

[Chen et al. 2016] Chen, K., Tan, G., Lu, M., and Wu, J. (2016). Crsm: a practical crowdsourcing-based road surface monitoring system. *Wireless Networks*, 22(3):765–779.

[Chen et al. 2015] Chen, L., Zhang, D., Pan, G., Ma, X., Yang, D., Kushlev, K., Zhang, W., and Li, S. (2015). Bike sharing station placement leveraging heterogeneous urban open data. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '15, pages 571–575, New York, NY, USA. ACM.

[Chiasserini et al. 2005] Chiasserini, C., Fasolo, E., Furiato, R., Gaeta, R., Garetto, M., Gribaudo, M., Sereno, M., and Zanella, A. (2005). Smart broadcast of warning messages in vehicular ad hoc networks. In *Workshop Interno Progetto NEWCOM (NoE)*.

[Cintra 2013] Cintra, M. (2013). A crise do trânsito em são paulo e seus custos. *GV-executivo*, 12(2):58–61.

[da Cruz et al.] da Cruz, S. M. S., Andrade, L. S., and Sampaio, J. O. Explorando dados abertos governamentais sobre as mobilidade urbana na cidade do rio de janeiro.

[Elbery et al. 2013] Elbery, A., ElNainay, M., Chen, F., Lu, C.-T., and Kendall, J. (2013). A carpooling recommendation system based on social vanet and geo-social data. In *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, SIGSPATIAL'13, pages 556–559, New York, NY, USA. ACM.

[Faezipour et al. 2012] Faezipour, M., Nourani, M., Saeed, A., and Addepalli, S. (2012). Progress and challenges in intelligent vehicle area networks. *Communications ACM*, 55(2):90–100.

[Fazio et al. 2013] Fazio, P., de Rango, F., and Lupia, A. (2013). Vehicular networks and road safety: An application for emergency/danger situations management using the wave/802.11 p standard. *Advances in Electrical and Electronic Engineering*, 11(5):357.

[Fleming 2001] Fleming, W. J. (2001). Overview of automotive sensors. *IEEE Sensors Journal*, 1(4):296–308.

[Ganti et al. 2010] Ganti, R. K., Pham, N., Ahmadi, H., Nangia, S., and Abdelzaher, T. F. (2010). Greengps: A participatory sensing fuel-efficient maps application. In *Proceedings of the 8th international conference on Mobile systems, applications, and services*, MobiSys '10, pages 151–164, New York, NY, USA. ACM.

[Gerla and Kleinrock 2011] Gerla, M. and Kleinrock, L. (2011). Vehicular networks and the future of the mobile internet. *Computer Networks*, 55(2):457 – 469. Wireless for the Future Internet.

[Gozálvez et al. 2012] Gozálvez, J., Sepulcre, M., and Bauza, R. (2012). Ieee 802.11 p vehicle to infrastructure communications in urban environments. *IEEE Communications Magazine*, 50(5).

[Hameed Mir and Filali 2014] Hameed Mir, Z. and Filali, F. (2014). Lte and ieee 802.11p for vehicular networking: a performance evaluation. *EURASIP Journal on Wireless Communications and Networking*, 2014(1):89.

[Hartenstein and Laberteaux 2008] Hartenstein, H. and Laberteaux, K. (2008). A tutorial survey on vehicular ad hoc networks. *Communications Magazine, IEEE*, 46(6):164–171.

[He et al. 2014] He, W., Yan, G., and Da Xu, L. (2014). Developing vehicular data cloud services in the iot environment. *IEEE Transactions on Industrial Informatics*, 10(2):1587–1595.

[Heinovski et al. 2016] Heinovski, J., Klingler, F., Dressler, F., and Sommer, C. (2016). Performance comparison of ieee 802.11p and arib std-t109. In *2016 IEEE Vehicular Networking Conference (VNC)*, pages 1–8.

[Huang and Wang 2016] Huang, W. and Wang, L. (2016). Ecds: Efficient collaborative downloading scheme for popular content distribution in urban vehicular networks. *Computer Networks*, 101:90 – 103. Industrial Technologies and Applications for the Internet of Things.

[Hussain et al. 2012] Hussain, R., Son, J., Eun, H., Kim, S., and Oh, H. (2012). Rethinking vehicular communications: Merging vanet with cloud computing. In *Cloud Computing Technology and Science (CloudCom), 2012 IEEE 4th International Conference on*, pages 606–609. IEEE.

[IPEA 2012] IPEA (2012). Instituto brasileiro de pesquisas econômicas.

[ISO 21217:2010 2010] ISO 21217:2010 (2010). Intelligent transport systems — communications access for land mobiles (CALM) — architecture. ISO 21217:2010, ISO TC204, Geneva, Switzerland.

[Jeong et al. 2010] Jeong, J., Guo, S., Gu, Y., He, T., and Du, D. H. (2010). Tsf: Trajectory-based statistical forwarding for infrastructure-to-vehicle data delivery in vehicular networks. In *Distributed Computing Systems (ICDCS), 2010 IEEE 30th International Conference On*, pages 557–566. IEEE.

[Jiang et al. 2008] Jiang, D., Chen, Q., and Delgrossi, L. (2008). Optimal data rate selection for vehicle safety communications. In *Proceedings of the Fifth ACM International Workshop on VehiculAr Inter-NETworking, ACM Conference on, (VANET'08)*, pages 30–38.

[Jiang and Delgrossi 2008] Jiang, D. and Delgrossi, L. (2008). Ieee 802.11p: Towards an international standard for wireless access in vehicular environments. In *Proceedings of the Vehicular Technology Conference, IEEE Conference on (VTC Spring'08)*, pages 2036–2040.

[Karagiannis et al. 2011] Karagiannis, G., Altintas, O., Ekici, E., Heijenk, G., Jarupan, B., Lin, K., and Weil, T. (2011). Vehicular networking: A survey and tutorial on requirements, architectures, challenges, standards and solutions. *Commun. Surveys Tutorials, IEEE*, 13(4):584–616.

[Kenney 2011] Kenney, J. (2011). Dedicated short-range communications (dsrc) standards in the united states. *Proceedings of the IEEE*, 99(7):1162–1182.

[Kumar et al. 2012] Kumar, S., Shi, L., Ahmed, N., Gil, S., Katabi, D., and Rus, D. (2012). Carspeak: a content-centric network for autonomous driving. *SIGCOMM Comput. Commun. Rev.*, 42(4):259–270.

[Lee et al. 2006] Lee, U., Park, J.-S., Yeh, J., Pau, G., and Gerla, M. (2006). Code torrent: Content distribution using network coding in vanet. In *Proceedings of the 1st International Workshop on Decentralized Resource Sharing in Mobile Computing and Networking*, MobiShare '06, pages 1–5, New York, NY, USA. ACM.

[Lequerica et al. 2010] Lequerica, I., Longaron, M. G., and Ruiz, P. M. (2010). Drive and share: efficient provisioning of social networks in vehicular scenarios. *IEEE Communications Magazine*, 48(11):90–97.

[Levy-Bencheton and Darra 2015] Levy-Bencheton, C. and Darra, E. (2015). Cyber security and resilience of intelligent public transport: good practices and recommendations. Technical report, European Union Agency For Network And Information Security (ENISA).

[Li and Wang 2007] Li, F. and Wang, Y. (2007). Routing in vehicular ad hoc networks: A survey. *Vehicular Technology Magazine, IEEE*, 2(2):12 –22.

[Lin et al. 2009] Lin, J., Chen, S., Shih, Y., and Chen, S.-h. (2009). A study on remote on-line diagnostic system for vehicles by integrating the technology of obd, gps, and 3g. *World Academy of Science, Engineering and Technology*, 32(8):435–441.

[Lorch et al. 2006] Lorch, J. R., Adya, A., Bolosky, W. J., Chaiken, R., Douceur, J. R., and Howell, J. (2006). The smart way to migrate replicated stateful services. *SIGOPS Oper. Syst. Rev.*, 40(4):103–115.

[Massaro et al. 2017] Massaro, E., Ahn, C., Ratti, C., Santi, P., Stahlmann, R., Lamprecht, A., Roehder, M., and Huber, M. (2017). The car as an ambient sensing platform [point of view]. *Proceedings of the IEEE*, 105(1):3–7.

[Monteiro de Lira et al. 2016] Monteiro de Lira, V., Renso, C., Perego, R., Rinzivillo, S., and Cesario Times, V. (2016). The comewithme system for searching and ranking activity-based carpooling rides. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '16, pages 1145–1148, New York, NY, USA. ACM.

[Nair et al. 2013] Nair, R., Miller-Hooks, E., Hampshire, R. C., and Bušić, A. (2013). Large-scale vehicle sharing systems: analysis of vélib'. *International Journal of Sustainable Transportation*, 7(1):85–106.

[Nam et al. 2011] Nam, T., Aldama, F. A., Chourabi, H., Mellouli, S., Pardo, T. A., Gil-Garcia, J. R., Scholl, H. J., Ojo, A., Estevez, E., and Zheng, L. (2011). Smart cities and service integration. In *Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times*, dg.o '11, pages 333–334, New York, NY, USA. ACM.

[of Transport 2016] of Transport, U. S. D. (2016). National its architecture. `http://www.iteris.com/itsarch/index.htm/`. [Online; acessado Dez-2016].

[of Transportation 2015] of Transportation, U. D. (2015). Traffic congestion and reliability: trends and advanced strategies for congestion mitigation.

[Olariu et al. 2011] Olariu, S., Khalil, I., and Abuelela, M. (2011). Taking vanet to the clouds. *International Journal of Pervasive Computing and Communications*, 7(1):7–21.

[Pan et al. 2013] Pan, B., Zheng, Y., Wilkie, D., and Shahabi, C. (2013). Crowd sensing of traffic anomalies based on human mobility and social media. In *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 344–353. ACM.

[Pan et al. 2012] Pan, J., Khan, M. A., Popa, I. S., Zeitouni, K., and Borcea, C. (2012). Proactive Vehicle Re-routing Strategies for Congestion Avoidance. In *2012 IEEE 8th International Conference on Distributed Computing in Sensor Systems*, pages 265–272.

[Paromtchik and Laugier 1996] Paromtchik, I. E. and Laugier, C. (1996). Motion generation and control for parking an autonomous vehicle. In *Proceedings of IEEE International Conference on Robotics and Automation*, volume 4, pages 3117–3122 vol.4.

[Peng and Li 2016] Peng, L. and Li, H. (2016). Searching parking spaces in urban environments based on non-stationary poisson process analysis. In *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, pages 1951–1956.

[Peng et al. 2006] Peng, Y., Abichar, Z., and Chang, J. M. (2006). Roadside-aided routing (rar) in vehicular networks. In *Communications, 2006. ICC'06. IEEE International Conference on*, volume 8, pages 3602–3607. IEEE.

[Plößl and Federrath 2008] Plößl, K. and Federrath, H. (2008). A privacy aware and efficient security infrastructure for vehicular ad hoc networks. *Computer Standards and Interfaces*, 30(6):390 – 397. Special Issue: State of standards in the information systems security area.

[Qin and Zhang 2011] Qin, H. and Zhang, W. (2011). Charging scheduling with minimal waiting in a network of electric vehicles and charging stations. In *Proceedings of the Eighth ACM international workshop on Vehicular inter-networking*, VANET '11, pages 51–60, New York, NY, USA. ACM.

[Qu et al. 2010] Qu, F., Wang, F. Y., and Yang, L. (2010). Intelligent transportation spaces: Vehicles, traffic, communications, and beyond. *IEEE Communications Magazine*, 48(11):136–142.

[Rajabioun and Ioannou 2015] Rajabioun, T. and Ioannou, P. A. (2015). On-street and off-street parking availability prediction using multivariate spatiotemporal models. *IEEE Transactions on Intelligent Transportation Systems*, 16(5):2913–2924.

[Rescorla and Korver 2003] Rescorla, E. and Korver, B. (2003). Guidelines for writing rfc text on security considerations. Technical report, IETF.

[Rettore et al. 2016] Rettore, P. H., André, B. P. S., Campolina, Villas, L. A., and A.F. Loureiro, A. (2016). Towards intra-vehicular sensor data fusion. In *Advanced perception, Machine learning and Data sets (AMD'16) as part of the 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC 2016)*, , Rio de Janeiro.

[Shirazi and Morris 2017] Shirazi, M. S. and Morris, B. T. (2017). Looking at intersections: A survey of intersection monitoring, behavior and safety analysis of recent studies. *IEEE Transactions on Intelligent Transportation Systems*, 18(1):4–24.

[Silva et al. 2014] Silva, T. H., De Melo, P. O. V., Almeida, J. M., and Loureiro, A. A. (2014). Large-scale study of city dynamics and urban social behavior using participatory sensing. *IEEE Wireless Communications*, 21(1):42–51.

[Sommer et al. 2012] Sommer, C., Joerer, S., and Dressler, F. (2012). On the applicability of two-ray path loss models for vehicular network simulation. In *IEEE Vehicular Networking Conference (VNC '12)*, pages 64–69.

[Studer et al. 2009] Studer, A., Shi, E., Bai, F., and Perrig, A. (2009). Tacking together efficient authentication, revocation, and privacy in vanets. In *Sensor, Mesh and Ad Hoc Communications and Networks, 2009. SECON'09. 6th Annual IEEE Communications Society Conference on*, pages 1–9. IEEE.

[Tasseron and Martens 2017] Tasseron, G. and Martens, K. (2017). Urban parking space reservation through bottom-up information provision: An agent-based analysis. *Computers, Environment and Urban Systems*, 64:30 – 41.

[Tasseron et al. 2016] Tasseron, G., Martens, K., and van der Heijden, R. (2016). The potential impact of vehicle-to-vehicle communication on on-street parking under heterogeneous conditions. *IEEE Intelligent Transportation Systems Magazine*, 8(2):33–42.

[Tian et al. 2017] Tian, B., Morris, B. T., Tang, M., Liu, Y., Yao, Y., Gou, C., Shen, D., and Tang, S. (2017). Hierarchical and networked vehicle surveillance in its: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 18(1):25–48.

[Timpner et al. 2016] Timpner, J., Schürmann, D., and Wolf, L. (2016). Trustworthy parking communities: Helping your neighbor to find a space. *IEEE Transactions on Dependable and Secure Computing*, 13(1):120–132.

[Tostes et al. 2013] Tostes, A. I. J., de LP Duarte-Figueiredo, F., Assunção, R., Salles, J., and Loureiro, A. A. (2013). From data to knowledge: city-wide traffic flows analysis and prediction using bing maps. In *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*, page 12. ACM.

[Trullols et al. 2010] Trullols, O., Fiore, M., Casetti, C., Chiasserini, C., and Ordinas, J. B. (2010). Planning roadside infrastructure for information dissemination in intelligent transportation systems. *Computer Communications*, 33(4):432 – 442.

[Tseng et al. 2010] Tseng, Y.-T., Jan, R.-H., Chen, C., Wang, C.-F., and Li, H.-H. (2010). A vehicle-density-based forwarding scheme for emergency message broadcasts in vanets. In *Mobile Adhoc and Sensor Systems (MASS), 2010 IEEE 7th International Conference on*, pages 703–708. IEEE.

[Uzcategui and Acosta-Marum 2009] Uzcategui, R. and Acosta-Marum, G. (2009). Wave: A tutorial. *Communications Magazine, IEEE*, 47(5):126–133.

[Wu et al. 2014] Wu, E. H. K., Sahoo, J., Liu, C. Y., Jin, M. H., and Lin, S. H. (2014). Agile urban parking recommendation service for intelligent vehicular guiding system. *IEEE Intelligent Transportation Systems Magazine*, 6(1):35–49.

[Yang et al. 2004] Yang, X., Liu, J., Zhao, F., and Vaidya, N. (2004). A vehicle-to-vehicle communication protocol for cooperative collision warning. In *First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services (MobiQuitous'04)*, pages 114–123.

[Yang et al. 2016] Yang, Z., Hu, J., Shu, Y., Cheng, P., Chen, J., and Moscibroda, T. (2016). Mobility modeling and prediction in bike-sharing systems. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*, pages 165–178. ACM.

[Ye and Xu 2017] Ye, Z. and Xu, M. (2017). Decision model for resolving conflicting transit signal priority requests. *IEEE Transactions on Intelligent Transportation Systems*, 18(1):59–68.

[Yousefi et al. 2006] Yousefi, S., Mousavi, M. S., and Fathy, M. (2006). Vehicular ad hoc networks (vanets): Challenges and perspectives. In *ITS Telecommunications Proceedings, 2006 6th International Conference on*, pages 761 –766.

[Zaldivar et al. 2011] Zaldivar, J., Calafate, C. T., Cano, J. C., and Manzoni, P. (2011). Providing accident detection in vehicular networks through obd-ii devices and android-based smartphones. In *Local Computer Networks (LCN), 2011 IEEE 36th Conference on*, pages 813–819. IEEE.

[Zhu et al. 2013] Zhu, J., Feng, Y., and Liu, B. (2013). Pass: Parking-lot-assisted carpool over vehicular ad hoc networks. *International Journal of Distributed Sensor Networks*, 2013:1–9.

[Zuchao Wang et al. 2013] Zuchao Wang, Min Lu, Xiaoru Yuan, Junping Zhang, and Van De Wetering, H. (2013). Visual traffic jam analysis based on trajectory data. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2159–2168.

Chapter 3

Experimental Wireless Networking Research using Software-Defined Radios

Adriele Dutra Souza (UFV), Ariel F. F. Marques (UFLA), Daniel F. Macedo (UFMG), Diarmuid Collins (Trinity College Dublin), Gilson Miranda Júnior (UFLA), Jefferson R. S. Cordeiro (UFMG), Johann M. Marquez-Barja (Trinity College Dublin), José Augusto M. Nacif (UFV), Kristtopher Kayo Coelho (UFV), Luccas R. M. Pinto (UFLA), Luiz A. da Silva (Trinity College Dublin), Luiz F. M. Vieira (UFMG), Luiz H. A. Correia (UFLA), Marcos A. M. Vieira (UFMG), Pedro Alvarez (Trinity College Dublin), Wendley S. Silva (UFMG)

Abstract

Thanks to the popularization of software-defined radios (SDR), it is possible today to perform high-quality research in wireless protocols in real deployments. Although this technology is still a bit expensive, there are a number of initiatives that provide free access to SDR for research. Further, the number of free software libraries available for SDR has reduced the amount of effort required to conduct research using SDR. This short course will show by examples how to perform experimental research in wireless networking using software-defined radios that are available for free on open testbeds being developed on FUTEBOL, a joint Brazil-European Union project. We will adopt a hands-on approach, in which the students will perform many small assignments on real hardware. Those assignments will demonstrate the maturity of SDR for research in wireless networking, and introduce the user to the many software tools and open source implementations of a variety of wireless standards.

3.1. Introduction

Software-Defined Radios (SDR) are a collection of hardware and software technologies where some or all of the radio's operating functions are implemented through modifiable software or firmware operating on programmable processing technologies (e.g. an FPGA, a generic CPU) [Wireless Innovation Forum 2017].

The US military was the first to employ SDRs, in order to provide flexible radios for large-scale operations [Dillinger et al. 2003]. The objective was to ensure the

interoperability of military radios among government agencies (firefighters, police, intelligence agencies). Such a need comes from a practical reason since each agency performs its purchases independently, and thus the communication technologies employed may be incompatible from the point of view of frequency used, signal modulation technology, among others. Thus, the term "digital radio" was coined to define radios that adapt to different operating standards.

In 2011, the Wireless Innovation Forum commissioned a study to evaluate the rate of adoption of SDR in the telecommunications industry [Forum 2011]. The results indicate that more than 90% of the mobile infrastructure on that year employed SDR technologies of some kind. For markets where interoperability is a mandatory requirement, as in military and public safety applications, they have found that almost all transceivers and base stations employ SDR.

Software-defined radios are now a reality, as seen by the number of commercial and free platforms available in the market[1]. Given the viable commercial and academic platforms that are capable of implementing 3G and IEEE 802.11a/b/g/n radios and other technologies, several research groups have purchased SDR equipment for their research. SDR has lowered the cost of conducting experimental research on the physical layer and link layer. Thus, experiments that were previously possible only in laboratories of large companies can be done in laboratories of universities at a reasonable cost. To showcase the importance to the SDR in the area of wireless networks, only the WARP platform at Rice University counted 43 scientific papers that used their hardware in 2016[2].

However, as we will show in this chapter, SDR is a fairly accessible technology for experimental research in wireless networks. There are many software libraries available, which awesome allow SDR boards to run popular wireless standards. An SDR can be used to emulate from big to small devices (from RFID tags to a 4G eNodeB), from simple to complex communication standards (from a remote garage controller to a digital TV transmission). Further, initiatives such as GENI[3] in the US, Fed4Fire[4] in Europe, and FUTEBOL[5] in Brazil, provide free access to SDR hardware for research purposes.

This chapter will focus on the kinds of experiments that can be performed in USRP boards [Ettus 2017], as well as how to setup and run an experiment on the SDR testbeds made available by the FUTEBOL project. At the end of this chapter, the reader should be able to understand the limitations of SDR, how to choose the type of SDR for his/her experiment, as well as how to perform an experiment remotely on the FUTEBOL testbed. Readers that want to learn about the theoretical aspects of SDR, as well as the basics of how to program SDR using GNU Radio, should refer to [Silva et al. 2015].

[1]See a list of some of the existing platforms: `https://en.wikipedia.org/wiki/List_of_software-defined_radios`

[2]`http://warpproject.org/trac/wiki/PapersandPresentations#LatestPapers`

[3]`http://geni.net/`

[4]`http://fed4fire.eu/`

[5]`http://www.ict-futebol.org.br`

3.1.1. Existing research using SDRs

SDR technologies enable different types of research that previously were mostly performed using simulations and/or analytical methods. Since SDR is a versatile hardware, the same board can be used over and over again for different wireless projects, becoming a must-have for groups that perform systems research in wireless networks. Below we present a list of recent advances in wireless communications that employed SDR as an evaluation platform:

- **Using network coding to increase wireless capacity.** With network coding, it is possible to send a packet that is the combination of several packets into wireless links, increasing the overall flow of the network. Earnings depend on the traffic pattern, ranging from a small percentage up to several times [Katti et al. 2008, Vieira et al. 2013].

- **Recover lost frames using signal processing.** Radios store incoming signals from a collision, and attempt to subtract received frames correctly (after a retransmission), often allowing the frame to be involved in a collision without it having to be retransmitted [Lin et al. 2008].

- **Rateless codes** In traditional wireless communication, data is transmitted using a certain modulation. Each modulation requires a certain Signal to Interference plus Noise Ratio (SINR) threshold for proper decoding of its data, and in order to deal with variations in SINR, existing protocols (e.g. Wi-Fi and WiMax protocols) use automatic modulation change mechanisms. Rateless codes [Gudipati and Katti 2011, Perry et al. 2012, Shokrollahi 2006] use a variable amount of symbols to encode the data. The transmitter sends the data using different codes, and the receiver uses those codes to identify which word is most likely to have been used to generate the received signals. Rateless codes require special protocols of the [Iannucci et al. 2012] binding layer, which can also be tested using SDR. The main benefit of rateless codes is that it provides bandwidth at the link much closer to Shannon's theoretical capacity.

- **Full-duplex radios.** Existing commercial radios are half-duplex since one receiving antenna would be saturated with signals transmitted by another adjacent antenna. However, research employing dedicated hardware and SDRs can achieve such levels of noise cancellation as to allow full-duplex radios to be compliant with IEEE 802.11b [Hong et al. 2012] and IEEE 802.11ac [Bharadia et al. 2013].

- **Cooperative MIMO for enterprise WLANs.** OpenRF [Kumar et al. 2013] uses the concept of coding vectors in order to perform beamforming, which cancels out the interference of neighboring APs. With the help of a centralized controller and a local scheduling algorithm running on each access point, it is possible to exploit the MIMO capabilities of the IEEE 802.11n standard to improve the SINR in the stations. This, in turn, reduces latency variations and increases network throughput.

Although the most common use of SDR is in telecommunications research, it can also be used by networking researchers. Nowadays it is possible to find full protocol

stacks that are either open source or available freely for research. This is allowing SDR to be used on networking-related papers, dealing with issues such as MAC layer design, management of wireless networks, performance evaluation of 4G networks, security, etc. Here are some examples.

- **Cloud RAN**. USRPs are being used to emulate Cloud RAN deployments [Beyene et al. 2014]. Aspects such as where to run the physical and MAC layers (near the eNodeBs or in data centers) and how to control the RAN system can be studied using real deployments.
- **Cognitive Radios**. Research on cognitive radios requires equipment that is able to change its operating frequency and survey the usage of the spectrum by employing different algorithms to detect the existence and types of transmissions on the medium. One such study is presented by Souryal et al., which measured the usage of the spectrum using USRPs as their platform [Souryal et al. 2015].
- **New MAC protocols**. SDR can be used to change parameters of the MAC layer or even implement a completely new protocol. Usually, the MAC protocol is implemented in the firmware of the wireless transceivers, and as such, it is very hard to change it. Thus, without SDR such types of research would probably be performed using simulators. However, this task is feasible in SDRs. One example is LA-MAC, a load-aware MAC protocol that was tested using USRPs [Hu et al. 2009].
- **Investigating the security of wireless protocols**. SDR can be used to capture traffic and then analyze it to identify vulnerabilities in wireless protocols. It can also be used to build proof-of-concept attacks against those protocols, as well as to propose defenses. Gollakota et al. employed USRPs to propose a device called *shield*, which prevents outsiders from eavesdropping the messages of implantable medical devices (IMDs) [Gollakota et al. 2011].

3.2. Getting your hands dirty

This section presents the Universal Software Radio Peripheral boards, which are the most popular SDR platform among researchers. Alternative platforms will also be presented, as well as the practical issues that arise when using SDR to perform experimental research. Next, we will describe the lifecycle of an experiment in remote testbeds.

3.2.1. An introduction to USRP boards

Universal Software Radio Peripheral (USRP) is a framework for the development of digital radios, providing a complete infrastructure for signal processing. The system is characterized by its high flexibility and cost-benefit. USRPs are developed by Ettus Inc [Ettus 2017], which is a subsidiary of National Instruments.

USRPs are an attractive platform for SDR research for many reasons. First, Ettus open sourced schematics for some of the USRP models, and the driver that allows the communication of the boards with a computer is also open source. The USRP hardware driver (UHD) is compatible with many operating systems, such as Windows, Linux, and

Mac. Second, USRPs are compatible with GNU Radio [Gilmore and Blossom 2017], which is a GNU library of software that implements several algorithms related to signal processing and communications. Further, there is a large community of people using USRPs for research or as a hobby, and as such it is relatively easy to find support online. Finally, popular scientific software such as MatLab and others support USRPs.

USRPs are composed of an FPGA, components for baseband processing, and daughterboards. In general, the hardware processes the RF signals, converting them into digital signals to be processed either at the FPGA or at a host computer. The USRP communicate with a PC using a high-speed bus, which may be a USB interface or a network interface. The daughterboards are interchangeable cards that provide the filters and amplifiers that are necessary to support a certain application, that is, a certain range of frequencies. Some of the models do not support daughterboards, and as such, they have a fixed range of frequencies.

There are a number of USRP models, with varying interfaces and capabilities. The Networked series connects to the PC using Ethernet. The bus-based series connects to the PC using USB. This series has a "mini" line, which are USRPs with a small form factor (at the moment of the writing, the size of a business card). The X series products are the higher end products, with more capable FPGAs as well as higher sampling rates. Finally, the Embedded series provides a more rugged product that is coupled with an ARM processor. It is ideal for operations on the field.

USRPs are supported by a number open source software libraries or initiatives. There are open source implementations of many communication standards provided as out-of-tree GNU Radio components (that is, they are not officially supported by GNU Radio), such as IEEE 802.11, IEEE 802.15.4, RFID. It is worth noticing that USRPs can be used today even to implement a full mobile network using open source telecommunication stacks such as OpenBTS [ope 2017] and OpenAirInterface [OAI 2017].

3.2.1.1. Alternatives to USRP

Although USRPs are very popular SDR platforms, other platforms can also be used in research experiments. This section lists those that are most commonly found in universities and research centers around the world.

- **SORA** stands for *Microsoft Research Software Radio* [Tan et al. 2009], and is an SDR platform developed by Microsoft Research (MSR) Asia in Beijing. The SORA hardware is very simple, having only a baseband decoder, and all processing is done by an x86 CPU. Due to that architecture, the platform has very stringent bandwidth requirements and developers must optimize their implementations using assembly for better performance.

- **WARP** is a research-oriented SDR that is also used in many testbeds [Murphy et al. 2006, Amiri et al. 2007]. Its hardware is more capable than most USRPs since it is able to decode up to 40MHz channels. Further, it has libraries implementing high-speed communication standards such as IEEE 802.11. As in the USRPs, it has an

FPGA that can be used for time-critical parts of the code. However, the hardware is costlier than USRPs.

- **Soft-MAC platforms**. Many platforms allow only the MAC layer to be modified [Tinnirello et al. 2012, Neufeld et al. 2005, Nychis et al. 2009, Rao and Stoica 2005]. The benefits are a lower cost of the devices, as well as the use of simpler programming languages. Some of those platforms even run on commodity wireless interfaces, since they are essentially a firmware update of a commercial interface. However, the kind of programmability that is allowed is limited. For example, in FLAVIA [Tinnirello et al. 2012] the only actions allowed are those defined by the developers of the platform. As such, it would not be possible for example to test a transmission power control protocol on such platforms.

- **RTL-SDR**. The chip RTL2832U is employed in many USB-based digital TV decoders, and it is in effect an SDR [RTL-SDR]. This is a very cheap SDR platform, which can be found for less than 40 dollars. However, its capabilities depend largely on the USB stick's manufacturers. Most are able only to receive signals, and the range of allowed frequencies is usually limited to those of digital TV. Although RTL-SDR is aimed at hobbyist, it could be used for very simple research projects or for teaching purposes. The advantage is that the software specific for RTL-SDR is usually very simple to use.

3.2.1.2. Limitations of SDR

Although SDR is very flexible, there are limitations associated with its generic hardware and its high CPU demands. This section will discuss those limitations so that experimenters understand when not to use SDR to conduct research, and what types of issues must be taken into account in order to select the proper SDR platform and configuration.

High demand for computing and I/O resources. In order to implement high-speed networks, SDR requires fast I/O from the SDR board to the processing unit, as well as a fast processing unit. The amount of processing increases with the complexity of the modulation as well as the bandwidth of the channels. Due to that, many SDR platforms support FPGAs so that the time critical and high bandwidth operations are processed in hardware, near the transceiver.

Limitations due to the choice of filters and antennas. Every daughterboard has limitations as to the frequency range that it is able to operate. Likewise, omnidirectional antennas are tailored to a certain operating frequency. This is not that important when the experimenter has direct access to the equipment since he/she can change the antenna and the daughterboard. However, when using a testbed, the antenna, and the daughterboard are usually fixed. For example, when using an antenna optimized for 2.4Ghz signals in an experiment that employs 900MHz will reduce the quality of the communication. Further, some experiments will not be feasible, since the filter does not allow the SDR to process the required frequency. As an example, an USRP using the Ettus WBX daughter-

board[6] cannot be used to decode AM radios, because the WBX cannot pick up signals on frequencies lower than 50MHz.

Limited oscillator resolution. The resolution of the oscillator dictates the capability of the board to implement protocols that require more precise timings. For high-speed networks, for example, 3G and 4G networks, the devices must employ more precise oscillators (e.g a GPS-based oscillator). When the oscillator is not precise enough, the devices' clocks drift from one another, and as such the symbols will be decoded incorrectly.

Delays of the operating system and I/O buses. The delays incurred by the I/O subsystem as well as the operating system overhead must be taken into account when running high-speed networks. For example, in order to support IEEE 802.11g in SORA, Microsoft Research Asia changed the Windows scheduler so that a core is always dedicated to SDR processing [Tan et al. 2009]. OpenAirInterface recommends the use of low latency Linux kernels[7].

3.2.2. Using testbeds

The creation of experimental facilities for more realistic research is a concern of the networking community. Researchers have identified the need for experiments within realistic conditions in order to reduce the gap between experimentation and real network deployments. Due to that, the community has deployed a number of testbeds. The funding agencies have recognized that need as well, so that initiatives such as GENI in the USA and FIRE in Europe, and more recently the RNP calls of joint Brazil-Europe projects, allocate budget for projects that will create and maintain open testbeds. As a consequence, there are a number of testbeds around the world that are available for experimentation. Those testbeds provide resources such as virtual machines, sensor nodes, physical machines, software-defined networking switches, and routers, as well as USRPs.

The access to those devices is free of charge and is performed remotely. The conditions to request access as well as the steps required to setup and run an experiment vary from testbed to testbed, however they follow a set of steps which are described below:

1. The research team identifies the requirements of the experiment.

2. With the requirements, the team identifies which testbed, or which set of testbeds, could be used to perform this experiment.

3. The Principal Investigator (PI), which is the lead researcher on the team, creates a project within the testbeds. Usually, this requests involves a simple description of the experiments as well as a rough estimate on the type and number of devices employed, as well as the duration of the experiment. Other steps may be necessary, for example, if the experiment involves privacy and ethics issues, or if there is a need for a license to use a chunk of the spectrum. In this case, the researcher may be required to request approval from the appropriate university of government bodies.

[6] https://www.ettus.com/product/details/WBX

[7] https://gitlab.eurecom.fr/oai/openairinterface5g/wikis/OpenAirKernelMainSetup

4. Once the project is approved, the PI creates accounts for the other researchers in the experiment.

5. The researchers create *slices* on the testbed, which are composed of devices and links among those devices. The slices have a maximum lifetime, however, the slices can be destroyed and recreated several times during the lifetime of a project. The slices are described by a configuration file, which defines the type of devices to be used, their location as well as the software to be employed.

6. The slice is activated, and the researchers access the slice to install their tools and code so that the experiment can be run.

7. The experiment is run, and the researchers download the relevant data to be analyzed.

8. The slice is released, either because the experiment ended, or because the maximum time of the slice has expired.

For a more detailed description of the lifecycle of an experiment on a testbed as well as the software required to setup and run a testbed, please refer to [Gomez 2013]. From now on, in this chapter we will describe how to setup and run experiments in testbeds federated in Fed4Fire, more specifically using the resources provided by the FUTEBOL project. However, the process will be similar in other testbeds and for other types of resources.

3.2.3. Using FUTEBOL to conduct research

The EU-BR FUTEBOL project envisages the creation of a federated control framework to integrate testbeds from Europe and Brazil for network researchers from academia and industry. FUTEBOL's major goal is to allow the access to advanced experimental facilities in Europe and Brazil for research and education across the wireless and optical domains. To that end, the project will deploy testbed facilities in a number of European countries as well as in three sites in Brazil, as shown in Figure 3.1. FUTEBOL resources are available in the UFES, UFMG, and UFRGS in Brazil, and in VTT (Finland), IT Aveiro (Portugal), TCD Dublin (Ireland) and University of Bristol (England). Those resources are interconnected using FIBRENET in RNP, and Géant in Europe. Authorization and management of users credentials are performed using Fed4Fire, meaning that any researcher in academia or in the industry with a valid Fed4Fire will be able to access the FUTEBOL resources.

For the moment, USRPs are available in FUTEBOL in the Trinity College Dublin (TCD) as well as in the UFMG islands. However, in the near future, USRPs will also be available in UFRGS and in UFES. For more information on the types of USRPs available as well as their topology, please refer to the website of each island. Links are available at the main FUTEBOL website.

Creating an account. The first step to use FUTEBOL is to create an account and a project in Fed4Fire. The Principal Investigator (PI) should request an account, which can be created at `https://authority.ilabt.iminds.be/`. With an account

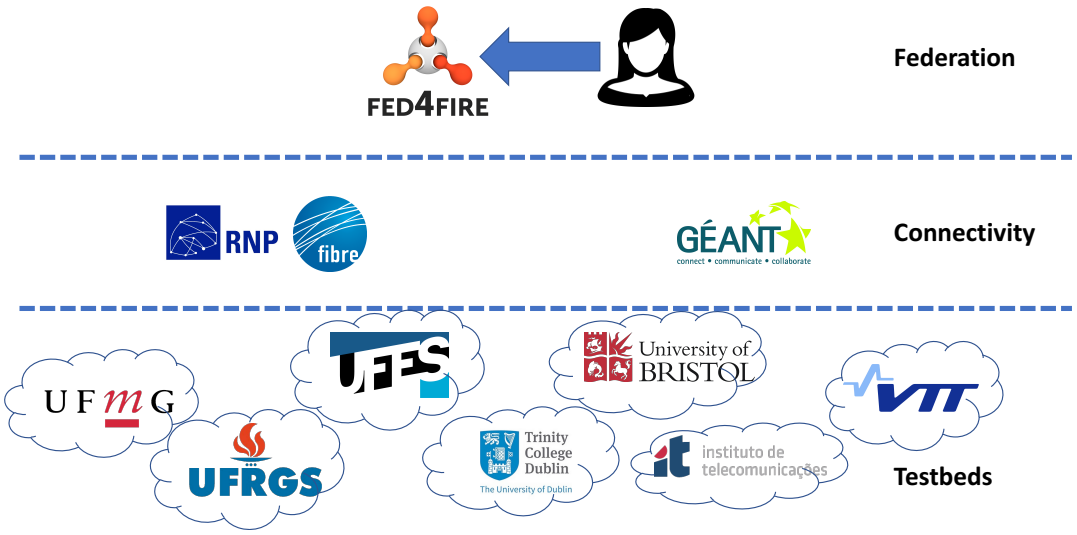


Figure 3.1: The overall architecture of the FUTEBOL federated testbed.

in hand, the PI can create a project that will use Fed4Fire. This is a simple form that requests a description of the project and a project name. Once this is done, the PI requests the researchers associated with the project to create accounts on Fed4Fire and associates them as members of the project.

Downloading your certificate. After creating a login and receiving an authorization message, the reader can download the certificate and store it locally on your local computer. Save the file with `.pem` extension. This certificate will be requested to access the remote environment through JFed.

Using JFed. Before using JFed, the reader needs to install Java 8 manually. This link provides more details about Java installation: `http://jfed.iminds.be/java8_on_linux/`. To install JFed, the user can use the `.deb` format, available at `http://jfed.iminds.be/downloads/`.

After the installation, we can run JFed and use the .pem certificate, choosing the option *Login with PEM-certificate*, as seen at Fig. 3.2.

The following sections describe experiments using USRPs and assume that the user knows how to create an experiment in the FUTEBOL testbed. This involves creating an account in Fed4Fire, using JFed or other tools to define the number of USRPs that the user needs, and then request the reservation of those resources. For a step-by-step description of those steps, please refer to `http://futebol.dcc.ufmg.br/tutorials.html`. This website provides a sample file that books one virtual machine with one USRP in the UFMG testbed and can be modified to book a large number of USRPs.

3.3. Cognitive radios

In recent years, users have used communication services based on social networks, web chats, email and an infinity of applications that require the devices to have greater processing power, memory, as well as fast and efficient connections. Current mobile devices are equipped with multi-core processors, higher capacity memory systems and a

Figure 3.2: jFed login screen

diversity of communication technologies such as Bluetooth, Wi-Fi, and LTE. The demand for communication devices with these characteristics has increased the use of the licensed or primary frequencies and also of the secondary or unlicensed frequencies (ISM - Industrial Scientific and Medical). The intensive use of these devices has caused interferences among primary and secondary frequencies and consequently the spectrum pollution.

The coexistence of different networks and devices that operate at the same frequency, or in adjacent frequencies, may lead to harmful interference, which limits the capabilities of the applications and, in some cases, results in the complete shutdown of those networks. In 2006, the authors [Zhou et al. 2006] predicted that if nothing were done to avoid interference and coexistence problems, the growth of wireless networks could cause the complete overlapping of communication channels. In addition, studies of [McHenry et al. 2006] showed that at some places the 2,400 MHz frequency spectrum, which is used by several Wi-Fi devices, has an occupation of 90%.

To avoid these problems the telecommunication regulatory agencies have auctioned frequency bands each time highest (tens of GHz), but this requires high investment by telecom operators and exhaustive development of standards and devices by industry. On the other hand, the growth of mobile and ubiquitous computing has increased the demand for wireless communications. Several solutions have been proposed to reduce interference in wireless channels such as smart antennas, multiple radios (MIMO), filters, transmission power control, but these solutions do not efficiently explore the frequency spectrum. Other proposals found in literature use solutions based on Dynamic Spectrum Allocation (DSA) to avoid interferences in wireless communication systems and optimize the frequency spectrum. DSA uses mechanisms that include spectrum sensing, choosing the best channel/frequency available and dynamically reconfigure the radio device. These mechanisms have been the groundwork for the development of cognitive or intelligent radios [Correia et al. 2015].

The term Cognitive Radio was defined by Mitola as "*radio technologies that can make possible more intensive and efficient spectrum use by users licensed within their own networks, and by spectrum users sharing spectrum access on a negotiated or an opportunistic basis*" [Mitola 1999]. Moreover, these technologies include, among other things, the ability of devices to determine their location, sense spectrum use by neighboring devices, change frequency, adjust output power, and even alter transmission parameters and characteristics. Cognitive radio technologies enable spectrum exploring in space, time, and frequency dimensions that until now have been unavailable.

In [Commission 2003] the Federal Communications Commission states that cognitive technologies can reconfigure radios according to environment characteristics in real-time, offering to regulatory agencies the potential for more flexible, efficient, and comprehensive use of available spectrum while reducing the risk of harmful interference.

Cognitive radios (CR) emerge as a viable solution to avoid interference, improve network throughput and optimize frequency spectrum usage. Spectrum usage can be optimized by opportunistic frequency exploration in spatial and temporal dimensions, for example: if in a region there is a license for primary frequency use, but the frequency is not exploited, then secondary users could use it; or else, random or seasonal usage of licensed frequencies allow secondary users to opportunistically exploit them. In both cases, it is mandatory for secondary users to sense the spectrum so that, at any indication of transmission of primary users, the channel is released and there is an immediate migration to another frequency. To develop CR is important to create an abstract model with all its components, tools, services and applications. In this way, several CR can be connected to form a cognitive radio network (CRN) or a CRAHN (Cognitive Radio Ad Hoc Network). This model should consider that all elements of the CRN can perform spectrum sensing, have the capacity to exchange information, whether centralized or not and choose the best communication channel.

There are still many technical issues to overcome in cognitive radio. Among many challenges that still need to be solved, two of the most important are presented by [Yucek and Arslam 2009]: the problem of the hidden primary user, and the problem of the spread spectrum primary user. These two issues can lead a cognitive radio to incorrectly choose a frequency that seems to be empty, interfering in primary user's signal. To define the hidden primary user problem consider that there is a primary user A in range to transmit to B, and a secondary user C that is in range of B but not in range of A. C senses the spectrum and because it is out of range of A, it may conclude that there is no primary user. As B is in range to communicate with both A and C, but C cannot detect A, if C begins its transmission it will interfere with transmissions from A to B. Another problem is the spread spectrum primary user. In this case, primary user's low power transmission may seem like background noise, as the cognitive radio may sense the spectrum and interpret primary user's signal as noise. In order to avoid this problem, the cognitive radio should sense a large of the frequency spectrum to identify the primary user.

Several papers propose the development of cognitive radio networks. Akyildiz et al. proposed a framework for spectrum management in cognitive radios [Akyildiz et al. 2008]. This framework is based on a cross-layer model in which the MAC layer reconfigures the radio based on application requirements as well as network state. The framework

model proposed by Akyildiz et al. for spectrum management in cognitive radios is divided into four stages:

1. Spectrum sensing: A CR should be able to monitor the frequency spectrum to avoid interference with primary frequencies, and to search for unused frequencies (spectrum holes). The CR should consider the primary users in the region registered on regulatory agencies, as well as the time of sensing. Information collected by CR can be treated individually by nodes or concentrated and treated centrally by a node.

2. Spectrum decision: With the information about spectrum holes, the CR can choose an unused frequency. The decision method for selecting the best channel can use algorithms based on RSSI (Received Signal Strength Indicator), Artificial Neural Networks (ANN), Correlation, Analytic Hierarchy Process (AHP), Random Forests (RndF) and others. In addition, the decision method can also be influenced by local policies and regulations.

3. Spectrum sharing: The CRN is formed by several CRs that are trying to access the spectrum, often in the presence of other devices that are operating at near frequencies. The network must be coordinated to avoid collisions and to prevent overlapping of spectrum portions. This is achieved by the exchange of messages between nodes containing the best selected channel (or a list of best channels) to be used without causing interference or collisions.

4. Spectrum mobility: When the CR receives information about the best available channel, it reconfigures its radio to new frequency. Spectrum mobility is also employed to avoid that if a portion of the spectrum in use is required by a primary user, the communication must be immediately interrupted and can be continued in another available channel.

This framework proposed by Akyildiz et al. is only a theoretical model that did not present any real implementation.

A framework with a focus on spectrum decision and using Bayesian networks was developed to model spectrum correlation in CRN [Li and Qiu 2010]. The work was based on a graphical modeling and used a numerical simulator to implement analytical models and demonstrate the problems of interference between primary and secondary users. In spite of this, this framework did not consider the complexity and the dynamism of frequency spectrum.

A framework developed for SDR (Software Defined Radio) using GNU Radio was proposed by [Jagannath et al. 2015]. The objective was to implement modules for development and testing of new techniques for automatic classification of multiple signals. Despite the accuracy signal classification, and their importance in spectrum sensing, the authors did not present their application in CRN.

The authors [Marques et al. 2016] proposed an architecture for development of spectrum decision methods for CRN. The architecture was implemented in GNU Radio using broad spectrum bands on real hardware. From this research, it was possible to construct a generic framework component that can be altered to test new spectrum sensing

and spectrum decision methods. In addition, it allows the inclusion of application requirements and definition of other quality of service policies. Its modular organization facilitates the testing of different medium access control protocols and spectrum sharing message exchange.

The following experiments will use this framework to demonstrate the viability of CR implementation in SDR. The framework is based on the abstract model proposed by Akyildiz et al. and therefore follows the four stage model for CR development.

3.3.1. Framework architecture

The framework architecture is based on a cross-layer model that integrates all modules and resources for spectrum exploration. The main feature of this model is to enable dynamic spectrum access, and it is similar to the four stage model proposed by Akyildiz, composed by Sensing, Decision, Sharing and Mobility. The abstraction of the cross-layer communication model for CRN is shown in Figure 3.3, with all layers, modules, blocks and interfaces used to define the communication in a CRN.

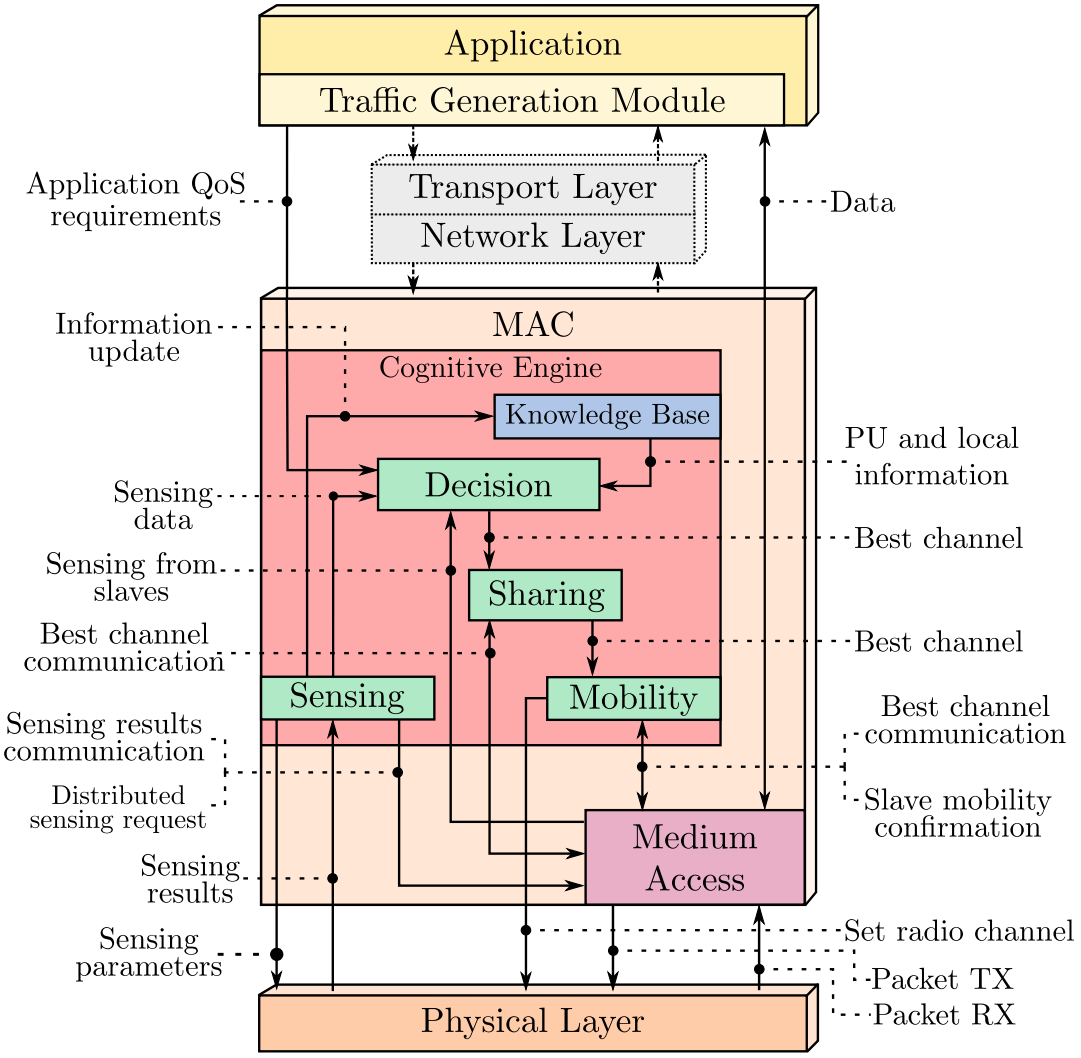


Figure 3.3: Framework architecture.

The application layer can define the quality of service requirements and message traffic type. The quality of service policies can be established by the application to define requirements in terms of bandwidth, latency, and others. These policies directly influence MAC layer's behavior, for example, radio parameters, duty cycle operation, and sensing frequency. The traffic generation module is responsible for generating traffic patterns used for message exchanges. This module was developed to facilitate experimentation, generating traffic for evaluation of framework modules (or blocks in GNU Radio), and simulate the behavior of a real application. Three different distributions were implemented for packet sending interval: uniform, constant, and exponential. In addition, this module includes continuous and discrete distributions [Saucier 2000]. For the experiments presented in this work, only the uniformly distributed traffic model was considered.

Transport and Network layers were not implemented at this stage of the framework. As the experiments focused on spectrum sensing, decision, sharing, and mobility, the services provided by these layers were not essential. It is important to note that the CRN implemented in this work is based on unicast communications. Nevertheless, the modular design of the framework allows the insertion of additional modules to include transport and routing capabilities to form a CRN with multihop communication (CRAHN).

On MAC layer two major modules were defined: cognitive engine and medium access. The cognitive engine is composed of the four stages of a cognitive radio and a knowledge base. Licensed users or primary users (PU) register the contracted frequencies in regulatory agencies to operate telecommunication services. The knowledge base may consist of: a database provided by regulatory agencies, information collected through local spectrum sensing, and information entered manually by the administrator, based on policies or regulatory laws of countries or regions.

Knowledge base information can be dynamically updated by data collected from local spectrum sensing. This information can be preprocessed or not and will be used as input to spectrum decision methods (SD). In this framework, a database provided by ANATEL [8] was inserted as a block, which can be used to get information of PUs in a specific region at a given time. In addition, the local spectrum sensing data also can be entered into this database.

The Spectrum Sensing (SSe) communicates with the physical layer by sending sensing parameters and commands, and collecting the results. Spectrum sensing can be done in two ways: distributed, executed by all nodes; or centralized, executed only by the master node. In distributed mode, the master node requests sensing from all slave nodes by messages transmitted through medium access block. The slaves execute spectrum sensing and send information back to the master node, which combines all the received information in the spectrum decision module.

The main block of the cognitive engine is the Spectrum Decision (SD). All intelligence methods can be implemented within this block. This framework provides four methods for spectrum decision: Artificial Neural Network (ANN), Analytical Hierarchical Process (AHP), Random Forest (RndF) and a simple method based in received signal

[8] Agência Nacional de Telecomunicações - (Brazilian Regulatory agency)

strength (CogMAC) [Marques et al. 2016]. These decision methods are affected by the input parameters: QoS, spectrum sensing and knowledge base. Based on these inputs, the methods decide which is the best channel that avoids interference with primary users and use the spectrum opportunistically. After the SD choose the best channel, it is necessary to share this information with the other nodes.

The Spectrum Sharing (SSh) module communicates the best channel to the slave nodes. In this framework a 6 GHz common control channel (CCC) is used to initialize communication between nodes, to exchange control messages, and as a fallback channel. The choice of this channel is based on empirical data and on the list of PUs provided by ANATEL. The message exchange is also controlled by the Medium Access module.

Information about the best channel is also passed to Spectrum Mobility (SM) block. The master sends messages informing the best channel to all slave nodes, and then migrates to the best channel after receiving mobility confirmation of at least one slave.

The medium access module is responsible for collision avoidance, and controls transmission of packets originated from upper layers and from the cognitive engine. Before transmitting a packet, it performs a Clear Channel Assessment (CCA) to verify medium state. If the channel is free the message is sent by radio. Otherwise, if the medium is busy, the message is delayed by a backoff algorithm. The transmission is retried until the message is delivered or application's timeout occurs. Furthermore, packet reception is also controlled by medium access module.

The physical layer has functions of adjusting radio parameters and transmitting and receiving packets over a wireless channel. In addition, all spectrum sensing is performed on this layer within a range of 800 MHz to 5.8 GHz.

The communication model is shown in the next section.

3.3.2. Communication model and message types

The framework uses a communication model based on the Master-Slave paradigm. Communication is established between two adjacent nodes (one hop distance), so there is no multihop communication. Figure 3.4 shows the message flow between the master node and two slaves nodes.

The master node sends messages in broadcast using the CCC (6 GHz) to start the neighbor discovery phase (ND). Nodes in range reply with $ACK(ND)$. After the discovery phase, the master node sends sensing requests messages (SSe_n) to all its 1-hop neighbors, waits for confirmation messages and enters idle mode. If a neighbor node does not reply after several retransmissions it is considered disconnected.

Slave nodes (S_n) perform spectrum sensing and send the collected data to the master node. All spectrum sensing is performed in 1 MHz steps within the frequency range of 800 MHz to 5.8 GHz. Spectrum decision (SD) phase initiate when the master node receives the spectrum sensing data ($DATA(SSe_n)$) from slave nodes. The best channel decision can be performed by different algorithms. In this framework, there are four algorithms and a knowledge base that contains the historic of primary users in that region (provided by a regulatory agency). The decision method receives as input parameters of

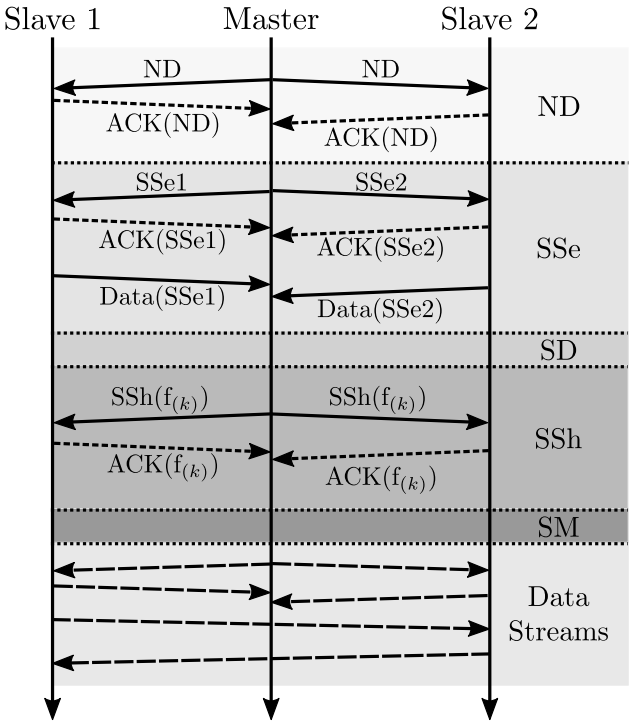


Figure 3.4: Message flow between three nodes.

QoS, collected sensing data from slaves, and PU historic. After processing, the method returns the best channel.

The best channel selected is individually informed to each slave node by the master by sending $SSh(f_{(k)})$ messages. Upon receiving at least one $ACK(f_{(k)})$ the master enters spectrum mobility phase (MD) and configures its radio to the best channel.

After the spectrum mobility phase (MD), all nodes can communicate with their neighbors. This communication can be established between master with any slave or between slaves.

3.3.3. Cognitive radio experiments

In real environments interference and noise influence radio communication. These characteristics have high complexity to be mathematically modeled, resulting in inaccurate transmission analysis through simulations.

This section presents experiments using real hardware, the USRP B200 and B210 daughterboards. These are Software Defined Radio (SDR) devices that have the capability of accessing frequency spectrum in ranges between 70 MHz and 6 GHz.

The network communication model is based on the master and slave paradigm, and the experiments use one master and one slave node. The master node is responsible for centralizing all information of spectrum sensing and apply spectrum decision methods to select the best channel according to application requirements. Slave node performs sensing and sends the collected information to the master node. Although, the network allows the use of multiple slave nodes.

The main objective of these experiments is to present a CRN composed by two nodes, which employ intelligence methods to dynamically access the spectrum, selecting channels with low noise ratio, and avoid interference on primary users.

This CRN uses the four stage model presented on Figure 3.3. In these experiments, only the sensing and decision methods are used, nevertheless, the other methods are fully implemented in the framework. The experiment will be performed in two steps, with each one focusing on a specific method: the first experiment demonstrates spectrum sensing phase, and the second presents the spectrum decision.

Preparing the Environment

The requirements to perform the experiments are described below:

- GNU Radio version 3.7.x.
- Ubuntu 14.04.
- Framework code, available on `http://github.com/GrubiCom/FrameworkCRN`
- 2 USRPs models B200 or B210, with proper cabling and antennas.
- 2 computers with VOLK library support[9].

The framework can be downloaded from GitHub with the command:

```
git clone http://github.com/GrubiCom/FrameworkCRN
```

Two scripts in the framework are used to configure the environment. The first, `setup_dependencies.sh`, installs the software necessary for running the framework. The second script, `build_blocks.sh`, configures additional GNU Radio blocks of the framework. The following commands must be run on the machines used for master and slave USRPs:

```
user@ubuntu:~$ sudo chmod +x setup_dependencies.sh
user@ubuntu:~$ sudo setup_dependencies.sh
user@ubuntu:~$ chmod +x build_blocks.sh
user@ubuntu:~$ ./build_blocks.sh
```

Cognitive Radio Experiment 1: Spectrum Sensing

The objective of this experiment is to perform spectrum sensing using USRP with the master-slave architecture. At the end of this experiment, it will be possible to verify the state of the sensed spectrum. The GNU Radio blocks that will be used in the master-slave architecture are represented in the figures 3.5 and 3.6.

In Figure 3.5, the block diagram of the slave nodes is presented. This type of node does not implement spectrum decision, since it does not execute this phase.

[9]https://gnuradio.org/doc/doxygen/volk_guide.html

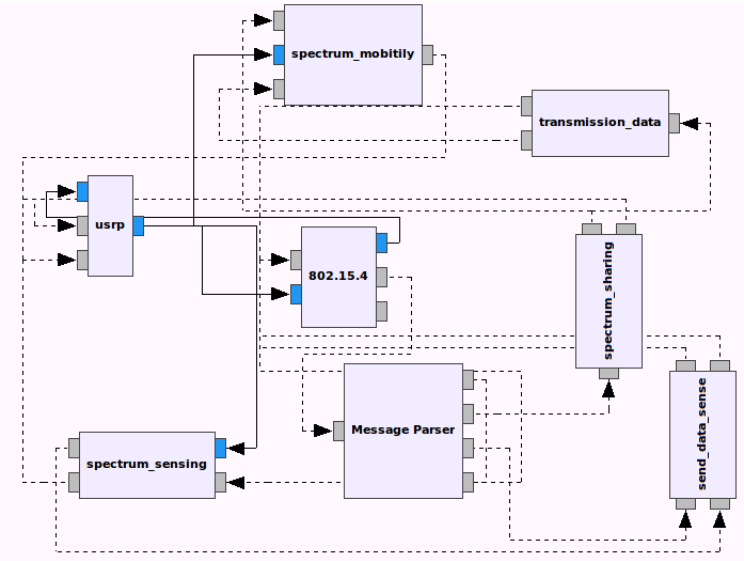


Figure 3.5: Slave node block diagram.

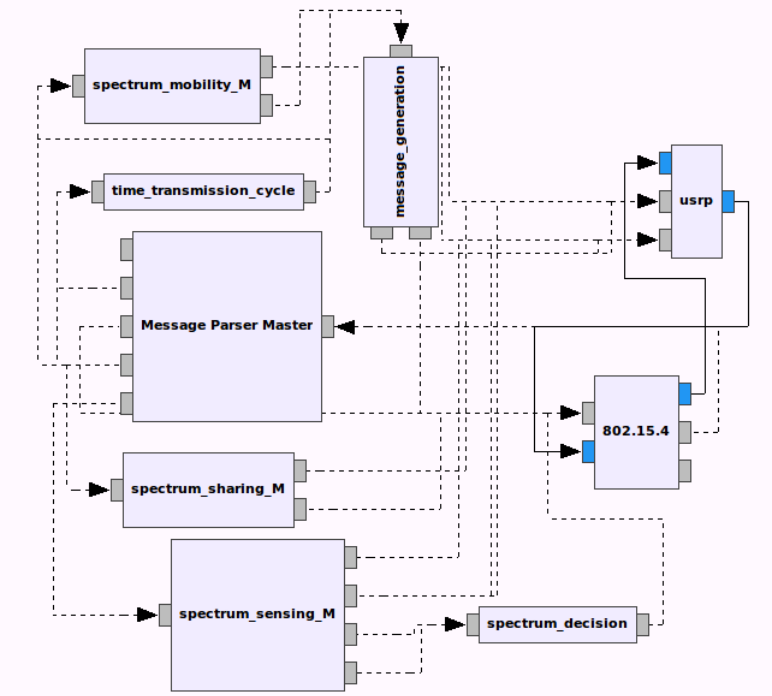


Figure 3.6: Master node block diagram.

Block diagram of the master node is presented in Figure 3.6. This type of node, on the other hand, implements all four stages of cognitive radio, and concentrates the sensing data to execute spectrum decision. For this experiment, the necessary files are located on folder `gr-pmt_cpp/grc`. Spectrum sensing can be run in two different modes:

- In graphic mode, go to File menu and open `slave.grc` file on slave computer, then open `master.grc` on master computer. To execute the code of each node,

click on the green arrow on GNU Radio interface (Execute the flow graph).

- Using the terminal, running the command `grcc -e slave.grc` on slave computer, and `grcc -e master.grc` on master computer.

During the execution in graphic mode, the terminal on GNU Radio's interface of the master node shows a time counter while it waits for sensing replies. On the slave node, the terminal shows the current sensed frequencies.

After finalizing the spectrum sensing, the slave stores the sensed results on `/tmp/` folder. The file `sense.txt` contains raw data, with up to 16 samples of each frequency, during 8 ms. For each frequency, the slave selects the sample with highest noise, and stores on `send.txt` file, which will be transmitted to the master node for spectrum decision.

On the master node, data received is stored on `/tmp/res_sense.txt`. The collected data can be visualized on slave using the script `slave_freq_plot.sh`, while on master node the results can be viewed with `master_freq_plot.sh`.

At this point, the first experiment is finalized, and the CRN is ready to start the second stage, spectrum decision.

Cognitive Radio Experiment 2: Spectrum Decision

Following the experiment 1, it is necessary to select the best channel. This is performed by the master node, during the spectrum decision phase. In this phase, the master combines sensing information received with information on its knowledge base, and application QoS requirements.

The decision method used in this experiment is based on ANN. In Figure 3.6, the decision block is named `spectrum_decision`, being able to decide the best channel with accuracy of 99.9996%. This stage is carried on immediately after spectrum sensing, and no additional executions are needed. After spectrum decision, the best channel frequency is stored on `/tmp/master_channels.txt`, and can be visualized with `best_freq_plot.sh`, both on master and slave computers.

After choosing the best frequency for data transmission, the master node informs the frequency to slave nodes. Slave nodes acknowledge the reception of this information, and upon receiving at least one confirmation, the master node modifies its channel to the best frequency. With the nodes operating in the new channel, data transmission starts and lasts for 60 seconds, when the network repeat the cognitive process.

3.4. Dynamic change of the MAC protocol in WPANs

This section will present how the MAC protocol influences the performance of a WPAN network. It is known that each family of protocols (contention-based and contention-free) perform best on certain scenarios: contention-free MAC protocols (e.g TDMA) are best for crowded networks, however they may have a large overhead and underutilize the medium on uncongested networks. On the other hand, contention-based MAC protocols (e.g. CSMA/CA) present a very low overhead, however, their performance degrades when many stations compete for the medium.

In this experiment, we'll use a rule-based system to change from one MAC protocol to the other in an IEEE 802.15.4 network. We will code our own rules into a set of existing GNU Radio modules, and will evaluate how our system behaves for a varying number of stations transmitting data.

3.4.1. MAC protocols

Multiple access methods in wireless networks are used when the medium is shared and some form of organization is needed so that different nodes can transmit their data satisfactorily. There are several protocols that fulfill this role and they can be divided into two broad categories: multiple access protocols contention-based and contention-free, as shown in Figure 3.7.

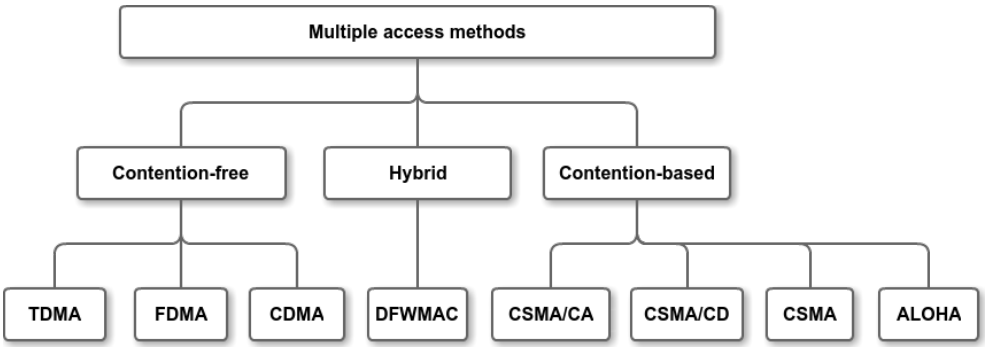


Figure 3.7: Multiple access control methods

In contention-based methods, the transceivers do not have strong restrictions to access the medium. In this case, access does not require coordination to use the channel and the decision to transmit is taken locally. There is also no specific time delimitation within which the transmission should be performed. For these reasons, contention mechanisms are adopted in the protocols to reduce the number of collisions and to avoid channel saturation and thus the network to function satisfactorily. Generally, when a collision occurs, the node waits a random time and repeats the transmission attempt.

The first protocols that used this approach were Pure ALOHA and Slotted ALOHA. This approach was improved by including the verification of the medium before the transmission attempt, resulting in the CSMA (Carrier sense multiple access) and later, its variations with collision detection (CSMA/CD) and collision avoidance (CSMA/CA). In the approach of these protocols, the attempt of transmission is direct, depending only on the carrier sense and the waiting of some previously defined timings. This is positive because in case of no collisions, the use of time is optimized. However, with this approach, the number of collisions tends to increase with increasing devices trying to communicate. At certain point, the number of collisions may be large enough to hamper the operation of the network [Takagi and Kleinrock 1985, Ziouva and Antonakopoulos 2002].

In contention-free methods, the medium is accessed in a coordinated way, and there is no need for mechanisms to resolve access conflicts and collisions. In this case, the allocation control of the channel can be done by a centralized entity, which coordinates the transmission order, or by some distributed approach as token passing or distributed queue. Some examples of protocols that do not use contention mechanisms are Frequency

Division Multiple Access (FDMA), Time Division Multiple Access (TDMA), and Code Division Multiple Access (CDMA) [Busch et al. 2004].

TDMA-based protocols, for example, divide time into slots where each device transmits individually. Since a slot is reserved for only one device, there are a small number of collisions, no matter how many devices are trying to communicate. Therefore, it is a good approach when the network has many nodes trying to transmit. Despite this advantage, when there are few transmission attempts, the control messages used in the slots allocation procedure consume a relative large time. So, these protocols are best when the network is congested, but they lose performance when there are few nodes trying to communicate.

There is also a hybrid approach, which uses combined contention-based and contention-free methods. In this approach, each mode operates for a certain time and there is a switch between the two modes at the end of each period. In the Contention Period (CP), the medium access organization is distributed, and in the Contention Free Period (CFP), the medium use is coordinated by a access point. DFWMAC is an example of protocol that uses this approach [Diepstraten and WCND-Utrecht 1993].

3.4.2. Experiments

In the experiments performed in this part of the course, we use the FS-MAC platform [Cordeiro et al. 2017] to automatically switch between a contention-based (CSMA/CA) and a contention-free protocol (TDMA). This platform works with these two protocols putting each one in operation according to some predefined rules. These rules are part of a fuzzy system that uses the information about *number of senders* and *packet delivery latency* to infer the contention level of the network and define which protocol should operate at each moment. Further details on the architecture and operation of the FS-MAC platform can be found at the address:
`https://github.com/jeffRayneres/FS-MAC`.

In the following sections we describe the settings required to use the FS-MAC platform, then we describe the experiments.

3.4.2.1. Configuration

To perform the experiments, we must install the FS-MAC platform and its dependencies. The dependencies, with their respective addresses, are:

- *gr-ieee802-15-4*: `https://github.com/bastibl/gr-ieee802-15-4`
- *gr-foo*: `https://github.com/bastibl/gr-foo`
- *gr-eventstream*: `https://github.com/osh/gr-eventstream`
- *gr-uhdgps*: `https://github.com/osh/gr-uhdgps`

The *gr-ieee802-15-4* project provides the ZigBee stack used as the basis for building the FS-MAC platform, and the *gr-foo* project is a dependency on that stack. The

gr-eventstream and *gr-uhdgps* projects are used in the CSMA/CA protocol of the FS-MAC platform in the Carrier sense process. A detailed description of the *gr-ieee802-15-4* installation can be found in [Silva et al. 2015]. In addition to these projects, we must also install the libraries *liblog4cpp5-dev* and *python-matplotlib*.

The FS-MAC platform can be downloaded from the address `https://github.com/jeffRayneres/FS-MAC`, or the user can use the command:

```
git clone https://github.com/jeffRayneres/FS-MAC
```

After downloading the project in the FUTEBOL testbed, go to the /gr-fsmac directory and run the following commands:

```
mkdir build
cd build
cmake ..
make
sudo make install
sudo ldconfig
```

After executing these commands, the FS-MAC platform will be installed and the environment will be prepared for running the experiments.

In the experiments we use resources provided by the FUTEBOL project, so in directory */gr-fsmac/examples* of the FS-MAC platform there is an RSpec file to be used with jFed for resource allocation. To improve execution, we do not use the graphical mode of the GNU Radio Companion, so all information generated during the experiments, including which protocol is in operation, will be displayed on the terminal window.

3.4.2.2. Dynamic Change Experiment 1

The objective of this experiment is to exchange the MAC protocol in operation on the network automatically according to some rules based on the amount of contention on the wireless medium. The file *decision.py* in the */gr-fsmac/python* directory of the FS-MAC platform, implements *fuzzy* system that receives as input the number of senders in the network and the average latency of package delivery. Its output is the effectiveness of a certain MAC protocol. The system employs fuzzy rules that determine which protocol should operate at a given moment, according with the network contention level. The fuzzy system was modeled as follows:

Linguistic variables: The model considers three linguistic variables, which are (i) Average latency of packets delivery (*AL*), ii) Number of senders (em NS) and iii) Adaptability of the protocol (*ADP*). They all accept the fuzzy terms *LOW* and *HIGH*.

Fuzzy rules:

```
CSMA
```

```
If NS is LOW and AL is HIGH then ADP is HIGH
If NS is LOW and AL is LOW then ADP is HIGH
If NS is HIGH and AL HIGH then ADP is LOW
If NS is HIGH and AL is LOW then ADP is HIGH

TDMA
If NS is LOW and AL is HIGH then ADP is LOW
If NS is LOW and AL is LOW then ADP is LOW
If NS is HIGH and AL is HIGH then ADP is HIGH
If NS is HIGH and AL is LOW then ADP is LOW
```

Membership functions: The membership functions are shown in the graphics of Figures 3.8, 3.9, 3.10 and 3.11 (source [Cordeiro 2017]):

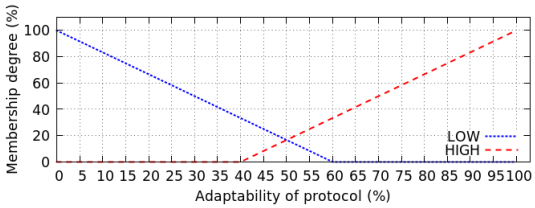


Figure 3.8: Membership functions for linguistic variable “Adaptability”.

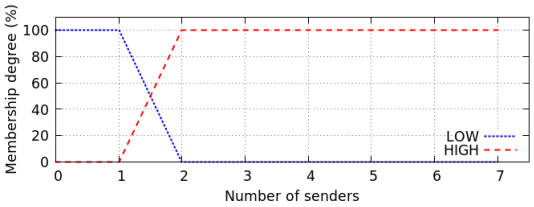


Figure 3.9: Membership functions for linguistic variable “Number of senders”.

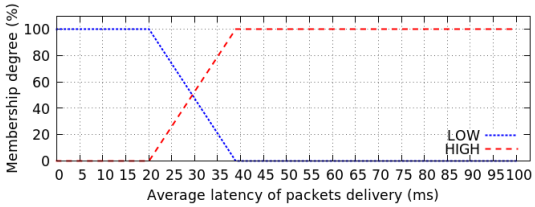


Figure 3.10: Membership functions for linguistic variable “Average latency - CSMA”.

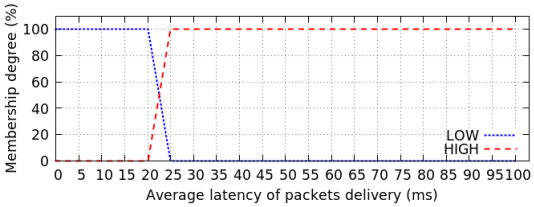


Figure 3.11: Membership functions for linguistic variable “Average latency - TDMA”.

In this experiment, we use three USRPs to simulate sensor nodes. Each USRP is connected to a computer installed with GNU Radio version 3.7. We call *StationN* the set formed by a USRP connected to a computer, so in the experiments we use a group formed by Station1, Station2 and Station3.

In the directory */gr-fsmac/examples* of the FS-MAC platform, there are three files configured to be used in the stations. The names of these files are *transceiverStation1.py*, *transceiverStation2.py* and *transceiverStation3.py*. These files are scripts generated by GNU Radio Companion, they contain all the necessary settings for the transmissions in this experiment. These settings include:

- Preparation of a message with 110 bytes to be sent.

- Setting the send interval to 20 ms.
- Setting the transmit power to 60 dBm.
- Setting the frequency used to 2.48 GHz.
- Setting the transmitter and receiver MAC address.

In addition to transmitting data packet and acknowledgement packet, Station1 operates as FS-MAC Coordinator. In the FS-MAC platform, the Coordinator node is the one that coordinates the exchange of the protocol in operation when necessary.

The experiment starts with the transmission of packets from Station1 to Station2. After some time, without interruption of Station1's transmission, the new transmission starts from Station2 to Station3. After some time, keeping the transmissions in progress, a new transmission starts from Station3 to Station1. When each sender is included, we check in the terminal which protocol is in operation. With the rules configured in the *fuzzy* system, in the *testbed* that we use, the contention level that indicates the moment of exchange of the protocol must occur when the network changes from one to two senders. Thus, when there is only one sender, the system operates with CSMA/CA, with two senders or more, the system automatically starts to operate with TDMA.

3.4.2.3. Dynamic Change Experiment 2

The purpose of this experiment is to change the rules of the *fuzzy* system so that the MAC protocol exchange in the network occurs at a different contention level from the previous experiment. To do this, we need to change the *decision.py* file in the */gr-fsmac/python* directory of the FS-MAC platform, we need to modify the functions *senders function()* (line 270) and *data function()* (line 250). These functions are responsible to calculate the membership of the values of *Packet delivery latency* and *Number of senders* to the HIGH and LOW sets. They reflect the membership functions of Figures 3.9, 3.10 and 3.11. We'll modify them to reflect the membership functions shown in Figures 3.12, 3.13 and 3.14.

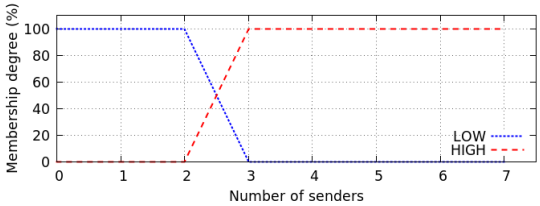


Figure 3.12: Membership functions for linguistic variable "Number of senders".

In this experiment we changed the membership functions so that the FS-MAC platform identifies the moment of exchange of the protocol in operation when the contention level is slightly higher than in the case of the previous experiment. After the functions have changed, we must reinstall the FS-MAC platform.

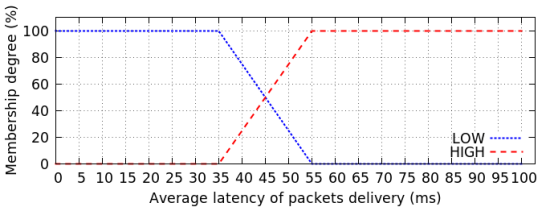


Figure 3.13: Membership functions for linguistic variable "Average latency - CSMA".

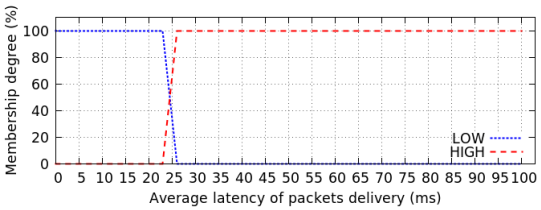


Figure 3.14: Membership functions for linguistic variable "Average latency - TDMA".

In terms of execution, this experiment follows the same dynamics of Experiment 1, that is, it starts with one sender and adds other sender up to a total of three. When each sender is added, we check in the terminal which protocol is in operation. With changes in the *fuzzy* system, the contention level that indicates the moment of exchange of the protocol must occur when the network changes from two to three senders. Thus, when there are one or two senders, the system operates with CSMA/CA, with three senders, the system automatically starts to operate with TDMA.

3.5. Reliability in WBANs

Wireless Body Area Networks (WBANs) consist of a wireless network composed of several wearable or implantable computing devices. Although WBANs can be applicable in various fields [Movassaghi et al. 2014], this section focuses on healthcare, in which the application is the monitoring and transmission of physiological signals to medical servers. In this part of our course, we will perform a reliability experiment in WBANs. First, we evaluate the classic IEEE 802.15.4 protocol stack, measuring the amount of transmitted data from both the sensor and the base station points of view. With this data in hand, we will evaluate the simplified protocol stack, quantizing the packet delivery efficiency. Next, we move forward using a more robust IEEE 802.15.4 protocol stack. This stack implements acknowledgment functionality providing single channel communication between devices with support to a three-way handshake. Finally, we will compare both protocols packet delivery efficiency.

3.5.1. General dependencies

In order to perform the experiment, we need to install some modules and complementary libraries.

First, we should download the module **gr-foo** developed by [Bloessl et al. 2013]. This module contains blocks responsible for the configuration of sending and displaying the data packets. For this experiment, we will use only the part in charge of visualizing the received data packets. The reader can get this module by running the following command on the terminal:

```
git clone https://github.com/bastibl/gr-foo.git
```

Other non-native GNU Radio plug-ins to download are **gr-eventstream** and **gr-uhdgps**. Both composing the set of blocks entrusted by performing the Carrier Sense

function. However, prior to the installation of these, it is necessary to install the following additional libraries with their respective commands:

```
sudo apt-get install liblog4cpp5-dev
sudo apt-get install python-matplotlib
```

To obtain the **gr-eventstream** module, execute the following command:

```
git clone https://github.com/osh/gr-eventstream.git
```

Similarly the command for the **gr-uhdgps** module is:

```
git clone https://github.com/osh/gr-uhdgps.git
```

Finally, we should download the module **gr-traffic_generator**, which is in charge of generating dynamic messages in a variable interval of time. The traffic generator block receives as parameters one value for the size of the message and another value for a time interval. These parameters can be generated from the distribution type (Pareto, Poisson, Weibull, Zipf or Uniform) with the help of the Distribution block, or defined by the user. The reader can download this module by running the following command:

```
git clone https://github.com/AdrieleD/gr-traffic_generator.git
```

The installation of each additional module downloaded is simple and follows the GNU Radio standard. Simply access their respective folders by creating a folder named build in the root of the project, access that new folder and execute the following commands:

```
cmake ../
make
sudo make install
sudo ldconfig
```

To uninstall, from within the **build** folder, just run the command:

```
sudo make uninstall
```

After the installation of each of the modules and complementary packages, the environment is able to receive the packages related to the respective experiments. Each experiment will depend on the main module containing its corresponding protocol stack, one simplified and one more robust, respectively.

3.5.2. WBAN Experiment 1

The objective of this experiment is to transmit data from a sensor node to a receiving station. This experiment uses real communication devices, where it is simulated sensor nodes and a base station. Furthermore, we will use a simplified protocol stack, which is only responsible for packaging and sending the data. At the end of the experiment, it will be possible to visualize the data transmitted from one device to another and compare them to each other. The purpose of this verification is to evaluate possible data loss and interference by using the simplified stack. We should download and install the **gr-mac1** by executing the following command:

```
git clone https://github.com/AdrieleD/gr-mac1.git
```

After performing all the steps for installing the protocol stack module, we can view the **gr-mac1** library in GRC. To complete the installation of this module, we need to install the hierarchical block, opening the *gr-mac1/examples/ieee802_15_4_OQPSK_PHY.grc* file in GRC and compile it (Generate the flow graph / F5 key) or using the following command line (we recommend the user to restart the GRC environment after this procedure):

```
grcc ieee802_15_4_OQPSK_PHY.grc
```

Please open the *gr-mac1/examples/transceiver_OQPSK_TX.grc* file to check if installation has been successful. All blocks should be properly connected and with no error messages. It is important to emphasize that to carry out the experiment, we need at least two computers, one to act as transmitting node and another as base station. Although the protocol stack used in both is the same, we need to make some changes according to the function that the application will perform. If you have access to the GCR graphical interface, you should edit the *gr-mac1/examples/transceiver_OQPSK_TX.grc*. When it comes to the receiving application, traffic generating blocks are not necessary, so they can be disabled or even deleted. If you are using the testbed terminal you do not need to perform any modification.

For the experiment to work properly, we should also update the address configuration, found in the *gr-mac1/lib/mac.cc* file. Both the source address and the destination address must be different. If such a change is necessary, it can be done by changing lines 60 and 355. In Table 3.1, we are show an example of how to configure two computers. In addition, we need to change the location where the file containing the experiment data will be saved. This configuration is performed in the File Sink block of the application which represents the base station. For the transmitting node, the use of the Wireshark Connector and File Sink blocks is not required. After any and all file editing of the blocks, it is essential to reinstall the blocks so that they can be effective. This is possible through the following commands:

```
cd gr-mac1/build
sudo make uninstall
sudo make install
sudo ldconfig
```

Table 3.1: The configuration of source and destination addresses for two computers.

	Computer 1	Computer 2
Source Address	60 char mac_addr_1 = 0x41;	60 char mac_addr_1 = 0x40;
Destination Address	355 addr0[0] = 0x40;	355 addr0[0] = 0x41;

After completing these steps, the environment is configured and able to start the experiment. Just open the *gr-mac1/examples/transceiver_OQPSK_TX.grc* file (Figure 3.15) and execute (F6 key) on the GNU Radio graphic interface. It is also possible to use the testbed terminal. In this case, you can compile and execute the respective files according to the desired computer function. The following commands initialize the computer as a receiver or transmitter, respectively:

```
grcc -e transceiver_OQPSK_RX.grc
grcc -e transceiver_OQPSK_TX.grc
```

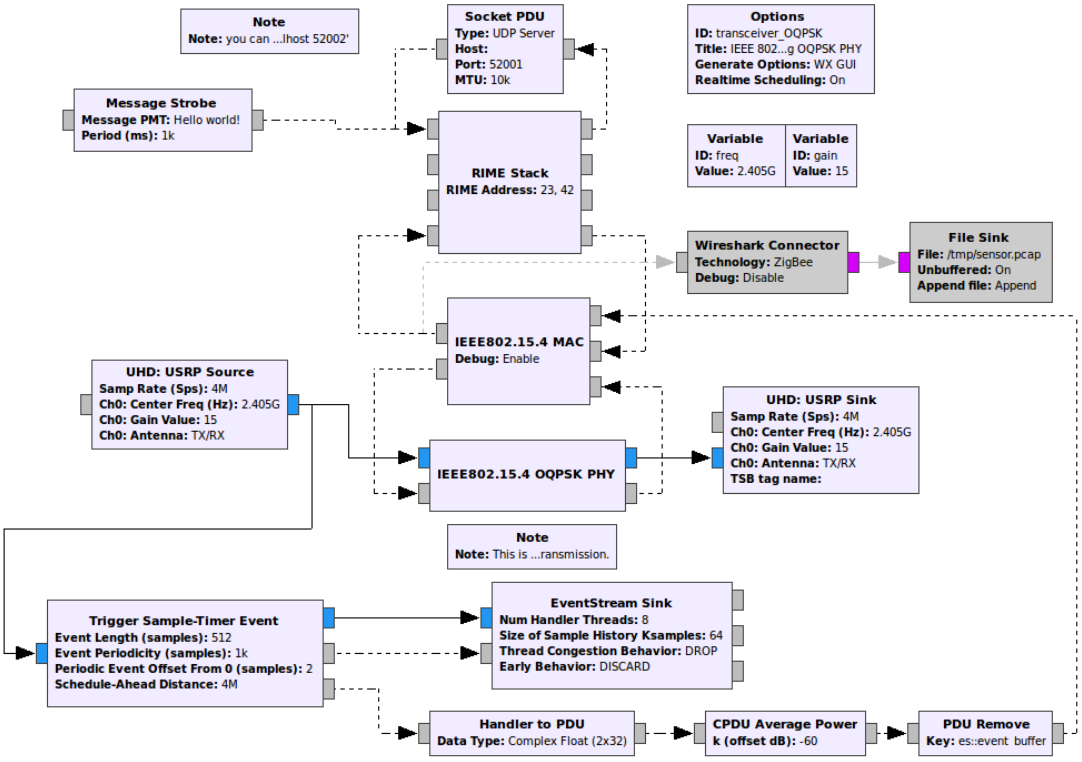


Figure 3.15: Graphical display of protocol stack for experiment 1.

In this experiment, the computer acting as base station should be initialized first. In this way we will start the communication between the nodes, always remembering that each computer with its respective USRP unit is equivalent to a node. Thus, for example, we can have 4 transmitter nodes and 1 base station, we need 5 computers and 5 USRPs.

The transmitting application is configured to send the "Hello world!" message endlessly with a one-second interval between messages. These settings can be changed

by configuring the properties of the Periodic Message Strobe block. Once the experiment is started, the sensor node configured as the base station is able to receive and confirm the messages transmitted by any transmitting node, as long as both are configured on the same channel. The selection of the channel occurs manually, manipulating the settings of the Variable block with ID “freq”.

Since the content of the message sent by the transmitting entities is static and predetermined, the visualization of these via an interface or the persistence in file becomes irrelevant. However, the confirmation of the sending and receiving of the packet by the base station is essential so that the full transfer of data between the devices can be identified. The persistence of data by the receiving station is a must. Once stored, they can be analyzed, compared and used to produce information. This data retention is the responsibility of the Wireshark Connector block, which saves the messages in a .pcap file. In this way, the reader can check the newly generated .pcap file if the message sent by the transmitter was correctly received by the USRP. This file can be better analyzed with the help of Wireshark (Figure 3.16), where we can view the transmitted messages and their respective text content.

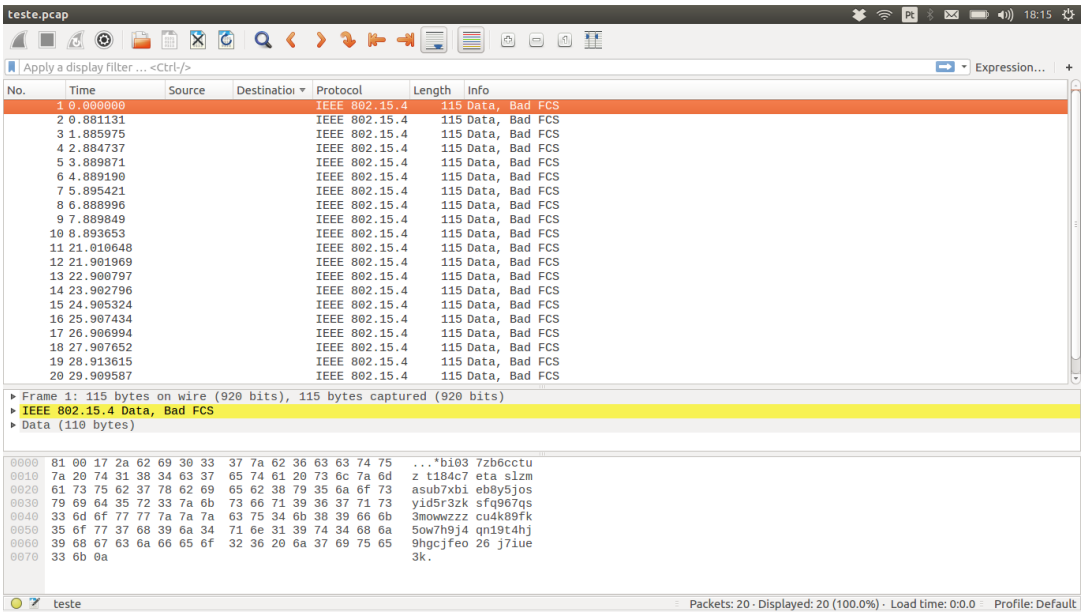


Figure 3.16: Viewing data packets through Wireshark.

3.5.3. WBAN Experiment 2

This experiment is an improvement of the previous one. We will also perform a data transmission between the devices, sensors and base station, but using of a more **robust protocol stack**. The stack used in this protocol was implemented based on the stack of the first experiment, but it has features that provide a higher quality and reliability in data transmission. Included in these functionalities is the Frequency Hopping technique, which provides interference inhibition and better use of the frequency spectrum. Three-way Handshake is another utility, responsible for establishing a single channel of communication between two devices. Another objective of this experiment is to evaluate, as in the previous one, the quality of data delivery. We also provide a quality comparison

of the data transmitted by both protocols. First, we should download the **gr-mac2** module through the following command:

```
git clone https://github.com/AdrieleD/gr-mac2.git
```

We should also carry out the complementary modules installation process. As well as in the previous experiment, after completing all the steps of installing the module referring to the protocol stack, it becomes possible to view the gr-mac2 library in GRC. To complete the installation, it is essential to compile the hierarchical block *gr-mac2/examples/ieee802_15_4_OQPSK_PHY.grc* in GRC (Generate the flow graph / F5 key). Similarly to experiment 1 it is advisable to restart the GRC environment and check the installation success by opening the *gr-mac2/examples/transceiver_OQPSK_TX.grc* file where all blocks must be properly connected and without error messages. You can also perform this step executing the following command in the testbed terminal:

```
grcc ieee802_15_4_OQPSK_PHY.grc
```

This experiment requires at least two computers, one acting as a transmitter and another as a receiver. The stack used for both projects is the same, but we should make some configurations to act according to their specific functions. Please refer to **WBAN Experiment 1** instructions on how to configure the nodes as transmitter or receiver.

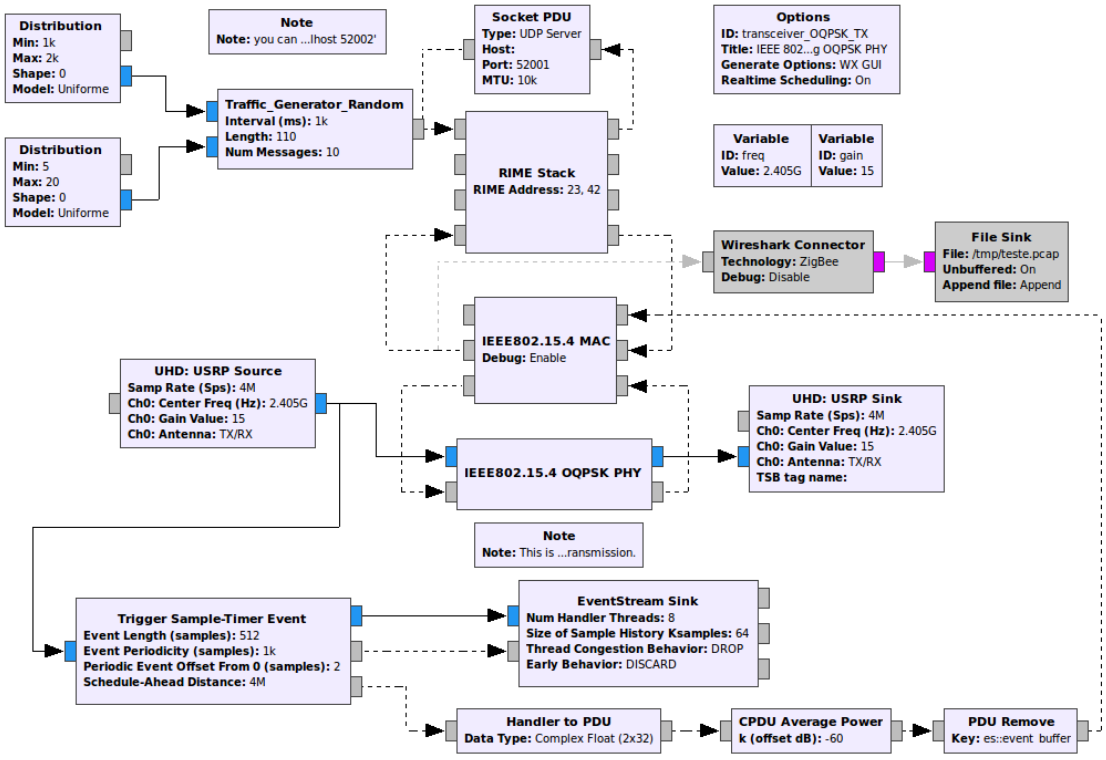


Figure 3.17: Graphical display of protocol stack for experiment 2.

After we have completed all the above procedures, the environment is configured and the user can start the experiment. To do this, simply click the Execute the flow graph

icon or use the F6 hotkey. In this experiment, the initialization order of the devices is not important, but it is suggested that the node with base station characteristics be started first. Similarly to the first experiment, each node is conditioned to the existence of a computer with a USRP unit. You can also execute the experiments using the following commands to initialize the computer as a receiver or transmitter, respectively:

```
grcc -e transceiver_OQPSK_RX.grc
grcc -e transceiver_OQPSK_TX.grc
```

Unlike experiment 1, the content and time between messages are dynamic. They can be changed according to the configuration of the Distribution, **Traffic_Generator** and **Traffic_Generator_Random** blocks. In order to be viewed by the transmitting application, it is necessary to include a Message Debug block connected to the output of the traffic generator module. By enabling the "Debug" option of the IEEE 802.15.4 MAC block, it is also possible to observe the attempts to connect and transmit the packets with their respective confirmations, in addition to the end of the connection. A log is made available at the end of all submissions with information regarding data transmission.

In the base station application also with “Debug” option enabled, is visualized the establishment of the communication such as its closure, messages referring to received data and information about the “jumps” between the channels. For a detailed verification of the received data, the newly generated .pcap file is analyzed with the help of Wireshark, where we can see the transmitted messages and their respective text content as Figure 3.18 depicts. In this Figure, we can see: 1) Command to finish communication; 2) Finished MAC; 3) Total communication time; 4) Total number of sent packets; 5) Total number of confirmed packets; 6) Number of retransmissions; 7) Packet throughput; 8) Bytes throughput; 9) Packet delivery rate; 10) Packet latency.

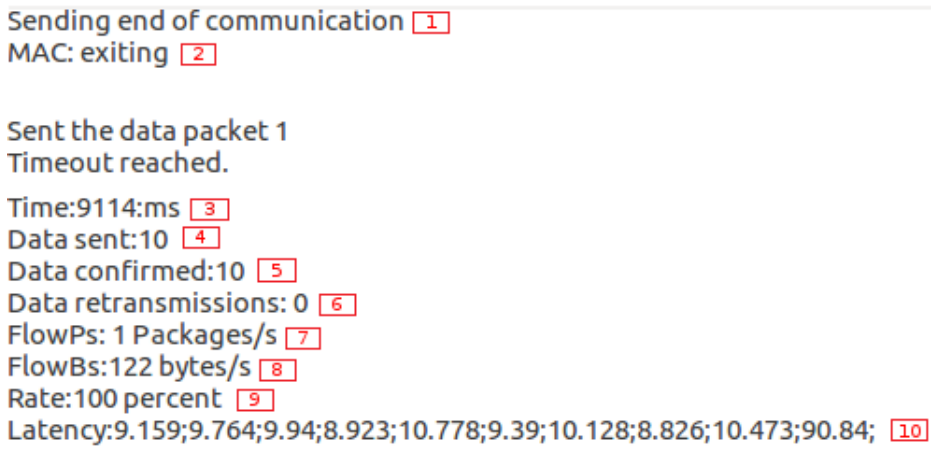


Figure 3.18: Log of the transmitted data.

3.5.4. Comparison between protocols

As is remarkable, the stack structure of the protocols used in the two experiments is similar (Figures 3.15 and 3.17). Taking into account a basic application, both would perform significantly. However, more complex applications, such as WBANs, require

functionality that ensures reliability, accuracy, and agility in data transmission and confirmation. In addition to the basic characteristics necessary for communication between one or more transmitter nodes and a base station, the protocol of experiment 2 has other functionalities such as Frequency Hopping, Handshake. Table 3.2 details both protocols characteristics.

Table 3.2: Comparison between the protocols used in experiments 1 and 2, respectively.

Features	Experiment 1	Experiment 2
Static message and fixed interval	x	*
Data frame	x	x
Data package	x	x
Broadcasting	*	*
Message addressed	x	x
Carrier Sense	x	x
ACK frame	x	x
ACK package	x	x
Dynamic message and variable interval		x
Control frame		x
Control package		x
RTS/CTS		x
Handshake		x
Frequency Hopping		x

* It has the functionality, however, it is not justified the use of it.
x It has the functionality

In the Handshake process, the transmitting node requests the establishment of communication by sending the request command (RTS). Considering the immediate availability of the base station, upon receiving this command a confirmation is sent along with the CTS. After this confirmation, there is a dedicated communication between the two nodes and the data will be transmitted until the connection is terminated. The Frequency Hopping technique allows the node to make a previous evaluation of the quality of a channel, in this case, the transmitting node. If it has some interference or noise, 5 MHz "jumps" are performed until a channel free of these factors that are harmful to communication is found. For the base station, the "jumps" are performed every 50 ms within the 2.4 GHz band, in order to sweep the entire free frequency spectrum. In both nodes, the starting channel starts from a random choice.

Through the improvements mentioned above, it is possible to notice some improvements regarding the performance in data transmission. The transmission carried out in one fixed channel causes a saturation of this channel and the underutilization of the others. This causes congestion and packet loss caused by interference from other nodes or external devices. Packet loss for any reason leads to data retransmissions, which increase traffic on the channel and contribute to low data throughput. In addition to contributing to an increase in energy consumption. Problems of this nature are easily bypassed with the deployment of Frequency Hopping. On the Handshake side, it is possible to smooth

application layer problems since the received packets are always from the same transmitter. Thus, there is no need for further checks and data ordering by devices. Other benefits, achieved through the establishment of a secure communication channel, were accuracy and reliability as the problems with interference were smoothed.

3.6. Experimenting with Orthogonal Frequency-Division Multiplexing (OFDM) Modulation

In this section, we introduce multi-carrier modulation and Orthogonal Frequency-Division Multiplexing (OFDM). OFDM is a modulation that is used in popular high-speed network standards, such as WiFi, DSL, and 4G. Trinity College Dublin's experiment consists of sending packets from the transmitter (Tx) N210 USRP to the Receiver (Rx) N210 USRP using an OFDM signal generated by GNU Radio [Blossom 2004]. The experiment will illustrate the capability of changing center frequency, bandwidth, gain, modulation depth, and cyclic prefix dynamically, and the impact of those parameters on the transmitted and received signal. Through this course, the reader will gain an appreciation for the factors that are most important in the use of OFDM for wireless communication by exploring its configuration and use on real radios. The goals of this section include:

- Understand the need for OFDM signaling in telecommunications networks.
- Comprehend and understand some of the theory behind multi-carrier modulation and the design of OFDM.
- Observe and examine OFDM performance and behavior under different conditions using real radio experimentation equipment.
- Exposure to advanced Future Internet Research and Experimentation (FIRE) testbed infrastructure.
- Experience using GNU Radio, an open-source software toolkit that provides signal processing blocks to implement software radios.

Figure 3.19: OFDM Standards.

3.6.1. Multi-carrier Systems and OFDM Standards

The Collins Kineplex System was the first multicarrier system based on orthogonal subcarriers in HF military radio links. It was built in 1957. In 1966, a team at Bell labs filed a patent (granted in 1979) and published the first article on OFDM systems in IEEE Trans. Communications Technology in 1967 entitled: Performance of an efficient parallel data transmission system [Saltzberg 1967]. OFDM was quickly recognized as an efficient data transmission method and research and standards continued to evolve in the 1980's, 1990's and 2000's with the addition of the following (see Figure 3.19):

- Power-Line-Communication.
- Broadcast: DVB-C2.
- ADSL/-2/-2+.
- Digital Subscriber Line (DSL) technologies.
- Broadcast: DAB, DVB-T/-T2, DVB-H, ISDB-T.
- Wireless Personal Area Network (WPAN): WiMEdia.
- Wireless Local Area Network (WLAN): IEEE802.11a/g/n/ac/ad, IEEE 802.15.4g, HiperLAN/2.
- Wireless Metropolitan Area Network (WMAN): IEEE 802.16a WiMAX.
- Mobile telephony: LTE (3.9G), LTE Advanced (4G)

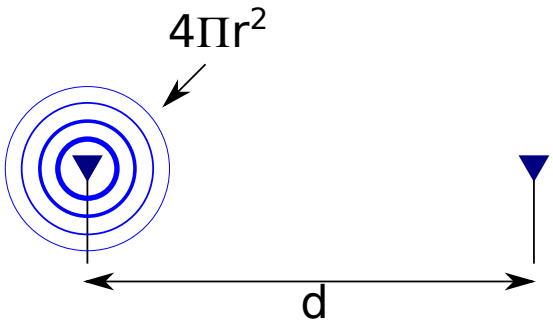


Figure 3.20: Signal Transmitter and Receiver.

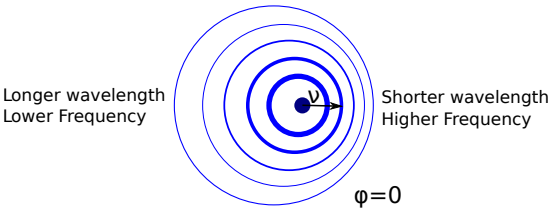


Figure 3.21: The Doppler Shift.

3.6.2. The Wireless Channel

A signal undergoes changes when it is transmitted on its way to the receiver, see Figure 3.20. A key metric of the joint impact of a wireless channel is the variation and attenuation in received signal envelope power over time and/or space, which is called fading. There are two types of fading used to describe the signal level at the receiver, large scale-fading and small-scale fading.

3.6.2.1. Large-Scale Fading

In large-scale fading, signal power falls quadratically with distance as a result of attenuation and diffraction, which occurs due to the signal traveling over large distances and using different frequencies i.e. signal path loss. Large objects such as trees, buildings, mountains, and so forth cause shadowing, and as a result received power can vary dramatically.

3.6.2.2. Small-Scale Fading

In small-scale fading, which is due to reflectors, scattering and receiver motion, multiple versions of the transmitted signal can be received from different path lengths spread over time. There are several types of small-scale fading. These include Multipath and Motion.

Multipath. If the channel is considered as a linear-time invariant system, the convolution of the channel impulse response h(t,τ) with the input stimulus x(t) (the transmitted signal) yields the system output y(t) (the channel output, i.e. the received signal). Delay spread $\sigma \downarrow \tau$ is the maximum difference between times of arrival of multipath components. The following YouTube video illustrates multipath small-scale fading [Ó Coileáin 2016].

Motion. If transmitter, receiver and/or interacting objects are in motion with the speed v under relative angle ϕ, the received signal gets shifted in frequency by Δf due to the Doppler effect, i.e., Doppler shift (see Figure 3.21). Different propagation directions result in different Doppler shifts per multipath component. Received envelope power depends on constructive or destructive addition of signals. The following short YouTube video gives a high level explanation of the Doppler Effect [Alt-Shift-X 2013].

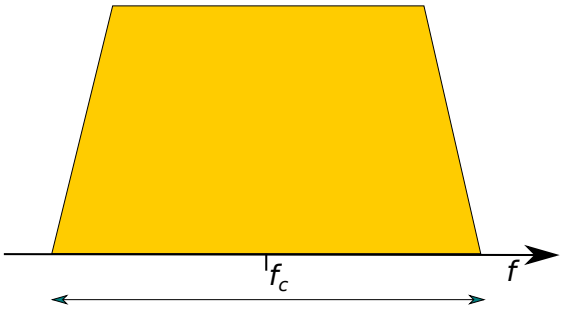


Figure 3.22: Single carrier or mono-carrier system.

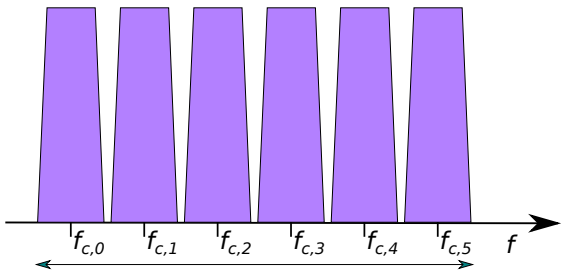


Figure 3.23: Multicarrier system.

3.6.3. Multi-carrier Systems

Multimedia applications require higher and higher data rates from wireless and wired communications systems. Mobile radio channels are fading channels that can be flat or frequency selective. For high bandwidth applications, channels are frequency selective. In conventional single-carrier modulation techniques this can only be achieved by, see

Figure 3.22: transmitting shorter symbols => limited in the case of multi-path propagation (Inter-symbol interference (ISI)); and transmitting more bits per symbol => limited by noise and other distortions. In single carrier/mono-carrier system with symbol width 1/W, data is transmitted using only one carrier. Disadvantages include:

- Event frequency selective fading.
- Equalization is complex.
- Very short pulses.
- Inter-symbol interference (ISI) is long.
- Poor spectral efficiency because of guard bands.

Multicarrier modulation is a technique where multiple low data rate carriers are combined by a transmitter to form a composite high data rate transmission, see Figure 3.23. To improve the spectral efficiency, guard bands between carriers need to be eliminated. In a classic multi-carrier system, the available spectrum is split into several non-overlapping frequency sub channels. The individual data elements are modulated into these sub-channels and are thus frequency multiplexed.

Symbol width=N_c/W and data stream is split up into multiple lower data rate sub-streams, see Figure 3.23. They are modulated and transmitted in parallel on different sub carrier frequencies i.e. Frequency Division Multiplexing (FDM). By parallel data transmission on NC sub-carriers, symbol duration TS can be increased by factor N_c to achieve the same data rate. Longer symbols are less susceptible to inter-symbol interference (ISI). Other advantages include:

- Flat fading per subcarrier.
- N_c short equalizers.
- N_c long pulses.
- ISI is relatively short.
- Poor spectral efficiency because of guard bands.
- It is easy to exploit Frequency diversity.
- 2D coding techniques are allowed.
- Dynamic signaling is possible.

3.6.4. Orthogonal Frequency-Division Multiplexing (OFDM)

Orthogonal frequency-division multiplexing (OFDM) is a multicarrier modulation technique for encoding digital data on multiple carrier frequencies. It is an FDM scheme that uses a large number of sub-carrier signals. These signals are orthogonal to each and carry parallel channels of data. In classic multicarrier systems, guard bands have to be inserted, resulting in poor spectral efficiency. A more efficient approach is to allow the spectra of individual subcarriers to overlap, see Figure 3.24. Zero crossings occur at every multiple of and hence no inter-carrier interference is present i.e., no overlap at sampling frequencies. The following YouTube video gives a high-level overview of OFDM technology [Huawei 2014].

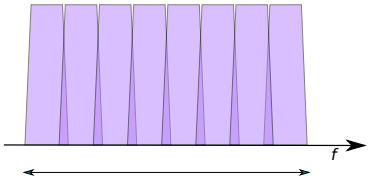


Figure 3.24: OFDM subcarrier tones are separated by the inverse of the signalling symbol duration.

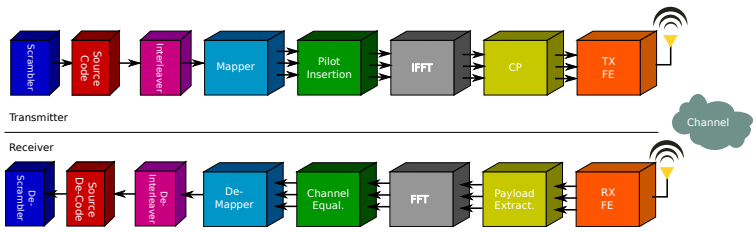


Figure 3.25: Framework architecture.

Problem: If individual subcarriers are overlapping isn't there interference between carriers?

Answer: No! If subcarrier tones are separated by the inverse of the signaling symbol duration, independent separation of frequency-multiplexed tones is possible. Additionally, sub-spectra may overlap in the frequency domain, which supports more efficient use of available spectrum and greater data rates are achievable.

In the remainder of this section, we give you a brief overview of some basic OFDM concepts that we will explore further in the experimentation section using TCDs IRIS FIRE testbed equipment [Collins 2016] [Collins et al. 2016]. These include symbol mapping/de-mapping, Inverse Discrete Fourier Transform (IDFT), Discrete Fourier Transform (DFT), equalization, cyclic prefix, and frequency sensitivity. Figure 3.25 illustrates the OFDM Systems Model, and how these concepts are interconnected.

Symbol Mapping / De-Mapping. Symbol mapping (or de-mapping) involves loading (or unloading) data bits received from the source encoder and the interleaver (or channel equalizer) to (or from) complex subcarrier modulations such as QAM, PSK, and so forth, see Figure 3.27, as **Inverse Discrete Fourier Transform (IDFT).** The output from the mapper constitutes as input to the Inverse Discrete Fourier Transform (IFFT), which accepts complex input data. In IDFT, data is parallelized then treated as samples in the frequency domain. The IDFT process transforms these into time domain signals. Rectangular time-domain pulse shaping spectra of the subcarriers become a cardinal sine function or sync function in the frequency domain.

If number of sub-carriers NC is chosen as a power of 2 (2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048), the IDFT can be replaced by an IFFT, yielding a very efficient

implementation of a OFDM modulator (FFT for demodulator at receiver). For example, 8-PSK, which has 8 Phase Shift Keying, has three bits per sub-carrier per symbol, see Figure 3.26.

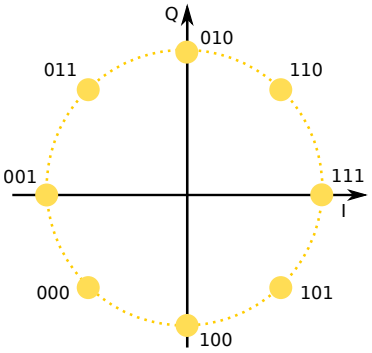


Figure 3.26: 8-PSK Constellation Diagram.

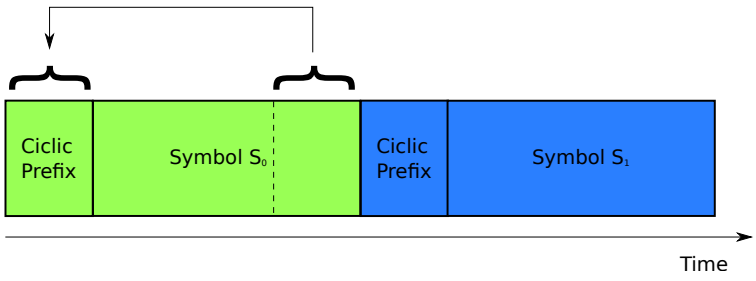


Figure 3.27: Cyclic prefix.

Equalization. The primary advantage of OFDM over mono-carrier schemes is its ability to cope with severe attenuation across channel frequencies such as small or large-scale fading (discussed above in Section 3.6.2) using a simplified equalization scheme. Equalization helps attenuate or adjust the balance between frequency components to flatten channel response, supporting the removal of frequency selective fading effects. This is achieved by:

- Insertion of known symbols (pilots) in the OFDM frame.
- Evaluating their distortions at the receiver.
- Assuming a relatively static channel, data symbols can be equalized.

In OFDM, each carrier becomes an infinite sinusoid (i.e. eigenfunction). As a result, the out of channel is a scaled version of the same function. The eigenvalues of the (circular) channel are the complex scalar terms that multiply each carrier. Thus symbols only experience magnitude and phase change, which makes equalization simple. Convolution in the time domain corresponds to multiplication in the frequency domain. However, this fact does not hold in discrete time. Circular convolution in the (discrete) time domain corresponds to multiplication in the (discrete) frequency domain. OFDM wants simple multiplication in the frequency domain. So, circular convolution is needed and not the regular convolution i.e., the real channel does regular convolution. The solution to this problem is to add a cyclic prefix, so regular convolution can be used to create circular convolution.

Cyclic Prefix. The cyclic prefix is added to the beginning of a symbol and is a repetition of the end of a symbol. Figure 3.27 shows the cyclic prefix added to a symbol over time. Its purpose is to help preserve sinusoids in multipath channels. Sinusoids are eigenfunctions of linear time-invariant channels. The cyclic prefix helps eliminate inter-symbol interference (ISI), which is the delayed replica of previous symbols interfering

with the current symbol. Additionally, they facilitate equalization by transforming linear convolution into circular convolution. Transmission time is limited to N symbols and this property is lost. The cyclic prefix restores this property by "simulating" an infinite-length sinusoid. Looking at the spectrum, Y(f)=X(f)·H(f), if H(f) is not approximately equal for all f, the original signal is destroyed. Cyclic prefix can make OFDM transmissions completely immune to ISI created by multipath propagation when cyclic prefix length T_cp is longer than the delay spread: T_cp $\geq \sigma \downarrow \tau$

Discrete Fourier Transform (DFT). At the receiver, OFDM de-modulation uses Discrete Fourier Transform (DFT) transformation to convert payload received to the frequency domain. Modulation symbols received from the DFT are de-mapped from complex subcarrier modulations such as QAM, PSK, and so forth, to bits, which are inputted into the deinterleaver and the source-decoder blocks.

3.6.5. OFDM Disadvantages: Timing and Frequency Sensitivity

OFDM transmissions are susceptible to timing and frequency offsets. Timing offsets are due to uncertainties of OFDM symbol boundaries, which can cause intersymbol interference, channel interference, and phase offset. Frequency offsets cause inter-carrier interferences (ICI), and a reduction of desired power in data received. Frequency offsets are caused by the Doppler shift or hardware imperfections e.g. imprecise up-down-conversion. This has the effect that operating on different frequency sub-carriers is no longer orthogonal. OFDM needs accurate frequency synchronization.

3.6.6. OFDM Data Rates

Doubling subcarriers in used bandwidth do not double the data rate. See Table 3.3 for comparisons of modulation depth and data rate.

Data Rate	Bandwidth	N	Code Rate	Modulation
6 Mbps	15	48	1/2	BPSK
9 Mbps	15	48	3/4	BPSK
12 Mbps	15	48	1/2	QPSK
18 Mbps	15	48	3/4	QPSK
24 Mbps	15	48	1/2	16-QAM
36 Mbps	15	48	3/4	16-QAM
48 Mbps	15	48	2/3	64-QAM
54 Mbps	15	48	3/4	64-QAM

Table 3.3: OFDM Data Rates and Modulations Depths

3.6.7. FIRE Testbed Environment

This OFDM course runs completely on Trinity College Dublin's (TCD's) IRIS testbed facility, which is located on TCD's campus in Dublin, Ireland. The testbed consists of 16 flexible Universal Software Radio Peripheral (USRP) N210 Ettus Research units aligned in a grid configuration, see Figure 3.28. Each USRP is connected to a virtual machine that runs a software-defined radio (SDR) system. In these experiments, we use the GNU Radio software development toolkit. GNU Radio offers signal-processing

Figure 3.28: Iris wireless laboratory ceiling mounted N210 USRP radio equipment.

blocks that implement software radios. A conceptual diagram of IRIS's virtualized cloud resources, radio hypervisor, user experiments and physical equipment is shown in Figure 3.29. The hardware that you will use will be configured automatically through a process called provisioning. This process will take care of the reservation of two virtual machines in TCDs FIRE testbed facility, the connection of the same to appropriate USRP hardware, installation of required operating system and tools, and initialization of experimentation services. These two virtual machines are under your sole control for use in experimentation. Data for monitoring wireless spectrum is sent to a database on the webserver which is displayed to you via a graph.

Each OFDM experiment requires a virtual machine and a USRP for transmitting a signal and a VM and a USRP for receiving a signal. Variable parameter changes from users are sent from the web interface to GNU Radio, which supports changing frequency, gain, modulation depth, and so forth. The Rx node sends data streams received for frequency, time, waterfall and constellation back to a gateway server for rendering in the web interface.

Testbed Configuration and Tools. Users can experiment with real radio hardware equipment on a Future Internet Research & Experimentation (FIRE) testbed facility [Collins 2016] or [Collins et al. 2016]. TCD/CONNECT has deployed the Smart Reconfigurable Radio Testbed based on the GNU Radio software-defined radio (SDR) system. The Iris testbed supports experimentation with a mature SDR running on a virtualized computational platform. The testbed is organized in experimentation units, each of which consists of three parts: a virtual computational platform, SDR software, and flexible radio front-end hardware. Through this organization, TCD encapsulates the elements required to use the GNU Radio SDR system to construct a broad range of radio

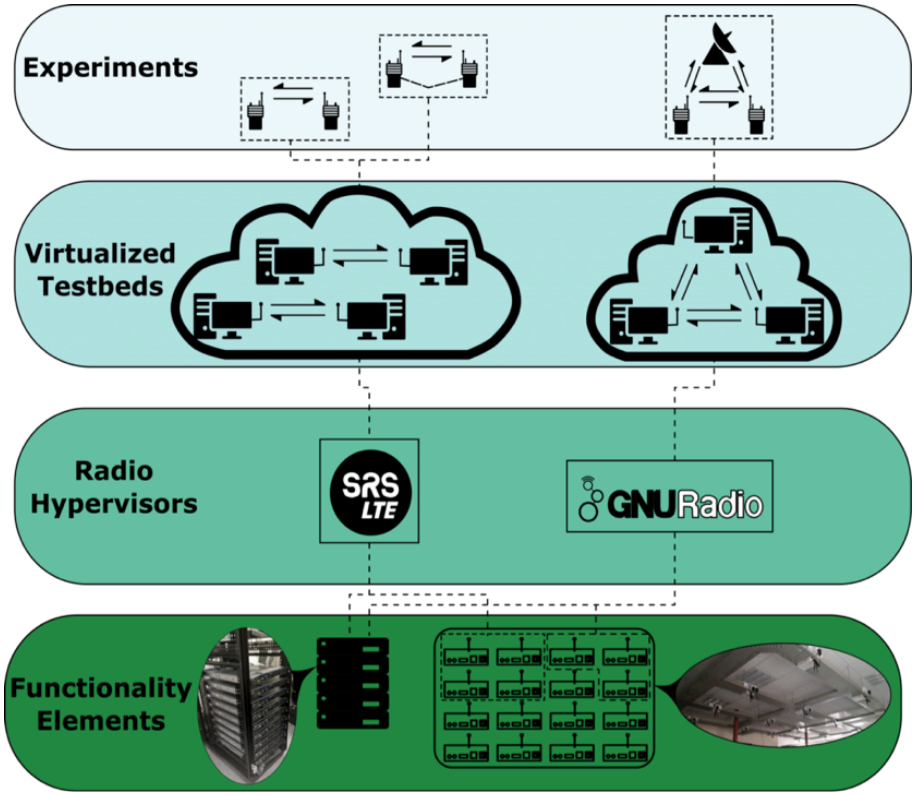


Figure 3.29: Conceptual diagram of IRIS's virtualized cloud resources, radio hypervisor, physical equipment, and user experiments.

systems. Each experimentation unit is designed to flexibly serve a range of needs: Linux (Ubuntu 16.04.01 LTS) provides a highly configurable computation platform, GNU Radio provides real-time radio reconfigurability, and a USRP offers a broad range of wireless interfaces. Radio hardware is housed on the ceiling of the dedicated indoor testing space to provide users with a clean operating environment. The management infrastructure allows users to deploy experimentation units to compose arbitrary radio systems and networks as desired. These facilities have enabled and facilitated several international research and education-related projects.

3.6.8. Exercises

The TCD OFDM course, which allows students to inspect the effect of configuring OFDM concepts and principles using the GNU Radio software radio equipment on the transmitter (Tx) and receiver (Rx) machines, is available at [Collins 2016] or [Collins et al. 2016]. There is no need to investigate every possible combination of configuration parameters. However, after each experiment, you should try to understand the effects and implications of your configuration changes on the OFDM radio and try to answer the accompanying questions.

3.7. Hints for using testbeds

This section will present some hints of how to use testbeds in the most efficient way (how to organize your code, how to run it...).

3.7.1. OFDM - Notes when running GNU Radio experiments

- When initializing experiments, please note that it can take up to ten minutes to provision the Tx and Rx nodes.
- Remember that each configuration change can take up to several seconds to take effect.
- It is important to remember that the transmitter and receiver need to be configured with the same frequency, modulation depth, bandwidth, and cyclic prefix when sending and receiving a signal. This is to give the message or packet the best chance of being received correctly.
- Due to the nature of radio communication - every packet may not be received. Consequently, don't be afraid to send lots of packets.
- After an experiment is launched, resources provisioned will only be available for two hours.

3.7.2. Make an installation script

Considering that during the stage of installation of the modules several commands are repeated, a script can be constructed to aid this process. Using Shell scripts we can automate the installation, assuming that the files will initially be uploaded to the VM or that they will be fetched using git commands.

3.7.3. Avoiding the graphical interface

For various reasons, such as screen freezes, the use of the graphical interface should be avoided when using remote testbeds. Here are some changes that need to be made to prevent a graphical interface from being displayed at runtime.

Before sending the `.grc` file to the remote environment, we need to delete or disable all blocks that instantiate the GUI. All GRC GUI blocks should be avoided or disabled, as shown in Fig. 3.30. In the `Options` block, at the top-left corner of GRC, we need to choose *No GUI* on *Generate Options*.

In the example of Fig. 3.30, `WX GUI Slider` and `WX GUI Chooser` are disabled to prevent the execution of the graphical interface. To disable a block just right click on the block and choose *Disable*. If the block that was disabled provides a variable for use in other GRC blocks, we need to add a `Variable` block with the same variable name that was previously contained in the block that used GUI. In addition, all `WX GUI` and `QT GUI` blocks should be disabled, as well as the all `WX GUI Sink` and `QT GUI Sink` instrumentation blocks. This procedure is required in an existing file or in a totally new project.

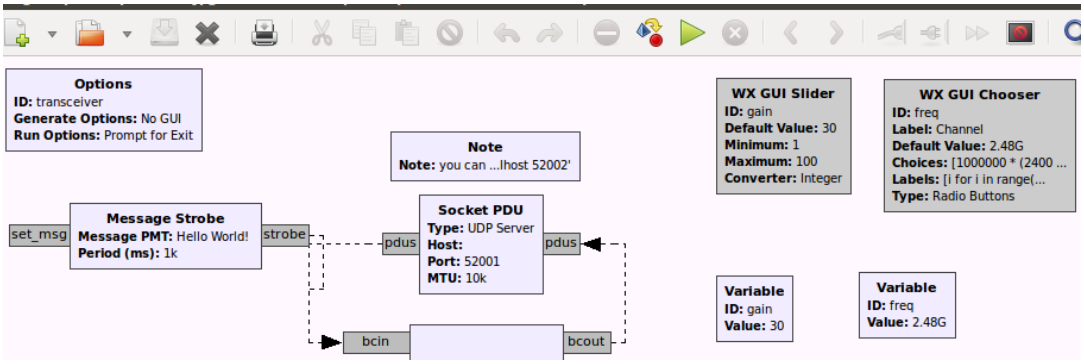


Figure 3.30: Changes on GRC to avoid graphical interfaces

After modifying all graphical GUI instances in GRC, it is important to note if there is any GUI execution in the source code. If so, it should be changed to terminal output.

After modifying the files to not display the graphical interface, the files can be sent to the remote environment and `.grc` can be executed remotely as follows:

```
vm$ grcc -e file.grc
```

If you only need to compile the file, then it is necessary to use the command:

```
vm$ grcc file.grc
```

Finally, all debug checks must be performed by the output terminal, for example, by using in the source code a `print "Message"` in case of Python, or in case of C language using `printf("Message")`.

3.7.4. Using the RSpec editor

In JFed, using the RSpec editor, we can create or modify an experiment and add new features. At the XML file, as shown in the following, we can change the fields *client_id* (line number 3), *component_manager* (line 3), *silver_type name* (line 4) and *disk_image name* (line 5).

```
1 <?xml version='1.0'?>
2 <rspec xmlns="http://www.geni.net/resources/rspec/3" type="request"
    generated_by="jFed RSpec Editor" generated="2017-02-03T16:50:46
    .711-02:00" xmlns:emulab="http://www.protogeni.net/resources/rspec/
    ext/emulab/1" xmlns:jfedBonfire="http://jfed.iminds.be/rspec/ext/
    jfed-bonfire/1" xmlns:delay="http://www.protogeni.net/resources/
    rspec/ext/delay/1" xmlns:jfed-command="http://jfed.iminds.be/rspec/
    ext/jfed-command/1" xmlns:client="http://www.protogeni.net/
    resources/rspec/ext/client/1" xmlns:jfed-ssh-keys="http://jfed.
    iminds.be/rspec/ext/jfed-ssh-keys/1" xmlns:jfed="http://jfed.iminds
    .be/rspec/ext/jfed/1" xmlns:sharedvlan="http://www.protogeni.net/
    resources/rspec/ext/shared-vlan/1" xmlns:xsi="http://www.w3.org
    /2001/XMLSchema-instance" xsi:schemaLocation="http://www.geni.net/
    resources/rspec/3 http://www.geni.net/resources/rspec/3/request.xsd
     ">
3   <node client_id="node01" exclusive="false" component_manager_id="
      urn:publicid:IDN+futebol.dcc.ufmg.br+authority+am">
4     <sliver_type name="usrp-vm">
```

```
5       <disk_image name="urn:publicid:IDN+futebol.dcc.ufmg.br+image+
            gnuradio"/>
6     </sliver_type>
7     <location xmlns="http://jfed.iminds.be/rspec/ext/jfed/1" x="410.0"
          y="85.0"/>
8   </node>
9   <node client_id="node02" exclusive="false" component_manager_id="
        urn:publicid:IDN+futebol.dcc.ufmg.br+authority+am">
10    <sliver_type name="usrp-vm">
11      <disk_image name="urn:publicid:IDN+futebol.dcc.ufmg.br+image+
            gnuradio"/>
12    </sliver_type>
13    <location xmlns="http://jfed.iminds.be/rspec/ext/jfed/1" x="10.0" y
          ="85.00000000000001"/>
14  </node>
15  <jfed-command:experimentBarrierSegment orderNumber="0" tag="Barrier
        segment 0"/>
16 </rspec>
```

The options to those fields are:

client_id: any nick;

component_manager: urn:publicid:IDN+futebol.dcc.ufmg.br+authority+am ;

sliver_type name: choose between available slivers: usrp-vm, iot-vm and raw-raspberry;

disk_image name: each sliver has a disk available, which depends on the testbed that you are using.

3.7.4.1. Adding one more node

In the `.rspec` file presented earlier, there are 2 nodes, called *node01* (see line 3) and *node02* (line 9). If we intend to add another node, we can copy the whole stretch of line 9 to line 14, modifying the node identifier (`node client_id`) and the `location xmlns` field, to avoid overlapping nodes in the jFed graphical environment.

3.8. Conclusions and further readings

This chapter presented how to use Software-Defined Radios (SDR) to conduct wireless research. SDRs are wireless transceivers that are able to run a number of wireless protocols, since they are implemented in software, not in hardware. Further, the GNU Radio project provides a number of algorithms and protocols that can be used to implement new protocols. SDRs are an important tool for wireless researchers because it allows the creation of new protocols, which can be tested in a realistic situation. By experimenting with a real device, the proposal is evaluated under more realistic settings than those found in most simulators, for example. As a consequence, the research becomes more relevant, and the gap from research to mass dissemination of the technology becomes shorter.

Although SDR devices are not cheap, it is possible to perform research using real hardware very easily by remotely using those resources, for free, over the Internet. There

are a number of testbeds spread all over the world, including testbeds in Brazil, which have USRPs available for researchers. In this short course, we focused on the FUTEBOL federation of testbeds, however other testbeds could be used.

In order to show that SDR is relevant for research in wireless communications and wireless networking, this chapter presented four simple experiments that highlight the capabilities of such a platform. Those experiments range from WBAN to cellular, from modulation techniques to MAC protocols, and provide a glimpse of the versatility of SDR. Further, all the code used in the experiments is available for use and modification by other researchers.

In the future, we expect more and more papers to be written using SDRs as their platform. Although today it may be a bit hard to find full stack implementations for some popular wireless standards, this limitation is quickly being addressed by the community. As a consequence, the complexity of building experiments with SDRs will go down with time. There is a push in the networking and telecommunications community towards experimental research, so wireless researchers should be aware of SDRs, how to use them and what are their limitations. Even if you do not plan to use it today for your experiments, you might need to use them in the near future.

Acknowledgements

The authors of this chapter have been funded by CAPES, CNPq, Fapemig and the FUTEBOL project. FUTEBOL has received funding from the European Union's Horizon 2020 for research, technological development, and demonstration under grant agreement no. 688941 (FUTEBOL), as well from the Brazilian Ministry of Science, Technology and Innovation (MCTI) through RNP and CTIC.

References

[OAI 2017] (2017). OpenAirInterface | 5G software alliance for democratising wireless innovation. `http://www.openairinterface.org`.

[ope 2017] (2017). OpenBTS | open source cellular infrastructure. `http://openbts.org`.

[Akyildiz et al. 2008] Akyildiz, I., Lee, W.-Y., Vuran, M. C., and Mohanty, S. (2008). A survey on spectrum management in cognitive radio networks. *Communications Magazine, IEEE*, 46(4):40–48.

[Alt-Shift-X 2013] Alt-Shift-X (2013). The doppler effect: what does motion do to waves? `https://youtu.be/h4OnBYrbCjY`.

[Amiri et al. 2007] Amiri, K., Sun, Y., Murphy, P., Hunter, C., Cavallaro, J., and Sabharwal, A. (2007). WARP, a unified wireless network testbed for education and research. In *Microelectronic Systems Education, 2007. MSE '07. IEEE International Conference on*, pages 53–54.

[Beyene et al. 2014] Beyene, Y. D., Jäntti, R., and Ruttik, K. (2014). Cloud-ran architecture for indoor das. *IEEE Access*, 2:1205–1212.

[Bharadia et al. 2013] Bharadia, D., McMilin, E., and Katti, S. (2013). Full duplex radios. In *Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM*, SIGCOMM '13, pages 375–386, New York, NY, USA. ACM.

[Bloessl et al. 2013] Bloessl, B., Leitner, C., Dressler, F., and Sommer, C. (2013). A GNU Radio-based IEEE 802.15. 4 Testbed. *12. GI/ITG FACHGESPRÄCH SENSORNETZE*, page 37.

[Blossom 2004] Blossom, E. (2004). Gnu radio: tools for exploring the radio frequency spectrum. *Linux journal*, 2004(122):4.

[Busch et al. 2004] Busch, C., Magdon-Ismail, M., Sivrikaya, F., and Yener, B. (2004). Contention-free MAC protocols for wireless sensor networks. In *International Symposium on Distributed Computing*, pages 245–259.

[Collins 2016] Collins, D. (2016). Connect smart reconfigurable radio testbed. `https://iris-testbed.connectcentre.ie/ofdm_v2/login.php`.

[Collins et al. 2016] Collins, D., Barja, J. M., Kaminski, N., Blumm, C., Silva, L. D., Sutton, P., and Gomez, I. (2016). Introduction to orthogonal frequency-division multiplexing (OFDM) modulation method. `http://www.forgebox.eu/fb/preview_course.php?course_id=180`.

[Commission 2003] Commission, F. C. (2003). *FCC 03-322*. FCC.

[Cordeiro 2017] Cordeiro, J. R. S. (2017). FS-MAC: um sistema para flexibilização da subcamada MAC em redes sem fio.

[Cordeiro et al. 2017] Cordeiro, J. R. S., Lanza, E., Macedo, D. F., and Vieira, L. F. M. (2017). Fs-mac: Uma plataforma para a flexibilização da sub-camada mac em redes sem fio. In *XXXV Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*.

[Correia et al. 2015] Correia, L. H., Tran, T.-D., Pereira, V. N., Giacomin, J. C., and Sá Silva, J. M. (2015). Dynmac: A resistant mac protocol to coexistence in wireless sensor networks. *Computer Networks*, 76(Complete):1–16.

[Diepstraten and WCND-Utrecht 1993] Diepstraten, W. and WCND-Utrecht, N. (1993). IEEE 802.11 wireless access method and physical specification. *Power*, 5:10.

[Dillinger et al. 2003] Dillinger, M., Madani, K., and Alonistioti, N. (2003). *Software Defined Radio: Architectures, Systems and Functions*. Wiley & Sons.

[Ettus 2017] Ettus (2017). Ettus Research. `http://www.ettus.com`.

[Forum 2011] Forum, W. I. (2011). Software defined radio - rate of adoption. `http://www.wirelessinnovation.org/sdr_rate_of_adoption`.

[Gilmore and Blossom 2017] Gilmore, J. and Blossom, E. (2017). GNU Radio - the free and open software radio system. `http://gnuradio.org/redmine/projects/gnuradio/wiki/`.

[Gollakota et al. 2011] Gollakota, S., Hassanieh, H., Ransford, B., Katabi, D., and Fu, K. (2011). They can hear your heartbeats: Non-invasive security for implantable medical devices. In *Proceedings of the ACM SIGCOMM 2011 Conference*, SIGCOMM '11, pages 2–13, New York, NY, USA. ACM.

[Gomez 2013] Gomez, O. M. (2013). Implementation of the ofelia control framework (ocf) for open flow-based testbed facilities. Master's thesis, Universitat Politècnica de Catalunya (UPC).

[Gudipati and Katti 2011] Gudipati, A. and Katti, S. (2011). Strider: automatic rate adaptation and collision handling. In *Proceedings of the ACM SIGCOMM 2011 conference*, SIGCOMM '11, pages 158–169, New York, NY, USA. ACM.

[Hong et al. 2012] Hong, S. S., Mehlman, J., and Katti, S. (2012). Picasso: flexible RF and spectrum slicing. *SIGCOMM Comput. Commun. Rev.*, 42(4):37–48.

[Hu et al. 2009] Hu, W., Li, X., and Yousefi'zadeh, H. (2009). La-mac: A load adaptive mac protocol for manets. In *GLOBECOM 2009 - 2009 IEEE Global Telecommunications Conference*, pages 1–6.

[Huawei 2014] Huawei (2014). Huawei Learning Service Express OFDM. `https://youtu.be/tPQ_ahjCujY`.

[Iannucci et al. 2012] Iannucci, P. A., Perry, J., Balakrishnan, H., and Shah, D. (2012). No symbol left behind: a link-layer protocol for rateless codes. In *Proceedings of the 18th annual international conference on Mobile computing and networking*, Mobicom '12, pages 17–28, New York, NY, USA. ACM.

[Jagannath et al. 2015] Jagannath, J., Saarinen, H. M., and Drozd, A. L. (2015). Framework for automatic signal classification techniques (fact) for software defined radios. In *CISDA*, pages 1–7.

[Katti et al. 2008] Katti, S., Rahul, H., Hu, W., Katabi, D., Médard, M., and Crowcroft, J. (2008). XORs in the air: practical wireless network coding. *IEEE/ACM Trans. Netw.*, 16(3):497–510.

[Kumar et al. 2013] Kumar, S., Cifuentes, D., Gollakota, S., and Katabi, D. (2013). Bringing cross-layer MIMO to today's wireless LANs. In *Proceedings of the ACM SIGCOMM 2013 conference*, SIGCOMM '13, pages 387–398, New York, NY, USA. ACM.

[Li and Qiu 2010] Li, H. and Qiu, R. C. (2010). A graphical framework for spectrum modeling and decision making in cognitive radio networks. In *2010 IEEE Global Telecommunications Conference GLOBECOM 2010*, pages 1–6.

[Lin et al. 2008] Lin, K. C.-J., Kushman, N., and Katabi, D. (2008). ZipTx: Harnessing partial packets in 802.11 networks. In *Proceedings of the 14th ACM international conference on Mobile computing and networking*, MobiCom '08, pages 351–362, New York, NY, USA. ACM.

[Marques et al. 2016] Marques, A. F. F., Miranda, G., Silva, L. M., Ávila, R. S., and Correia, L. H. A. (2016). Iscra - an intelligent sensing protocol for cognitive radio. In *IEEE – ISCC*, pages 385–390.

[McHenry et al. 2006] McHenry, M. A., Tenhula, P. A., McCloskey, D., Roberson, D. A., and Hood, C. S. (2006). Chicago spectrum occupancy measurements & analysis and a long-term studies proposal. In *Proceedings of the first international Workshop on Technology and policy for accessing spectrum (TAPAS), 2006*, pages 1–12.

[Mitola 1999] Mitola, J. (1999). Cognitive radio : model-based competence for software radios. NR 20140804.

[Movassaghi et al. 2014] Movassaghi, S., Abolhasan, M., Lipman, J., Smith, D., and Jamalipour, A. (2014). Wireless body area networks: A survey. *IEEE Communications Surveys & Tutorials*, 16(3):1658–1686.

[Murphy et al. 2006] Murphy, P., Sabharwal, A., and Aazhang, B. (2006). Design of warp: A wireless open-access research platform. In *European Signal Processing Conference*.

[Neufeld et al. 2005] Neufeld, M., Fifield, J., Doerr, C., Sheth, A., and Grunwald, D. (2005). Softmac-flexible wireless research platform. In *Proc. HotNets-IV*, pages 1–5.

[Nychis et al. 2009] Nychis, G., Hottelier, T., Yang, Z., Seshan, S., and Steenkiste, P. (2009). Enabling MAC Protocol Implementations on Software-defined Radios. In *6th USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, pages 91–105.

[Ó Coileáin 2016] Ó Coileáin, D. (2016). Multipath fading. `https://youtu.be/1rcCLfdR5qs`.

[Perry et al. 2012] Perry, J., Iannucci, P. A., Fleming, K. E., Balakrishnan, H., and Shah, D. (2012). Spinal codes. In *Proceedings of the ACM SIGCOMM 2012 conference on Applications, technologies, architectures, and protocols for computer communication*, SIGCOMM '12, pages 49–60, New York, NY, USA. ACM.

[Rao and Stoica 2005] Rao, A. and Stoica, I. (2005). An overlay MAC layer for 802.11 networks. In *Proceedings of the 3rd international conference on Mobile systems, applications, and services*, pages 135–148.

[RTL-SDR] RTL-SDR. Rtl-sdr.com. `http://www.rtl-sdr.com/`.

[Saltzberg 1967] Saltzberg, B. (1967). Performance of an efficient parallel data transmission system. *IEEE Transactions on Communication Technology*, 15(6):805–811.

[Saucier 2000] Saucier, R. (2000). Computer generation of statistical distributions. Approved for public release; distribution is unlimited.

[Shokrollahi 2006] Shokrollahi, A. (2006). Raptor codes. *IEEE/ACM Trans. Netw.*, 14(SI):2551–2567.

[Silva et al. 2015] Silva, W. S., Cordeiro, J. R. S., Macedo, D. F., Vieira, M. A. M., Vieira, L. F. M., and Nogueira, J. M. S. (2015). Introdução a Rádios Definidos por Software com Aplicações em GNU Radio. In *Minicursos do XXXIII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, chapter 5, pages 216–265. Sociedade Brasileira de Computação.

[Souryal et al. 2015] Souryal, M., Ranganathan, M., Mink, J., and Ouni, N. E. (2015). Real-time centralized spectrum monitoring: Feasibility, architecture, and latency. In *2015 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, pages 106–112.

[Takagi and Kleinrock 1985] Takagi, H. and Kleinrock, L. (1985). Throughput analysis for persistent CSMA systems. *IEEE Transactions on Communications*, 33(7):627–638.

[Tan et al. 2009] Tan, K., Zhang, J., Fang, J., Liu, H., Ye, Y., Wang, S., Zhang, Y., Wu, H., Wang, W., and Voelker, G. M. (2009). Sora: High performance software radio using general purpose multi-core processors. In *USENIX International Symposium on Networked Systems: Design and Implementation (NSDI)*, pages 75–90.

[Tinnirello et al. 2012] Tinnirello, I., Bianchi, G., Gallo, P., Garlisi, D., Giuliano, F., and Gringoli, F. (2012). Wireless MAC processors: Programming MAC protocols on commodity hardware. In *IEEE INFOCOM*, pages 1269 –1277.

[Vieira et al. 2013] Vieira, L. F. M., Gerla, M., and Misra, A. (2013). Fundamental limits on end-to-end throughput of network coding in multi-rate and multicast wireless networks. *Computer Networks*, 57(17):3267–3275.

[Wireless Innovation Forum 2017] Wireless Innovation Forum (2017). Wireless innovation forum. `http://www.wirelessinnovation.org`.

[Yucek and Arslam 2009] Yucek, T. and Arslam, H. (2009). A Survey of Spectrum Sensing Algorithms for Congnitive Radio Applications. *Proceedings of the IEEE*, 97(5):805–823.

[Zhou et al. 2006] Zhou, G., Stankovic, J. A., and Son, S. H. (2006). Crowded spectrum in wireless sensor networks. Workshop on Embedded Networked Sensors.

[Ziouva and Antonakopoulos 2002] Ziouva, E. and Antonakopoulos, T. (2002). CSMA/CA performance under high traffic conditions: throughput and delay analysis. *Computer Communications*, 25(3):313 – 321.

Capítulo 4

Fiscalização da Neutralidade da Rede: Conceitos e Técnicas

Ligia E. Setenareski (UFPR), Thiago Garrett (UFPR), Letícia M. Peres (UFPR), Luis C. E. Bona (UFPR), Elias P. Duarte Jr. (UFPR)

Abstract

Network Neutrality (NN) is becoming increasingly important as the global debate intensifies and governments worldwide implement regulations. According to NN, all types of traffic must be processed without discrimination, regardless of origin, destiny and/or content. The discrimination between different types of traffic compromises innovation, fair competition and the freedom of choice of consumers. However, ensuring that ISPs are not employing discriminating practices is still a challenge. This tutorial presents an overview of several existing solutions to detect "traffic differentiation". These solutions differ mainly on the monitoring topology, metrics and statistical methods employed. An introduction to the global debate around NN is also presented, as well as an overview of different regulations defined in Brazil and other countries around the world.

Resumo

A Neutralidade da Rede (NR) torna-se cada vez mais importante à medida que o debate sobre este princípio se intensifica, levando mais países a promoverem sua normatização. Segundo a NR, todo tipo de tráfego deve ser tratado da mesma forma, independente de sua origem, destino e/ou conteúdo. Discriminar tipos diferentes de tráfegos de dados compromete a inovação, concorrência justa e liberdade de escolha dos consumidores. Porém, fiscalizar se provedores de acesso estão praticando alguma diferenciação de tráfego ainda é um desafio. Este minicurso tem como objetivo principal apresentar diversas soluções para a detecção destas práticas. Estas soluções diferem principalmente na topologia utilizada para medição, nas métricas empregadas e nos métodos estatísticos utilizados. Também é apresentada uma introdução ao debate da NR, bem como um panorama da sua normatização no Brasil e no mundo.

Todos os autores são filiados à Universidade Federal do Paraná (UFPR). Emails: {ligia,tgarrett,lmperes,bona,elias}@inf.ufpr.br

4.1. Introdução

A importância da Internet na sociedade moderna tem aumentado significativamente, conforme a quantidade de usuários e serviços disponíves na rede cresce [K.G. Coffman 2002]. Adequar e manter a estrutura da rede para atender esta demanda crescente é um desafio, em especial porque além dos aspectos tecnológicos, devem ser considerados também aspectos econômicos. Os provedores de acesso (*Internet Service Providers*, ISPs) podem empregar técnicas de gerência de tráfego de rede para reduzir e/ou postergar investimentos na infraestrutura de rede [van Schewick and Farber 2009]. Entretanto, muitas destas técnicas de gerência de tráfego podem ser consideradas discriminatórias e podem ser utilizadas para outros fins, como para obter vantagens competitivas ou cobrar taxas extras de usuários e provedores de conteúdo.

Práticas discriminatórias consistem em manipular o tráfego de dados de forma a priorizar ou degradar algum tipo específico [Ravaioli et al. 2012] – baseando-se no tipo de informação sendo trafegada ou no destino/origem dos pacotes, por exemplo. Este tipo de manipulação configura a chamada diferenciação de tráfego (DT). Em geral a DT é empregada por um ISP devido a três fatores: (i) congestionamento, no qual limita-se a largura de banda utilizada por aplicativos que geram muito tráfego, como compartilhamento de arquivos P2P e *streaming* de vídeo [Mueller and Asghari 2012]; (ii) acordos comerciais, em que provedores de serviço pagam taxas extras para ter seu tráfego priorizado pelo ISP, as chamadas *fast-lanes* [Habibi Gharakheili et al. 2016]; e (iii) obtenção de vantagem competitiva, em que o tráfego do próprio ISP é priorizado ou o tráfego de concorrentes é degradado [Kendrick 2009, Lomas 2016].

A DT faz parte de um longo e controverso debate mundial sobre o princípio da Neutralidade da Rede (NR) [Joch 2009]. Este princípio já foi instituído em diversos países do mundo por meio de leis, diretrizes, regras e/ou princípios [Habibi Gharakheili et al. 2016]. Uma definição da NR comum às estas diversas normatizações diz que, em uma rede neutra, todo tipo de tráfego deve ser tratado da mesma forma, sem distinção por origem, destino e/ou conteúdo, ou seja, a DT não é permitida [Crowcroft 2007].

Para os defensores da NR, a DT ameaça três conceitos que foram fundamentais para o sucesso da Internet: inovação, concorrência justa e liberdade de escolha dos consumidores [Berners-Lee 2010]. Em um mundo sem NR, um ISP poderia ter controle sobre quais serviços teriam mais chances de serem consumidos pelos usuários e quais serviços teriam maior chance de serem bem sucedidos [van Schewick and Farber 2009].

A inovação em uma Internet não-neutra seria conduzida pelos ISPs e pelas grandes corporações [van Schewick and Farber 2009], as quais teriam recursos suficientes para ter seus dados priorizados. Novos serviços e soluções inovadoras encontrariam dificuldades para obter sucesso, já que não teriam capacidade de competir em iguais condições com os serviços já bem estabelecidos no mercado [Guo and Easley 2016, Cooper and Brown 2015]. Por outro lado, os mais conservadores afirmam que a falta de controles restritivos adicionais sobre os ISPs configura um mercado mais competitivo [Joch 2009]. Portanto, garantir que a Internet continue a ser um ambiente que fomente inovação é o tema central do debate da NR [Weitzner 2008].

Entretanto, apenas a existência das normatizações da NR não garante que os

ISPs irão respeitá-las. Além disso, podem haver práticas discriminatórias não previstas pelas normatizações [Knutson and Ramachandran 2016]. Assim, é importante a criação de soluções que auxiliem na fiscalização das normatizações da NR, aumentando a transparência das práticas de gerência de tráfego empregadas pelos ISPs. Estas soluções devem detectar a ocorrência de violações da NR, ou seja, a presença de práticas de DT.

Porém, a detecção de DT ainda é um desafio [Tariq et al. 2009]. Uma das dificuldades encontradas deve-se às diversas diferentes formas que ISPs podem implementar a DT. Um tráfego pode ser discriminado baseado no protocolo utilizado, no destino, origem ou conteúdo das mensagens, por exemplo. Além disso, diversas técnicas podem ser empregadas, como engenharia de tráfego (*traffic shaping*), diferentes rotas internas e vigilância de tráfego (*traffic policing*). Outro desafio é descobrir em qual ISP entre a origem e o destino do tráfego a diferenciação está ocorrendo, já que não se tem nenhum conhecimento prévio sobre a estrutura interna da rede. Além disso, diversos outros fatores além da DT podem afetar o desempenho de um tráfego de dados na Internet, como congestionamento, tráfego de fundo e balanceamento de carga, os quais podem ser mal interpretados como DT.

Este minicurso tem como objetivo principal apresentar soluções existentes para a detecção de DT. São apresentados a definição do problema, as técnicas utilizadas pelas soluções existentes, assim como o funcionamento, requisitos e limitações de cada solução. Além disso, o minicurso também apresenta uma visão geral do debate em torno da NR e das normatizações já implantadas no Brasil e em diversos países do mundo. A maioria das soluções existentes para detecção de DT são baseadas em medições de rede e inferência estatística. Em geral, estas soluções efetuam medições a partir de um ou diversos *hosts* e utilizando diferentes tipos de tráfego. Os dados obtidos são então comparados para inferir se houve ou não uma diferença significativa entre conjuntos diferentes de medições. Modelos estatísticos robustos são necessários para diferenciar variações causadas por DT das causadas por outros fenômenos, como congestionamento, entre outros. Dentre as soluções descritas neste minicurso, diversas geram tráfegos artificiais, correspondentes a diferentes aplicações entre dois *hosts* e comparam o desempenho fim-a-fim destes tráfegos. Já outras soluções obtém medições a cada *hop* do caminho entre os *hosts*, com o objetivo de identificar exatamente onde a DT ocorre. Há ainda soluções que capturam passivamente os tráfegos de diferentes aplicações, comparando-os posteriormente.

O restante deste minicurso está organizado da seguinte maneira. Primeiramente, apresentamos na seção 4.2, como motivação do trabalho, uma linha do tempo com diversos casos reais de violações da NR ocorridos em vários lugares do mundo nos últimos anos. Em seguida, a seção 4.3 apresenta uma visão geral do debate sobre a NR e das normatizações ao redor do mundo. A seção 4.4 apresenta uma fundamentação sobre como a DT é empregada na Internet e define o que é uma rede neutra e o problema de detecção de DT. A seção 4.5 descreve diversas soluções já existentes para o problema, seus requisitos e limitações. São apresentadas também uma comparação destas soluções e as diversas técnicas que podem ser utilizadas na detecção de DT, extraídas das soluções existentes. Concluímos então o minicurso na seção 4.6.

4.2. Casos Reais de Violações da NR

Denúncias de casos de violações da NR tornam-se cada vez mais comuns à medida que cresce a quantidade de usuários da Internet e cada vez mais países implementam a NR em seus territórios, fomentando o debate sobre o tema. Esta seção apresenta, em ordem cronológica, diversos casos reais de violações da NR ao redor do mundo. Estes casos consistem não apenas de trabalhos científicos, mas também de denúncias de usuários e da imprensa.

Começamos por um caso típico de violação da NR por meio do bloqueio de páginas Web. Em 21 de julho de 2005, os membros do sindicato canadense dos trabalhadores de telecomunicações, a *Telecommunications Workers Union* (TWU), entraram em greve contra a Telus, um ISP do país. Em 22 de julho (dia seguinte), a operadora Telus bloqueou o acesso de seus usuários à página Web *Voices for Change*, dirigida por e para os membros da TWU, alegando que o seu contrato de serviço com os usuários lhe permitia bloquear qualquer página Web [Austen 2005]. Em 28 de julho, a operadora Telus libera novamente o acesso à página Web após receber uma liminar.

Em 24 de julho de 2007, foi lançada a ferramenta Web Tripwires, que detecta modificação de conteúdo em páginas Web [Reis et al. 2008]. Os dados coletados nos primeiros 20 dias de funcionamento da ferramenta mostraram que ISPs provocaram mudanças intencionais no tráfego de 46 dos 50171 *hosts* medidos, entre outros resultados.

Também em 2007, em um fórum de discussões da página Web DSLReports [Topolski 2007], Topolski relata que a operadora Comcast utilizou equipamentos da Sandvine [Sandvine] para controlar sessões de comunicação de aplicativos P2P. Segundo Topolski, o equipamento da Sandvine verificava todos os pacotes que ingressam na rede da Comcast. Caso o tráfego de pacotes referentes a aplicativos P2P fosse maior que um limite estabelecido pelo ISP, o equipamento passava a interromper os fluxos de tais aplicativos. Topolski afirma que estas interrupções foram feitas por meio de pacotes forjados do tipo *reset* (RST) do protocolo TCP injetados no fluxo de comunicação dos aplicativos.

Em abril de 2009, Kendrick afirmou na página Web Gigaom [Kendrick 2009] que a operadora T-Mobile da Alemanha estava bloqueando o uso do aplicativo Skype em todas as suas redes, fato confirmado pela própria operadora. A operadora T-Mobile afirma que os motivos para bloquear todo tráfego VoIP em suas redes são apenas técnicos e não econômicos. Segundo a operadora, o elevado tráfego do aplicativo prejudicaria o desempenho da rede e caso o aplicativo passasse a não funcionar corretamente, os consumidores culpariam a T-Mobile.

Em junho de 2009, a *British Telecommunications* (BT), a empresa britânica de telecomunicações, foi acusada de degradar todo o tráfego de vídeos da página Web da emissora de TV britânica BBC, limitando a largura de banda máxima [Cellan-Jones 2009]. A BT alegou apenas que todas as suas práticas de gerência de tráfego buscam otimizar a experiência de todos os consumidores.

Em 2010, os autores em [Ling et al. 2010] apresentam os argumentos contra e a favor da NR, utilizando como exemplo o caso de bloqueio de serviços P2P efetuado pela operadora Comcast. Os autores afirmam que os serviços P2P não prejudicam a qualidade da Internet, apenas transferem a necessidade de investimento dos provedores de conteúdo

para os ISPs. Segundo os autores, o único dano causado pelos aplicativos P2P em uma rede neutra é aos ISPs, que não podem cobrar taxas extras de provedores de conteúdo para trafegar seus dados.

Em fevereiro de 2011 foi formada a GreatFire [GreatFire.org], uma organização sem fins lucrativos que monitora e publica o estado das páginas Web e palavras-chave censuradas na China por meio do chamado "Grande Firewall da China" (*Great Firewall of China*). A página Web da organização ajuda usuários chineses da Internet a acessar alguns conteúdos bloqueados, a testar suas conexões e publica os dados do monitoramento de páginas e palavras-chave bloqueadas. Dos 49720 domínios monitorados, por exemplo, 4329 são bloqueados na China, entre outras informações disponíveis.

Em abril de 2011, os autores da ferramenta CensMon [Sfakianakis et al. 2011], de detecção de censura, conduziram um experimento no PlanetLab. Foram utilizados 174 *hosts* do *testbed*, localizados em 33 países diferentes. A duração do experimento foi de 14 dias. Neste período a ferramenta testou 4950 endereços Web em 2500 domínios. Foram detectados 951 endereços e 193 domínios filtrados. A maior parte dos domínios bloqueados (176) foram detectados pelo *host* localizado na China.

A página Web europeia *Respect My Net* [Respect My Net] foi lançada em 22 de setembro de 2011. O objetivo desta página é permitir que usuários da Internet relatem violações da NR. A página contém uma lista de todos os casos relatados, com confirmações e provas fornecidas pelos usuários. Os casos não considerados como violações da NR – de acordo com as diretrizes da página – são removidos. A lista conta com um total de 219 relatos confirmados, que envolvem 18 países da Europa e 71 ISPs. Entre os casos relatados, três, tiveram um grande impacto: (i) degradação do tráfego do serviço YouTube na França, pelo ISP Free, com confirmação de 431 pessoas; (ii) bloqueio DNS à página Web *thepiratebay.org* na Bélgica, pela operadora Mobile Vikings, com confirmação de 18 pessoas; e (iii) bloqueio da porta 25 para todos os serviços SMTP pela operadora Belgacom, na Bélgica, com exceção do seu próprio serviço, com confirmação de 21 pessoas.

Em 2012 foi publicado um estudo sobre 2 casos de violações da NR utilizando a técnica de *Deep Packet Inspection* (DPI) nos E.U.A. e no Canadá [Mueller and Asghari 2012]. Nestes casos, ISPs destes países bloqueavam ou degradavam o tráfego de aplicativos P2P, gerando protestos, processos jurídicos, entre outros. O estudo descreve o impacto das práticas de DPI nos aspectos políticos e econômicos que envolvem a Internet (como inovação, competitividade e transparência, por exemplo). Os autores afirmam que utilizaram dados obtidos pela ferramenta Glasnost para o estudo.

Os autores da ferramenta Adkintun [Bustos-Jiménez et al. 2013], descrita mais à frente na seção 4.5, apresentam em [Bustos-Jiménez and Fuenzalida 2014] três casos referentes à utilização da ferramenta no Chile, entre os anos de 2011 e 2013. Em um destes casos, a ferramenta foi utilizada, a pedido do órgão regulador do país (SUBTEL), para avaliar o comportamento de dois ISPs chilenos, VTR e Movistar, que juntos controlam em torno de 80% dos serviços de banda larga no Chile. Os resultados mostraram que a velocidade de *download* durante o período da noite, para as duas operadoras, foi significativamente abaixo do contratado pelos usuários. No segundo caso, o canal estatal de televisão do país noticiou que o número de reclamações de usuários para a SUBTEL aumentou significativamente após o lançamento da Adkintun, assim como a qualidade de

serviço entregue pelos ISPs. O terceiro caso descrito trata de um processo contra a SUBTEL que acusa o órgão de não ter tomado medidas contra ISPs chilenos que não estavam cumprindo todas as exigências da Lei da NR. Esta acusação foi embasada nos dados coletados e publicados pela ferramenta Adkintun, mantida pela própria SUBTEL. Segundo os autores, este foi o primeiro caso no qual a infraestrutura de uma instituição governamental, voltada para garantir a NR, foi utilizada contra a mesma. Os autores afirmam ainda que a Adkintun foi totalmente implantada e tem coletado dados desde setembro de 2011 e já foi utilizado por mais de 10000 usuários.

A ferramenta HAKOMetar [Weber et al. 2013], descrita posteriormente na seção 4.5, foi utilizada por usuários finais na Croácia entre novembro de 2012 e março de 2013, período no qual o número total de medições excedeu 25000. Os resultados destas medições identificaram dezenas de casos em que largura de banda entregue aos usuários foi significativamente menor do que a contratada. Os autores relatam que estas medições motivaram reclamações dos usuários contra 3 das 16 operadoras medidas. Os dados obtidos pela HAKOMetar foram anexados a estas reclamações, as quais tiveram resultados positivos para os usuários.

Em 2013, Anderson descreve um estudo sobre a degradação de tráfego BitTorrent no Irã [Anderson 2013]. Foram analisados dados coletados por diversos clientes utilizando a ferramenta *Network Diagnostic Tool* (NDT), hospedada na plataforma de medição M-Lab. Os resultados da análise indicaram a presença de dois períodos longos em que houve degradação do tráfego BitTorrent. Entre 30 de novembro de 2011 e 15 de agosto de 2012 houve uma diminuição de 77% na taxa de transferência. Já entre 4 de outubro e 22 de novembro de 2012 a diminuição detectada foi de 69%.

Em 24 de junho de 2013, leitores do jornal *online Zambianwatchdog.com*, da Zâmbia, relataram terem recebido apenas mensagens de erro ao acessar a página Web [Mr T. 2013]. O jornal é considerado a maior página Web na Zâmbia após o Facebook, Google e YouTube. Foram executados testes com a ferramenta Ooni, os quais revelaram que a página era a única sendo bloqueada pelo governo do país.

A dissertação de mestrado de Shadi Esnaashari [Esnaashari 2014] apresenta a ferramenta *Web Censorship Monitoring Tool* (WCMT) utilizada entre julho e setembro de 2013 para identificar bloqueio de acesso a páginas Web e serviços da Internet na rede de diferentes organizações e ISPs em Wellington, na Nova Zelândia. Os resultados mostraram que todas as organizações e ISPs avaliados efetuaram bloqueio de algum conteúdo. Porém, houve uma variedade grande de conteúdos diferentes bloqueados em redes diferentes. O autor afirma que isto demonstra a falta de critérios das organizações ao definir o que deve ser bloqueado.

Shankesi propõe em 2013 na sua tese de doutorado, uma infraestrutura para detecção de manipulação de rede chamada Friendsourcing [Shankesi 2013]. O Friendsourcing baseia-se em colaboração coletiva (*crowdsourcing*), utilizando redes sociais para que um usuário receba auxílio de seus contatos para detectar se sua rede está sofrendo algum tipo de manipulação. O autor conduziu experimentos com 54 usuários reais na Índia. Os resultados mostraram que 64 endereços Web foram bloqueados por vários ISPs na Índia.

Em fevereiro de 2014, um usuário do fórum de discussões Reditt relatou um caso

de degradação de tráfego quando conectado em uma VPN utilizando a porta padrão do serviço OpenVPN [reddit 2014]. O usuário afirmou que, caso utilizasse outra porta, seu tráfego não era degradado. Diversos outros usuários confirmaram a denúncia. Também em 2014, Brodkin relata que a velocidade média de transferência no serviço Netflix teve uma queda nos últimos três a quarto meses na rede dos ISPs Verizon e Comcast [Brodkin 2014].

Em 1 de fevereiro de 2016, van Schewick envia ao Presidente da *Federal Communications Commission* (FCC) – o órgão regulador das telecomunicações nos E.U.A. um relatório no qual aponta que o serviço Binge On da operadora T-Mobile viola a NR, prejudicando a liberdade de escolha do usuário, a inovação, a concorrência e a liberdade de expressão na Internet [van Schewick 2016]. Segundo a autora, em novembro de 2015 a operadora T-Mobile, o terceiro maior provedor de acesso à Internet móvel nos EUA, lançou um novo serviço chamado Binge On no qual oferece transferência de vídeo ilimitada de provedores selecionados. Assim, os clientes podem acessar vídeos de 42 provedores, como Netflix, Amazon, Hulu, HBO, entre outros, sem o uso dos seus planos de dados, uma prática conhecida como "taxa zero". A autora afirma que esta prática configura um caso de DT, pois o ISP está favorecendo um conjunto de serviços em detrimento de outros. Já em 7 de fevereiro de 2016, a operadora Verizon é também acusada de violar a NR pela prática de "taxa zero" com seu serviço móvel de vídeo chamado Go90. Este serviço exclui o tráfego de vídeo da própria Verizon da franquia de dados de seus clientes [Lomas 2016].

Em 2 de março de 2016, a Public Knowledge, uma organização sem fins lucrativos que defende a NR e outros direitos do usuário na Internet, registrou uma queixa junto à FCC sobre o serviço Stream TV da operadora Comcast [Dreier 2016]. A denúncia diz que a Comcast não computa o tráfego do seu serviço Stream TV na franquia de dados de seus clientes, configurando assim a prática de "taxa zero" e, portanto, uma violação da NR. A Public Knowledge solicita então que a FCC interrompa o serviço discriminatório da operadora Comcast [Public Knowledge 2016].

Em 24 de março de 2016, o Netflix declarou que limita seu tráfego de vídeo em 600 Kbps para clientes acessando o serviço a partir de redes móveis [Knutson and Ramachandran 2016]. Segundo o Netflix, esta prática tem o objetivo de proteger seus clientes de cobranças adicionais por excederem suas franquias de dados. Em 25 de março de 2016, a *American Cable Association* (ACA) divulga uma declaração reprovando esta prática do Netflix [American Cable Association 2016]. Segundo a ACA, a FCC deve investigar os provedores de conteúdo e revisar sua regulamentação sobre a NR para incluir restrições também aos provedores de conteúdo e não somente aos ISPs. A FCC responde que, embora provedores de conteúdo não estejam incluídos na regulamentação da NR, a Netflix teve um comportamento que pode ser considerado incoerente. E conclui que esta revelação da Netflix põem em dúvida todo o fundamento e a razão de ser da decisão da NR [O'Rielly 2016].

Em 01 de abril de 2016, um grupo de mais de 50 organizações de interesse público e de defesa do consumidor pressionam a FCC para que tome medidas contra as práticas de "taxa zero" [Campbell 2016]. Segundo o grupo, estas práticas de ISPs como a Verizon, AT&T e T-Mobile, prejudicam a livre concorrência, a inovação, limitam a escolha do

usuário e elevam os preços.

A partir dos diversos casos apresentados, observa-se que houveram reclamações e estudos sobre violações da NR em diversos lugares do mundo e ao longo de todo o debate da NR. Assim, é possível afirmar que garantir o cumprimento da NR não é uma tarefa trivial, visto que diversos órgãos reguladores têm falhado em fiscalizar os ISPs.

4.3. Debate e Normatização Mundial da NR

Esta seção apresenta uma introdução ao debate da NR e descreve um panorama da normatização da NR em diversos países ao redor do mundo. Estas normatizações consistem em regras, princípios e/ou Leis instituídos com o objetivo de garantir um tráfego neutro na Internet. O processo de normatização da NR ocorreu de forma relativamente diferente ao redor do globo.

O debate mundial acerca da NR iniciou-se em 2002 quando a *Federal Communications Commission* (FCC), o órgão regulador das telecomunicações nos E.U.A, alterou a classificação do serviço de banda larga no país. O serviço anteriormente era equivalente a um serviço de telecomunicações comum, como a telefonia fixa, por exemplo. A nova classificação passou a ser de "serviço de informação" (*information service*) [Federal Communications Commission 2002], desvinculando a banda larga das leis que regulavam as telecomunicações. As leis que regem as telecomunicações garantiam uma Internet neutra no país. Assim, com a nova classificação, os ISPs ganharam o poder de priorizar ou bloquear um tipo de tráfego de dados em detrimento de outros. Neste contexto, inicia-se então o debate sobre a "Neutralidade da Rede", termo criado por Tim Wu [Wu 2002] ainda em 2002.

Em 2003, Tim Wu e Lawrence Lessig, um dos criadores da Creative Commons [Lessig 2001], enviaram uma carta à FCC apresentando uma proposta sobre a NR [Wu and Lessig 2003]. Esta proposta estabelecia um equilíbrio entre a proibição de ISPs em restringir o que os usuários fazem com suas conexões à Internet e a liberdade dos ISPs para gerenciar suas próprias redes.

A partir de então, cada vez mais indivíduos, empresas e instituições públicas e privadas passaram a fazer parte do debate da NR ao redor do planeta. Diversos provedores de conteúdo, como Google e Netflix, defendem a NR, enquanto a oposição é formada principalmente pelos ISPs. A comunidade científica mundial também ingressou no debate, trazendo conceitos, técnicas e outros aspectos relevantes que embasam as discussões e normatizações da NR ao redor do mundo.

Um conceito importante no debate da NR, e presente em diversas normatizações, é a gerência razoável do tráfego. O conceito da gerência razoável do tráfego determina quais práticas de gerência de tráfego os ISPs podem efetuar sem que a NR seja violada. A gerência de tráfego de um ISP pode ser considerada razoável se esta não for anticompetitiva, não causar danos indevidos aos consumidores e não prejudicar injustificadamente a liberdade de expressão [Jordan 2009b, Jordan 2009a].

Descrevemos abaixo, na subseção 4.3.1, em ordem cronológica, alguns dos pontos principais da normatização de diversos países.

4.3.1. Normatização Mundial da NR

Em 19 de setembro de 2006, o Japão, por meio do Ministério de Assuntos Internos e Comunicações (MIC), lança um programa (*New Competition Promotion Program 2010*) [Ministry of Internal Affairs and Communications 2006] que estabelece uma série de medidas a serem implementadas até 2010. O objetivo destas medidas é garantir a concorrência justa no mercado de telecomunicações e assegurar os direitos do consumidor. Este programa criou um grupo de trabalho para estudar o tema da NR e como ela deve ser implementada no país. Este grupo de trabalho apresentou o seu primeiro relatório em 20 de setembro de 2007 [Ministry of Internal Affairs and Communications 2007] e, em 07 de março de 2008, apresentou um segundo relatório [Ministry of Internal Affairs and Communications 2008] contendo recomendações para a manutenção da NR. Estas recomendações incluem: estudos sobre o mercado de telecomunicações e a utilização da infraestrutura de rede disponível; a criação de sistemas para fiscalizar a qualidade de serviço fornecida pelos ISPs; e o estabelecimento de regras que garantam um tráfego neutro, permitam novos modelos de negócio e protejam os usuários.

A Noruega lançou em 24 de fevereiro de 2009 suas diretrizes para a NR por meio de seu órgão regulador, o *Norwegian Communications Authority* (Nkom) [Norwegian Communications Authority a]. Estas diretrizes estabelecem que os usuários devem receber dos ISPs exatamente o serviço contratado, sem discriminação ou bloqueio de conteúdo. A Nkom também afirma que o modelo norueguês da NR [Norwegian Communications Authority b] busca mediar os interesses dos provedores de conteúdo, provedores de acesso e consumidores. Em 18 de novembro de 2014, Frode Sørensen, conselheiro sênior da Nkom, afirma que a prática de "taxa zero" viola as diretrizes norueguesas da NR [Sørensen 2014]. Nesta prática, o tráfego de um aplicativo específico não é considerado na franquia de dados do consumidor. Assim, segundo Sørensen, esta prática configura um caso de discriminação de tráfego, já que, caso o consumidor tenha utilizado toda a sua franquia, tal aplicativo não tem seu tráfego bloqueado ou estrangulado como acontece com os demais.

Em 21 de outubro de 2009, o Canadá, por meio da *Canadian Radio-television and Telecommunications Commission* (CRTC), publicou sua regulamentação em relação às práticas de gerência do tráfego da Internet empregadas pelos ISPs [Canadian Radio-television and Telecommunications Commission 2009]. A regulamentação estabelece que toda gerência de tráfego deve ser feita de forma transparente e sem discriminação, e que os ISPs devem manter investimentos na rede como a principal solução para evitar congestionamentos.

Em 18 de agosto de 2010, o governo do Chile decretou a Lei n. 20.453 [Subsecretaría de Telecomunicaciones 2010] que instituiu a NR no país. A Lei estabelece regras para órgãos públicos e empresas privadas que fornecem serviços de telecomunicações. Estas regras proíbem o bloqueio, interferência ou discriminação no acesso dos usuários a qualquer conteúdo legal, quaisquer que sejam os equipamentos utilizados, desde que não prejudiquem a rede. A Lei também exige que os provedores de acesso à Internet publiquem todas as características dos serviços prestados (largura de banda, disponibilidade, garantias, entre outros) e forneçam aos consumidores serviços de controle parental para filtrar conteúdos ilegais ou que os consumidores julguem impróprios.

Em 16 de novembro de 2011, a Secretaria Nacional de Telecomunicações do Chile (SUBTEL) afirma que os ISPs não estão fornecendo as informações exigidas por lei de forma suficientemente transparente e padronizada [Subsecretaría de Telecomunicaciones 2011]. Assim, a SUBTEL padronizou as informações que devem ser publicadas pelos ISPs, a fim de que os consumidores possam facilmente comparar os diferentes provedores de acesso. Em 27 de maio de 2014, a SUBTEL, com base na Lei de NR, proíbe a pratica de "taxa zero", sob pena de multa [Subsecretaría de Telecomunicaciones 2014]. Assim, os ISPs do Chile não podem mais comercializar serviços que incluem as chamadas "redes sociais gratuitas".

Em 16 de junho de 2011, o governo da Colômbia aprovou a Lei 1.450 referente ao Plano Nacional de Desenvolvimento para os anos de 2010 a 2014 [El Congreso de Colombia 2011]. O Art. 56 desta Lei trata da "Neutralidade na Internet", estabelecendo regras para os prestadores de serviço de Internet. Estas regras proíbem o bloqueio, modificação e discriminação de qualquer tráfego na rede, além de exigir que os ISPs publiquem todas as características dos serviços prestados. A Lei também exige que os ISPs forneçam serviços de controle parental aos consumidores e que implementem mecanismos para preservar a privacidade dos usuários. Em 16 de dezembro de 2011, o governo publica a Resolução 3502, que define regras para o cumprimento da NR estabelecida no Art. 56 da Lei 1.450, aprovada anteriormente [Comisión de Regulación de Comunicaciones 2011]. Estas regras tratam dos princípios e aspectos técnicos referentes à NR que devem ser seguidos. Estes princípios incluem: a livre escolha do usuário ao utilizar a rede, tráfego sem discriminação, transparência na gerência de tráfego e informação quanto aos serviços prestados pelos ISPs. Também são definidas quais práticas de gerência de tráfego são permitidas.

Em 3 de julho de 2012, a *Korea Communications Commission* (KCC), agência reguladora das telecomunicações da Coréia do Sul, publicou seu relatório anual referente ao ano de 2011 [Korea Communications Commission 2012]. Neste relatório, a KCC estabelece as diretrizes a serem seguidas para implementar a NR no país. Estas diretrizes proíbem o bloqueio e discriminação de tráfego e garantem direitos do consumidor como a transparência sobre as práticas de gerência adotadas pelos ISPs e as características dos serviços fornecidos, entre outros. O documento também define quais práticas de gerência de tráfego são aceitáveis, sem que a NR seja violada.

O debate sobre a NR no Brasil iniciou em 2009, quando o Comitê Gestor da Internet no Brasil (CGI.br), o órgão responsável pela governança da Internet no país, lança uma Resolução com os 10 Princípios para a Governança e Uso da Internet no Brasil [Comitê Gestor da Internet no Brasil 2009], o qual inclui a NR. A partir destes princípios iniciou-se um processo colaborativo que resultou no Projeto de Lei 2126/2011 [Poder Executivo 2011], apresentado em 24 de agosto de 2011, o qual torna-se Lei Ordinária 12965/2014 em 23 de abril de 2014, o chamado "Marco Civil da Internet" [Presidência da República 2014]. A Lei estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil. A NR é tratada no Artigo 9º desta Lei, estabelecendo que ISPs tem o dever de tratar todo pacote de dados da mesma forma, sem nenhum tipo de DT. A Lei ainda estabelece que a DT poderá ocorrer em casos especiais a serem regulamentados posteriormente pelos órgãos reguladores competentes, desde que seja transparente e não cause danos ao usuário. A regulamentação da Lei é decretada em 11 de maio de 2016 [Presidência da República 2016], com base em consultas públicas [Ministério da Justiça]. Quanto

à NR, esta regulamentação trata dos casos especiais em que a discriminação de tráfego é permitida e estabelece parâmetros para a fiscalização de violações. Em particular, a regulamentação proíbe as práticas de "taxa zero".

Em 14 de julho de 2014, o governo mexicano alterou a Lei Federal de Telecomunicações e Radiodifusão [Secretaria de Comunicaciones y Transportes 2014], incluindo itens que instituem a NR no México. Estes itens estabelecem que ISPs não podem bloquear, discriminar nem modificar qualquer conteúdo legal acessado pelos usuários e devem preservar a privacidade dos mesmos. A Lei também passa a exigir que ISPs mantenham uma qualidade mínima de serviço, sempre fornecendo aos usuários exatamente o que consta em contrato e mantendo públicas todas as características dos serviços oferecidos. Além disso, permite aos ISPs práticas de gerência de tráfego que tenham como objetivo garantir a qualidade ou a velocidade do serviço contratado pelo usuário, desde que isto não configure uma prática contrária à livre concorrência.

A NR foi instituída nos E.U.A. em 26 de fevereiro de 2015, pela FCC [Federal Communications Commission 2015]. Entretanto, o debate sobre a NR nos E.U.A iniciou em 2002, quando o orgão alterou a classificação do serviço de banda larga no país, como descrito acima. Em 5 de agosto de 2005, a Suprema Corte concordou com a posição adotada em 2002 pela FCC, de que o serviço de banda larga é um serviço de informação [Federal Communications Commission 2005]. Em 21 de dezembro de 2010, a FCC adota três regras básicas para preservar a Internet como uma plataforma aberta para a inovação, o investimento, a criação de emprego, o crescimento econômico, a concorrência e a livre expressão [Federal Communications Commission 2010]: a transparência (*Transparency*), nenhum bloqueio (*No blocking*) e nenhuma discriminação não razoável (*No unreasonable discrimination*). Em 14 de janeiro de 2014, estas três regras adotadas pela FCC em 2010 são julgadas pela Corte de Apelação do Distrito de Columbia. Como resultado do julgamento, esta Corte de Apelação sanciona a regra da transparência, mas anula as regras de não bloqueio e de discriminação não razoável, por considerar que não são práticas ilícitas quando classificadas em "serviços de informação" [Federal Communications Commission 2014]. Assim, somente na regulamentação de 26 de fevereiro de 2015 [Federal Communications Commission 2015] a FCC volta a classificar o serviço de acesso à Internet como um serviço de telecomunicações, o que lhe garante o fundamento jurídico necessário para preservar e proteger a Internet aberta. Dentre as regras estabelecidas destaca-se que os ISPs não podem bloquear nem degradar pacotes de dados, independente de sua origem, destino e conteúdo. Também fica proibida a chamada "priorização paga", que ocorre quando um ISP aceita pagamento (monetário ou não) para gerenciar sua rede de forma a beneficiar um determinado conteúdo, aplicação, serviço ou dispositivo. Desde a adoção destas regras, a FCC vem sofrendo pressão dos opositores, como por exemplo: em 25 de fevereiro de 2016, um projeto de Lei é proposto para proibir a FCC de reclassificar o serviço de banda larga e de impor regras sobre os prestadores de tal serviço [Lee 2016]; em 11 de abril de 2016, o presidente da FCC, Tom Wheeler, afirma que a política de "taxa zero" está sendo revista e que não há uma data final definida [Federal Communications Commission 2016]; e em 15 de abril de 2016, a Câmara dos Deputados dos Estados Unidos aprovou, com apoio bipartidário, uma Lei que proíbe a FCC de regular as taxas cobradas pelo acesso à Internet de banda larga [Kinzinger 2016].

Em 30 de junho de 2015, a Comissão Europeia publicou uma nova regulamen-

tação que institui a NR em todos os países da União Europeia [European Commission 2015]. Esta regulamentação foi resultado de anos de negociação entre a Comissão Europeia, o Parlamento Europeu e o Conselho Europeu, além de consultas públicas [European Commission 2009, European Commission 2010, European Commission 2014]. A regulamentação estabelece que não pode haver bloqueio, degradação e discriminação de nenhum conteúdo, aplicação ou serviço na Internet. O estabelecimento de diretrizes para a implementação e fiscalização desta regulamentação ficou a cargo do *Body of European Regulators of Electronic Communications* (BEREC), agência reguladora das telecomunicações da União Europeia criada em 25 de novembro de 2009 [European Parliament and Council of the European Union 2009, Body of European Regulators for Electronic Communications].

Em 8 de fevereiro de 2016, o órgão regulador das telecomunicações da índia, a *Telecom Regulatory Authority of India* (TRAI), publicou um regulamento [Telecom Regulatory Authority of India 2016] proibindo a "taxa zero" e a cobrança de taxas extras dos provedores de conteúdo e dos usuários que acessem conteúdos específicos. Estas práticas estavam sendo implantadas por um ISP indiano, o que fomentou o debate sobre a NR no país [Press Trust of India 2015].

Este panorama global da normatização da NR mostra que existe uma preocupação dos governos ao redor do mundo com a manutenção de uma Internet neutra. Outros países tem discutido a NR, como a Nova Zelândia [InternetNZ 2015] e a Rússia [Federal Antimonopoly Service 2016], por exemplo, mas não foram encontrados documentos indicando que a NR já tenha sido instituída nestes países. Com base neste panorama, é possível extrair aspectos comuns às diversas normatizações. A DT está presente em todas as normatizações, tornando-a assim um elemento chave no contexto da NR.

4.4. Fundamentação

Esta seção apresenta uma fundamentação sobre a DT, a NR e o problema de detecção de DT. A subseção 4.4.1 trata da DT, apresentando conceitos relacionados à organização e funcionamento da Internet e de que forma a DT pode ser implementada. A subseção 4.4.2 apresenta uma definição da NR e as propriedades que uma rede neutra deve ter. Finalmente, na subseção 4.4.3, o problema de detecção de DT é definido.

4.4.1. Diferenciação de Tráfego

A Internet é uma rede global formada pela interconexão de diversas redes independentes, chamadas de Sistemas Autônomos (*Autonomous Systems*, ASes). Cada AS representa um conjunto diferente de prefixos de roteamento na Internet e é controlado por um ou mais ISPs. Os ISPs são hierarquicamente divididos em três camadas, chamadas *Tiers*. Os ISPs do *Tier* 1 são redes de alta capacidade que interconectam globalmente as redes dos ISPs *Tier* 2, constituindo o "núcleo" da Internet. ISPs *Tier* 2 fornecem conectividade aos ISPs *Tier* 3, que são, por exemplo, os ISPs residenciais, os quais fornecem acesso à Internet para os consumidores finais.

Um pacote de dados enviado de um *host* final para outro, potencialmente pode atravessar ASes de ISPs de todas os *Tiers*, especialmente se os *hosts* estiverem geograficamente distantes. Assim, a DT pode ocorrer em qualquer uma dos *Tiers* e de inúmeras

formas, já que cada AS pode empregar tecnologias diferentes, assim como políticas internas de roteamento e gerência de tráfego distintas.

Pacotes de dados em um AS qualquer atravessam diversos roteadores deste AS, desde o ponto de entrada (o primeiro *hop*) até o ponto de saída (o último *hop*). Ao sair de um AS, um pacote de dados entrará na rede de outro ou terá chegado em seu destino. Assim, uma possível DT praticada por um ISP acontecerá em um ou diversos *hops* entre o ponto de entrada e o ponto de saída do AS.

Em geral, o tráfego é segmentado em classes e tratado de forma diferente conforme a classe atribuída. Existem inúmeros mecanismos que podem ser utilizados para classificar e posteriormente discriminar um tráfego de dados. A classificação de um tráfego de dados pode ocorrer, por exemplo, apenas no ponto de entrada da rede e ser inserida no cabeçalho dos pacotes (cabeçalho do protocolo interno do AS), informando os roteadores seguintes como estes pacotes devem ser tratados. É possível também que todos os roteadores, por onde os pacotes de um tráfego de dados passam, efetuem tanto a classificação quanto a discriminação em si. Técnicas de Redes Definidas por Software (*Software-Defined Networking, SDN*) também podem ser utilizadas [Qazi et al. 2013]. As possibilidades de implementação são diversas.

A classificação de um tráfego de dados pode basear-se em diversos critérios, como origem, destino, porta de origem, porta de destino, protocolo de aplicação, AS anterior (de onde o pacote veio) ou próximo AS (para o qual o pacote será roteado), entre outros. Há ainda a técnica *Deep Packet Inspection* (DPI), que consiste em analisar não apenas o cabeçalho dos pacotes, mas também os dados (*payload*). O objetivo do DPI é identificar com maior acurácia a qual aplicação correspondem os pacotes.

Os mecanismos mais comuns de DT são engenharia de tráfego (*traffic shaping*) [Kanuparthy and Dovrolis 2011] e vigilância de tráfego (*traffic policing*) [Flach et al. 2016]. Outros exemplos incluem: a injeção de pacotes TCP do tipo *reset* (RST) forjados, a fim de forçar o encerramento de conexões TCP, interrompendo a comunicação entre dois *hosts* finais; o encaminhamento de pacotes para rotas diferentes conforme a sua classificação, sendo que uma das rotas é propositalmente menos congestionada, configurando assim uma *fast-lane*; *middleboxes* [Detal et al. 2013], os quais podem interferir no tráfego entre dois *hosts* finais; Redes de Distribuição de Conteúdo (*Content Delivery Networks*, CDN) [Maille et al. 2016], as quais cobram para entregar conteúdo de terceiros, podendo assim caracterizar uma priorização.

A engenharia e a vigilância de tráfego são efetuadas, em geral, por equipamentos dedicados ou pelos próprios roteadores da rede de um AS. Estes mecanismos diferem na forma com que os pacotes são processados conforme chegam nos roteadores ou outros equipamentos.

A engenharia de tráfego baseia-se no enfileiramento de pacotes em *buffers*. Idealmente, um pacote de dados, ao chegar em um roteador, é imediatamente encaminhado para o próximo *hop* de sua rota – a qual, em geral, é decidida por meio de uma tabela de roteamento. Porém, caso o roteador esteja sobrecarregado e não consiga encaminhar o pacote imediatamente, o pacote é colocado em um *buffer*. Uma política de escalonamento é então empregada para decidir a ordem em que os pacotes pendentes serão retirados do

buffer e encaminhados. Caso o *buffer* fique cheio, uma política de descarte é empregada, descartando os novos pacotes que chegarem ou até mesmo pacotes já presentes no *buffer*.

Já a vigilância de tráfego emprega políticas de descarte assim que pacotes em excesso começam a chegar, diferentemente da engenharia de tráfego que as emprega apenas quando o *buffer* está cheio. Assim, a vigilância de tráfego não baseia-se no enfileiramento de pacotes, já que estes são descartados antes que acumulem.

As políticas de escalonamento de pacotes mais comuns são [Kanuparthy and Dovrolis 2010, Weinsberg et al. 2011]: (i) *First Come First Served* (FCFS), em que os pacotes que chegaram primeiro são escalonados primeiro; (ii) *Strict Priority* (SP), em que o escalonador sempre dá prioridade para uma classe específica; (iii) *Leaky Bucket*, em que cada classe tem um limite máximo de largura de banda; (iv) *Token Bucket*, em que cada classe tem um limite para a largura de banda média consumida pelos fluxos; e (v) *Weighted Fair Queuing* (WFQ), em que a largura de banda permitida para cada classe é dividida com base em pesos. Já as políticas de descarte mais comuns são: (i) *Drop-Tail* (DT), no qual em caso de *buffer* cheio os próximos pacotes a chegarem são descartados e *Weighted Random Early Detection* (WRED), em que pacotes de menor prioridade tem maior probabilidade de serem descartados.

4.4.2. Rede Neutra

O projeto original da Internet foi guiado por dois princípios fundamentais que são elementos-chave no contexto da NR [Krämer et al. 2013]: fim-a-fim (*end-to-end*) e melhor esforço (*best-effort*). O princípio fim-a-fim diz que as mensagens são fragmentadas em pacotes de dados que devem ser roteados através da rede de forma autônoma. Um *hop* intermediário (roteador) deve decidir apenas qual será o próximo *hop* para um pacote qualquer, enviando-o pelo menor caminho segundo sua tabela de roteamento. Assim, um roteador não tem controle do caminho completo que o pacote percorre da origem até o destino final. Já o princípio do melhor esforço garante que todos os pacotes de dados serão enviados pela rede tão rápido quanto possível. Se a taxa de chegada de pacotes em um roteador é maior que sua capacidade de envio, os pacotes serão enfileirados. Se a fila de pacotes encher, os próximos pacotes a chegar serão descartados, independentemente dos seus conteúdos, origens ou destinos.

No contexto da NR, estes princípios estabelecem que todos os pacotes de dados enviados pela rede devem ser tratados com igualdade e que nenhum *hop* intermediário pode exercer controle sobre a rede como um todo. Entretanto, não há uma definição amplamente aceita da NR, havendo algumas diferenças [Hahn and Wallsten 2006, Internet Society , Scott 2014, Ganley and Allgrove 2006, Crowcroft 2007]. Um conceito comum aos diversos trabalhos, normatizações e princípios fundamentais da Internet trata da DT. Assim, neste trabalho consideramos que para uma rede ser considerada neutra, todo pacote de dados deve ser tratado da mesma forma, ou seja, a DT não é permitida.

Portanto, em uma rede neutra, os roteadores de um ISP devem escalonar os pacotes a serem encaminhados seguindo a política FCFS (*First Come First Served*) e a política de descarte de pacotes deve ser DT (*Drop-Tail*) [Kanuparthy and Dovrolis 2010]. O próximo pacote a ser encaminhado é sempre o que chegou antes e em caso de *buffer* cheio os próximos pacotes a chegarem são descartados, independentemente da classifi-

cação. Assim, todo tipo de tráfego está sujeito às mesmas condições de atraso e perdas.

4.4.3. O Problema de Detecção de DT

O problema de detecção de DT tratado neste trabalho é defino como: inferir se um tráfego de dados está sendo tratado de forma diferente de outro(s) tráfego(s). Em outras palavras, o problema consiste em detectar se pacotes de diferentes tráfegos estão sujeitos a diferentes tratamentos de rede, apenas por terem propriedades distintas (gerados por aplicações diferentes, por exemplo).

4.5. Soluções para Detecção de Diferenciação de Tráfego

Segundo o *Body of European Regulators of Electronic Communications* (BEREC), a associação das agências reguladoras das telecomunicações da União Europeia, apenas a normatização da NR não garante seu cumprimento por parte dos ISPs. É importante que os usuários finais tenham conhecimento dos serviços efetivamente oferecidos a eles pelos ISPs contratados [Body of European Regulators for Electronic Communications 2012a, Body of European Regulators for Electronic Communications 2012b]. Assim, soluções para detectar possíveis práticas de DT são necessárias. Esta seção apresenta diversas soluções já existentes para o problema da detecção de DT, assim como um apanhado das técnicas utilizadas por estas soluções.

A DT pode afetar um tráfego de dados de diversas formas. Mecanismos de engenharia de tráfego, por exemplo, podem resultar em atrasos maiores para o tráfego discriminado. Já a vigilância de tráfego pode resultar em maiores taxas de perda de pacotes. Caso pacotes de diferentes tipos sejam encaminhados por rotas diferentes, estes podem apresentar um desempenho de rede diferente se uma das rotas está congestionada e a outra não. Assim, as soluções existentes para detecção de DT utilizam medições de rede para detectar estas diferenças de desempenho e inferir se um determinado tipo de tráfego está sendo discriminado em relação a outros.

Porém, diversos outros fatores além da DT podem resultar em uma diferença no desempenho medido para tipos de tráfegos diferentes – as chamadas variáveis de confusão. Exemplos incluem diferença de rotas, tráfego de fundo, mudanças constantes nas condições da rede, congestionamento, configuração dos *hosts* finais, além das próprias limitações das técnicas de medição utilizadas. Assim, modelos estatísticos robustos são necessários para se obter resultados confiáveis [Tariq et al. 2009].

As medições de rede efetuadas pelas soluções podem ser ativas ou passivas. Em uma medição ativa, as medições são obtidas a partir de tráfegos de dados artificiais entre um ou mais pares de *hosts*. Já na medição passiva, as medições são obtidas apenas observando-se tráfegos de dados reais, sem introduzir novos pacotes na rede. No caso da DT, dependendo de qual métrica esteja sendo utilizada, é possível que diferenças significativas nas medições para tráfegos diferentes sejam observáveis apenas quando a rota entre os *hosts* estiver congestionada. Assim, medições ativas, em geral, criam uma grande quantidade de tráfego para saturar a banda disponível no caminho entre dois *hosts* finais, forçando que atrasos e/ou perdas de pacotes aconteçam. As principais métricas utilizadas na soluções descritas nesta seção são: taxa de perda de pacotes, taxa de transferência e atraso.

As soluções apresentadas nesta seção diferem, em geral, na topologia de medição, nas métricas, nos modelos estatísticos e nos tipos de tráfegos de dados empregados nas medições. Diversas soluções efetuam medições ativas entre um ou mais pares de *hosts* – utilizando tráfegos correspondentes a aplicações diferentes – e comparam as medições obtidas a fim de detectar variações significativas. Outras soluções efetuam medições a cada *hop* do caminho entre um ou mais pares de *hosts*, com o objetivo de identificar exatamente onde a DT ocorre. Há ainda soluções que utilizam medições passivas de tráfegos de diferentes aplicações.

O restante desta seção está organizado da seguinte forma. As subseções 4.5.1 até 4.5.9 descrevem soluções que detectam DT. A subseção 4.5.10 apresenta um resumo comparativo das soluções. Na subseção 4.5.11, outros trabalhos relacionados à NR, mas que não se referem diretamente à detecção de DT, são apresentados.

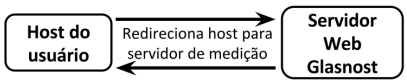
4.5.1. Glasnost, BTTest e BonaFide

Glasnost [Dischinger et al. 2010] é uma ferramenta que permite a usuários finais da Internet detectarem se seus ISPs estão praticando DT baseado nas aplicações em uso. O sistema já foi utilizado por milhares de usuários ao redor do mundo, incluindo usuários residenciais sem conhecimento técnico. A ferramenta foi inicialmente aplicada para detecção de DT de BitTorrent, mas pode também ser utilizada para qualquer outro protocolo de aplicação.

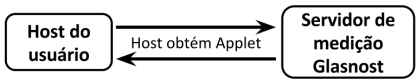
A ferramenta Glasnost foi projetada para ser de fácil utilização por qualquer usuário, independentemente de seu conhecimento técnico. O funcionamento da Glasnost é ilustrado na Figura 4.1. Primeiramente o usuário acessa a página Web da ferramenta[1] e é redirecionado para um servidor de medição, como mostra a Figura 4.1a. Existem vários servidores de medição e os usuários são redirecionados dinamicamente para um destes servidores, tornando difícil para os ISPs empregarem medidas contra servidores específicos. O navegador do usuário obtém então a aplicação cliente da Glasnost, como ilustrado na Figura 4.1b. A aplicação cliente é um *applet* Java executado pelo navegador do usuário que conecta-se ao servidor de medição e emula uma sequência de fluxos de dados, efetuando os testes de taxa de transferência para diferentes aplicações, como mostra a Figura 4.1c. Cada teste é composto por dois fluxos de dados em sequência. Um destes fluxos corresponde à aplicação sendo testada, sendo constituído pelo protocolo e dados específicos da aplicação. O outro fluxo é idêntico ao primeiro em quantidade de mensagens, ordem e tamanho dos pacotes, porém com conteúdo definido de forma aleatória, servindo como um *baseline* para comparação com o fluxo da aplicação. A partir da medição da taxa de transferência dos diferentes fluxos, é possível detectar se um ISP está praticando DT baseada no conteúdo das mensagens, como descrito abaixo.

Cada fluxo de dados entre a aplicação cliente e o servidor de medição dura diversos segundos, tempo suficiente para que o TCP chegue a uma taxa de transferência estável. Os testes são repetidos múltiplas vezes, a fim de diminuir o ruídos nas medições obtidas. Ao término da série de testes, o servidor de medição processa os dados obtidos e mostra uma página de resultados ao usuário. As métricas computadas são o valor mínimo, máximo e a mediana das taxas de transferência medidas.

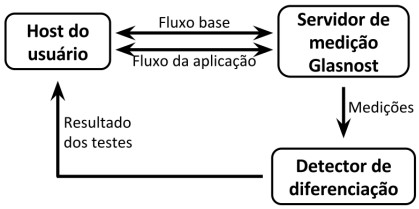
[1] `http://broadband.mpi-sws.org/transparency/glasnost.php`



(a) Usuário redirecionado para um servidor de medição.



(b) Obtenção da aplicação cliente.



(c) Execução de testes de taxa de transferência.

Figura 4.1: Funcionamento da ferramenta Glasnost.

Para detectar a DT, a Glasnost verifica se a diferença entre a taxa máxima de transferência dos dois fluxos de dados é maior que um limiar σ. Este limiar é um compromisso entre a capacidade de detectar DT e a produção de falso-positivos (falsas acusações de DT). Se o valor de σ for grande, como 50%, por exemplo, a ferramenta detecta DT apenas se a taxa máxima de transferência de um fluxo for metade da taxa máxima do outro fluxo. Por outro lado, se σ tiver um valor pequeno, 5%, por exemplo, a ferramenta pode erroneamente detectar DT quando houve apenas influência de algum tráfego secundário. Os autores afirmam que 20% é um bom valor para o limiar σ.

Os autores relatam que, em 2010, a Glasnost detectou que 10% dos seus usuários sofreram DT de BitTorrent. Dentre os casos detectados, a grande maioria ocorreu apenas no envio de dados (*upstream*), com poucos casos de DT detectados no recebimento (*downstream*) e 20% em ambos. Um resultado surpreendente é que, depois de concluir-se que um ISP estava praticando DT, apenas 21% dos usuários do ISP foram efetivamente afetados (mediana). Os autores listam 3 possíveis explicações para este cenário: (i) apenas usuários geradores de uma quantidade grande de tráfego foram afetados; (ii) apenas algumas partes do ISP foram afetadas; e (iii) a DT foi aplicada apenas durante períodos específicos, como horários de pico, por exemplo. Os autores também relatam que cerca de 6% dos usuários alegaram que a ferramenta não detectou DT que eles acreditavam estarem sofrendo. Uma possível explicação para isto é que a decisão de minimizar os falso-positivos pode aumentar os falso-negativos.

Uma ferramenta anterior à Glasnost, BTTest [Dischinger et al. 2008], foi criada por alguns dos autores da Glasnost e claramente serviu de base para a mesma. A BTTest detecta se um ISP está bloqueando tráfego BitTorrent. O funcionamento da BTTest é muito similar ao da Glasnost, exceto que a BTTest detecta apenas bloqueio de tráfego e apenas para BitTorrent. A BTTest foi disponibilizada por um período de 17 semanas, no qual mais de 47300 usuários finais utilizaram a ferramenta ao redor do mundo. Os dados obtidos neste período foram analisados e concluiu-se que em cerca de 8% dos testes foi detectado o bloqueio de tráfego BitTorrent, principalmente nos EUA. Além disso, a grande maioria dos bloqueios, cerca de 99%, ocorreu no envio de dados (*upstream*) e não

no recebimento (*downstream*).

Foi também desenvolvida posteriormente por outros autores outra ferramenta similar, BonaFide [Bashko et al. 2013]. BonaFide é uma adaptação da Glasnost focada em detectar DT em redes móveis. A ferramenta foi desenvolvida para o sistema Android e funciona de forma muito similar à Glasnost, mas com algumas modificações relacionadas às restrições presentes em dispositivos móveis. Na BonaFide, uma aplicação cliente executada no dispositivo móvel comunica-se com um servidor de medição executando assim os testes. Cada teste é constituído de 2 fluxos de dados, como na Glasnost. O BonaFide suporta diversos protocolos de aplicação, como VoIP e BitTorrent, por exemplo.

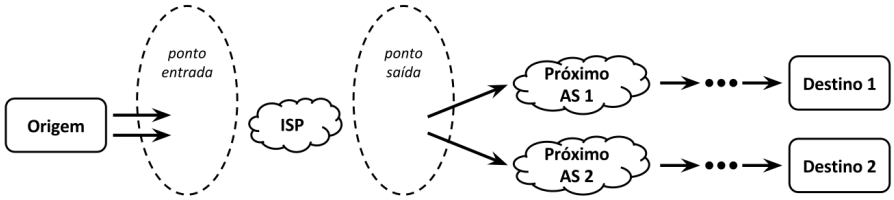
4.5.2. NetPolice e NVLens

NetPolice [Zhang et al. 2009] é uma ferramenta para detecção de DT em ISPs do "núcleo" da Internet (*Tier* 1). Os autores afirmam que detectar DT no núcleo tem impacto maior do que a detecção nos ISPs que atendem diretamente os usuários finais, já que a DT no núcleo potencialmente afeta uma quantidade maior de tráfego. A detecção de DT da NetPolice utiliza a taxa de perda de pacotes como métrica, que é medida a partir de diversos pontos de vista – *hosts* finais – em relação a um mesmo núcleo.

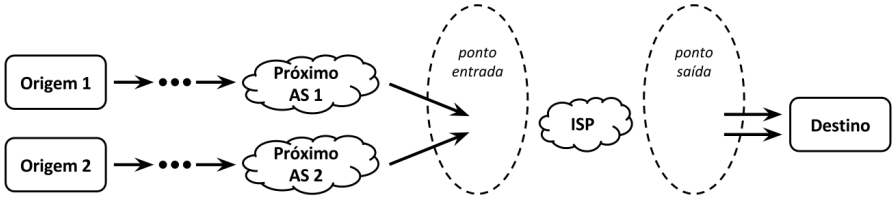
A NetPolice detecta DT baseada em conteúdo e em roteamento. A DT baseada em conteúdo ocorre quando os tráfegos gerados por diferentes aplicações são tratados de forma diferente, isto é, de acordo com a porta destino ou conteúdo dos pacotes, um ISP pode dar prioridade maior/menor aos mesmos ou até bloqueá-los. A DT baseada em roteamento ocorre quando pacotes são tratados de forma diferente dependendo dos seus dados de roteamento como, por exemplo, de qual AS veio o pacote ou para qual AS o pacote será encaminhado.

A Figura 4.2 mostra como a NetPolice detecta cada tipo de DT – baseada em conteúdo e roteamento. Na Figura 4.2a, as medições são feitas usando uma mesma origem e destinos diferentes, selecionados de forma que os *hops* imediatamente posteriores ao ponto de saída do ISP correspondam a ASes diferentes. Assim, é possível detectar se o ISP faz DT baseada no próximo AS para o qual o pacote será encaminhado. Na Figura 4.2b, as medições são feitas usando um mesmo destino e origens diferentes, selecionadas de forma que os *hops* imediatamente anteriores ao ponto de entrada do ISP sejam de ASes diferentes. Desta forma, é possível detectar se o ISP faz DT dependendo do AS anterior à sua rede. Na Figura 4.2c, as medições são feitas usando a mesma origem e destino, mas com pacotes de aplicações diferentes (porta destino e conteúdo). Assim, é possível detectar quando o ISP pratica DT baseada no conteúdo dos pacotes.

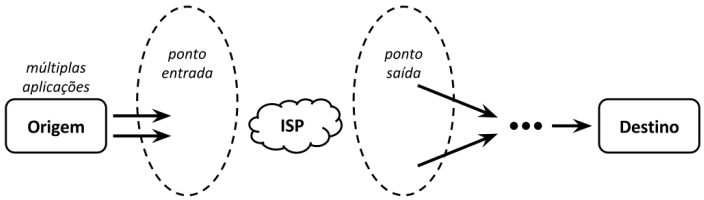
A detecção de DT da NetPolice baseada nas medições de perda de pacotes segue 4 etapas, ilustradas na Figura 4.3. A primeira etapa consiste em descobrir todos os caminhos que atravessam o ISP a ser avaliado, a partir de diversas origens (*probers*). Neste processo um grande número de rastreamentos de rota (usando o comando *traceroute*, por exemplo) é executado a partir de cada origem e para a maior quantidade possível de destinos na Internet (prefixos). Assim, além dos caminhos, são também obtidas as distâncias entre os pontos de entrada e saída do ISP, bem como os ASes anteriores e posteriores a estes pontos. Com esta informação, o NetPolice pré-calcula os valores de TTL (*Time to Live*) para alcançar cada par de pontos de entrada e saída do ISP alvo, a partir de todas as



(a) Detecção de DT baseada no próximo AS.



(b) Detecção de DT baseada no AS anterior.



(c) Detecção de DT baseada em conteúdo.

Figura 4.2: Detecção de diferentes tipos de DT na ferramenta NetPolice.

origens. O conjunto de caminhos e demais informações obtidas nesta etapa são chamados de *path view*.

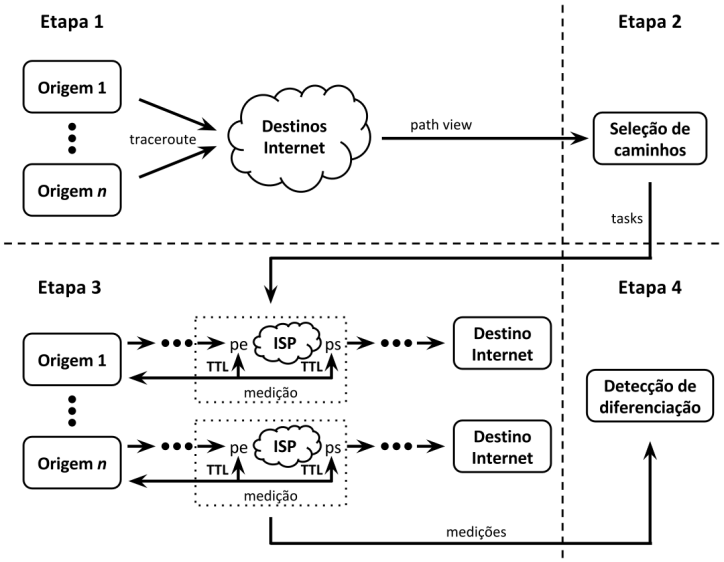


Figura 4.3: Funcionamento da ferramenta NetPolice.

Na segunda etapa são selecionados, a partir do *path view* criado na etapa anterior, quais caminhos serão efetivamente medidos, já que não é factível medir todos. Esta se-

leção de caminhos a serem medidos deve resultar em uma boa cobertura da rede interna do ISP alvo. A escolha deve ser inteligente, para que não sejam escolhidos origens e destinos que passem pelos mesmos caminhos internos do ISP ou caminhos que não atravessem o ISP. A seleção é modelada como um problema de otimização, com as seguintes restrições: cada tupla *(origem,entrada,saída)* deve ser percorrida pelo menos *R* vezes por caminhos para diferentes destinos; cada tupla *(entrada,saída,destino)* deve ser percorrida pelo menos *R* vezes por caminhos a partir de origens diferentes; e não podem haver mais que *m* caminhos a partir da mesma origem. O conjunto de caminhos a serem medidos é chamado de *tasks*.

Na terceira etapa são feitas as medições dos caminhos selecionados na etapa anterior para diferentes aplicações: HTTP, BitTorrent, SMTP, PPLive e VoIP. A medição de um caminho consiste em, uma vez a cada 200 segundos e para cada aplicação, enviar 2 pacotes: um pacote com o valor de TTL pré-calculado para alcançar apenas o ponto de entrada (*pe*) e gerar uma resposta ICMP de tempo excedido; e outro pacote com TTL pré-calculado para alcançar o ponto de saída (*ps*). Assim, subtrai-se a taxa de perda de pacotes do ponto de entrada do ISP da taxa do ponto de saída, obtendo-se a medição apenas para o caminho interno do ISP.

A quarta e última etapa consiste em inferir se o ISP está praticando DT baseada em roteamento ou conteúdo. Esta inferência utiliza o teste Kolmogorov-Smirnov (KS) [Noether 2012] para comparar as distribuições dos dados de medição obtidos. A detecção de DT por conteúdo é feita então comparando-se as distribuições de dados de cada aplicação com a distribuição de dados da aplicação HTTP, isto é, testes KS são aplicados para determinar se um conjunto de dados medidos para uma aplicação é significativamente diferente do conjunto de dados para a aplicação HTTP, caracterizando assim uma DT. A detecção de DT baseada em roteamento é feita de forma similar, mas comparando-se as distribuições de dados de caminhos diferentes para uma mesma aplicação.

Resultados experimentais com a NetPolice foram obtidos no PlanetLab. Nestes experimentos, 18 ISPs distribuídos em 3 continentes foram estudados em um período de 10 semanas. Os resultados mostraram que 4 ISPs realizaram DT em 4 aplicações e 10 ISPs realizaram DT baseada no AS anterior dos pacotes. As taxas de perda de pacotes medidas nestes casos chegaram a ser até 5% diferentes. Os autores também observaram, a partir dos resultados obtidos, que a DT pode depender da carga da rede. Já para alguns ISPs, os valores atribuídos ao campo TOS do cabeçalho dos pacotes tem forte relação com a DT (diferente priorização) e esta atribuição de valores é baseada apenas na porta de destino dos pacotes, não no conteúdo (não é feito DPI). Outra observação foi que a DT não é feita de forma homogênea em todos os roteadores dos ISPs.

Um trabalho anterior à NetPolice foi publicado pelos mesmos autores e apresenta uma versão anterior da ferramenta, com o nome de NVLens [Zhang et al. 2008]. No trabalho mais recente [Zhang et al. 2009], os autores detalharam diversos experimentos no PlanetLab, expandiram a análise dos dados obtidos e reformularam a última etapa do processo de detecção (teste para comparação das distribuições de dados).

4.5.3. DiffProbe

A DiffProbe [Kanuparthy and Dovrolis 2010] é uma ferramenta de detecção de DT que utiliza atraso e/ou descarte de pacotes como métrica. Esta detecção é feita por meio de medições feitas em fluxos de dados simultâneos entre um *host* cliente e um servidor. A ferramenta assume que um ISP classifica cada pacote como sendo de alta prioridade (classe H) ou baixa prioridade (classe L). Os autores afirmam que esta estratégia é genérica, abrangendo qualquer método específico de classificação que possa ser empregado por um ISP. Pacotes classificados como de baixa prioridade (L) podem sofrer atrasos e/ou perdas maiores dependendo das políticas de escalonamento e descarte empregadas por um ISP.

A DiffProbe requer 2 agentes: um cliente conectado à rede do ISP a ser verificado e um servidor. O funcionamento da ferramenta pode ser dividido em 3 etapas, ilustradas na Figura 4.4. A primeira etapa (1) consiste em um gerador de fluxos de dados. São gerados 2 fluxos diferentes: um fluxo de dados correspondente a uma aplicação suspeita de estar sofrendo DT (A) e outro fluxo de dados de medição (P). Os autores assumem que o fluxo P é da classe H (alta prioridade) e portanto não sofre nenhuma deterioração. A segunda etapa é responsável por executar os fluxos de dados. Os fluxos são enviados simultaneamente pela rede, primeiro do cliente para o servidor e em seguida no sentido contrário. A terceira etapa (3) da ferramenta é responsável por detectar se houve alguma DT. Nesta detecção são utilizadas as medições de atraso e perda de pacotes obtidas pelo cliente e servidor. A detecção baseia-se em uma comparação estatística entre as medidas correspondentes a cada fluxo. Estas três etapas da DiffProbe são descritas em mais detalhes abaixo.

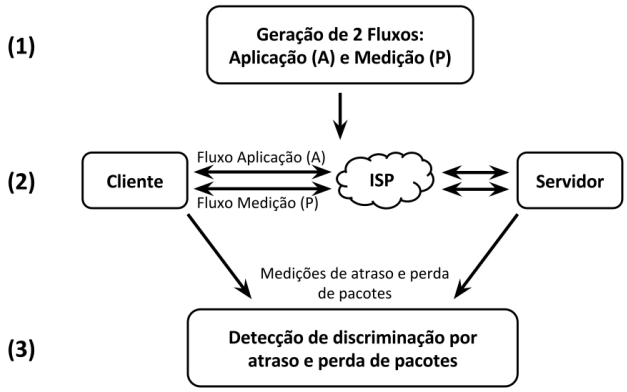


Figura 4.4: Funcionamento da ferramenta DiffProbe.

O fluxo de dados A é gerado a partir de um fluxo real, pré-armazenado, de uma aplicação. A DiffProbe dá duas opções de aplicação para o usuário: Skype e Vonage. Para criar o fluxo A são mantidos os protocolos de transporte, tamanhos de pacotes, portas, dados e intervalos de envio do fluxo original da aplicação. A geração do fluxo P é baseada no fluxo A, mas com restrições: o fluxo P deve ser suficientemente diferente do fluxo A, garantindo que P não seja classificado da mesma forma que A, ou seja, o fluxo P não pode ser classificado como de baixa prioridade. Por outro lado, o fluxo P deve ter características de rede (como tamanho dos pacotes) similares ao fluxo A para que possam ser posteriormente comparados. Na prática, a DiffProbe cria os pacotes do fluxo P

conforme o fluxo A é enviado. Um pacote qualquer do fluxo P tem o mesmo tamanho do último pacote do fluxo A enviado até então, com dados aleatórios e uma porta com baixa probabilidade de ser considerada de baixa prioridade.

A execução dos fluxos é feita em duas fases. Na primeira fase, os pacotes dos dois fluxos são enviados simultaneamente em uma taxa de envio igual. Na segunda fase, a taxa de envio do fluxo P é aumentada, enviando-se mais pacotes do fluxo P do que do fluxo A. O objetivo desta segunda fase é maximizar a chance de ocorrer enfileiramento de pacotes nos roteadores do ISP, já que, como dito anteriormente, não é possível detectar DT quando a carga da rede é baixa e os roteadores não precisam escalonar os pacotes a serem roteados. A DiffProbe não altera a taxa de envio do fluxo A, pois isso pode alterar a classificação do mesmo (se a classificação for baseada em fluxo, por exemplo). Com o aumento da taxa de envio de pacotes do fluxo P, são coletadas mais medidas para o fluxo P do que para o fluxo A. Assim, para os pacotes do fluxo P, a DiffProbe considera na detecção apenas as medidas referentes aos pacotes enviados imediatamente depois de algum pacote do fluxo A, resultando na mesma quantidade de medidas para os 2 fluxos. A primeira fase serve apenas para verificar se a DT é identificável: os maiores valores de atraso do fluxo P na segunda fase devem ser significativamente maiores que os atrasos médios do fluxo P na primeira fase para que a DT seja estatisticamente identificável.

A detecção de DT por atraso é feita comparado-se as distribuições dos atrasos medidos para cada fluxo: no caso de um escalonamento FCFS, os dois fluxos devem apresentar uma distribuição similar de atrasos dos pacotes. Caso exista despriorização dos pacotes do fluxo A, a distribuição dos atrasos do fluxo A será significativamente maior que a distribuição dos atrasos do fluxo P. O teste de igualdade para as distribuições de atrasos usado pela DiffProbe baseia-se na divergência de Kullback-Leibler. A detecção de DT por perda de pacotes também é feita comparando-se as distribuições das taxas de perda medidas para cada fluxo, de forma análoga. Porém, o teste de igualdade utilizado pela ferramenta para esta comparação é o teste Z para comparação de duas proporções (*two-proportion z-test*).

Os autores avaliaram a DiffProbe por meio de simulações, utilizando NS2, e da emulação de um ambiente real. Para este ambiente emulado, foram utilizados um cliente conectado a um ISP residencial e um servidor hospedado em uma universidade. A DT foi emulada por um roteador entre o *host* cliente e o ISP. Tanto as simulações quanto os experimentos no ambiente emulado mostraram que, quando a DT é identificável e o fluxo de medição foi capaz de gerar enfileiramento de mensagens, a detecção foi precisa.

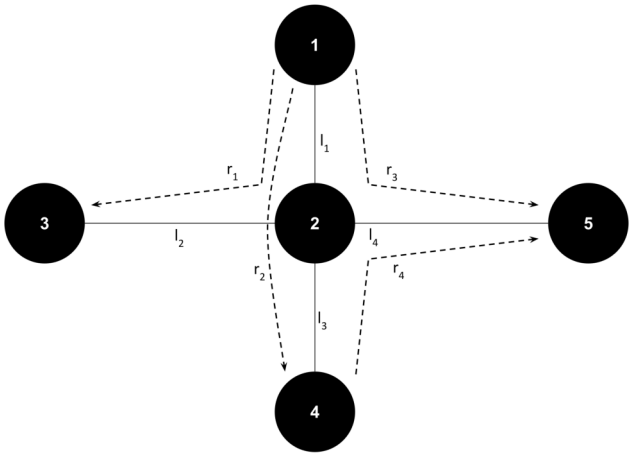
4.5.4. Inferência Baseada em Tomografia de Redes

Em [Zhang et al. 2014] os autores propõem um algoritmo baseado em técnicas de tomografia de rede para inferir se houve DT em uma rede qualquer. O algoritmo também é capaz de identificar especificamente em qual *link* ou sequência de *links* a violação da NR ocorreu, baseando-se apenas em observações externas (medições fim-a-fim), ou seja, sem medir diretamente *links* internos da rede. Os autores fornecem provas formais demonstrando em quais condições o algoritmo atinge estes resultados.

A tomografia de rede [Coates et al. 2002] consiste em inferir métricas sobre os *links* internos de uma rede (como atraso e taxa de perda de pacotes, por exemplo) apenas

a partir de medições fim-a-fim, ou seja, sem medir diretamente cada *link*. Em uma das técnicas existentes, de maneira simplificada, forma-se um sistema de equações $y = Ax$, em que y é um vetor com as medições fim-a-fim, A é uma matriz de roteamento que especifica a relação entre os *links* da rede e as rotas entre os pontos de medição (quais *links* estão em cada rota) e x é um vetor com as métricas a serem inferidas para cada *link*. Obtém-se então uma estimativa de x resolvendo o sistema ou, em caso de múltiplas soluções, escolhendo a solução mais adequada, como a solução que ocorre com a maior probabilidade ou com menor número de *links* problemáticos, por exemplo.

A Figura 4.5 exemplifica a técnica de tomografia de redes utilizada pelo algoritmo. A Figura 4.5a mostra uma rede com 5 *hosts*, numerados de 1 a 5, interligados pelos *links* $l_i, 1 \leq i \leq 4$. Neste exemplo, foram feitas 4 medições fim-a-fim cujas rotas estão representadas na Figura 4.5a por $r_j, 1 \leq j \leq 4$. A Figura 4.5b mostra a matriz de roteamento A para as 4 rotas medidas. Na matriz, as linhas são as rotas medidas e as colunas os *links*. O valor de uma posição da matriz é 1 se a rota (linha) atravessa o *link* (coluna) ou 0 caso contrário. A Figura 4.5c mostra o sistema de equações $y = Ax$ resultante. No sistema, $y = \{y_1, y_2, y_3, y_4\}$ e $x = \{x_1, x_2, x_3, x_4\}$, sendo y_j o valor medido para a rota r_j e x_i o valor da métrica a ser estimado para o *link* l_i. Se o *link* l_4 não for neutro, por exemplo, poderá haver uma inconsistência nas medições referentes às rotas r_3 e r_4, que compartilham este *link*. Neste caso, o valor de x_4 seria efetivamente diferente para cada uma das medições, resultando em um sistema de equações inconsistente e, portanto, sem solução.



(a) Exemplo de rede com 5 *hosts* e 4 medições fim-a-fim.

	l_1	l_2	l_3	l_4
r_1	1	1	0	0
r_2	1	0	1	0
r_3	1	0	0	1
r_4	0	0	1	1

(b) Matriz de roteamento A.

$$y_1 = x_1 + x_2$$
$$y_2 = x_1 + x_3$$
$$y_3 = x_1 + x_4$$
$$y_4 = x_3 + x_4$$

(c) Sistema de equações resultante das 4 medições fim-a-fim.

Figura 4.5: Técnica de tomografia de redes utilizada no algoritmo.

A técnica de tomografia de rede utilizada assume que a rede é neutra: cada *link* trata qualquer tráfego de qualquer rota da mesma forma. Caso isto não ocorra, torna-se impossível expressar as medições de diferentes rotas como função das métricas dos *links*

e o sistema de equações resultante não tem, portanto, solução. Assim, enquanto as técnicas convencionais de tomografia de redes tentam construir sistemas de equações com solução, o algoritmo proposto [Zhang et al. 2014] tenta construir sistemas de equações sem solução, revelando assim violações da NR. A ideia central deste algoritmo é que, quando uma rede não é neutra, observações feitas de pontos de vista distintos serão inconsistentes entre si.

A técnica de tomografia utilizada impõe restrições quanto à métrica empregada para realizar as medições entre os *hosts* finais. Esta deve ser aditiva, ou seja, considerando uma rota entre 2 *hosts* finais, a soma dos valores medidos para cada *link* da rota, utilizando tal métrica, deve ser igual ao valor medido entre os *hosts* finais (a rota toda). Atraso e taxa de perda de pacotes são exemplos de métricas aditivas.

O algoritmo recebe como entrada a topologia da rede e um conjunto de medições fim-a-fim com as respectivas rotas entre os *hosts* finais a partir dos quais as medições foram feitas. A saída do algoritmo é um conjunto de sequências de *links* não neutras, ou seja, em quais *links*, ou sequências de *links*, houve alguma violação da NR. As medições entre os *hosts* finais podem ser feitas utilizando em seus pacotes dados de diferentes aplicações, assim como dados de uma mesma aplicação com origem/destino diferentes. Desta forma, é possível detectar DT baseada tanto no conteúdo quanto na origem ou destino das mensagens.

Como mencionado acima, o algoritmo consiste em buscar sequências de *links* que geram um sistema de equações sem solução. Para cada sequência de *links* que esteja presente em mais de uma rota, forma-se um sistema de equações utilizando todas as medições cujas rotas atravessam esta sequência de *links*. Caso o sistema de equações construído não tenha solução, a sequência de *links* é não neutra. Caso contrário, a sequência é neutra ou a DT para esta sequência de *links* não é identificável (falso-negativo). Em outras palavras, o algoritmo confronta as medições cujos pacotes atravessaram um mesmo segmento da rede, tentando encontrar inconsistências que podem ser atribuídas a alguma DT ocorrida nestes segmentos.

O Algoritmo 1 especifica a estratégia de detecção de DT. O conjunto R é dado como entrada e contém todas as rotas medidas. Nas linhas 3 a 10, cada sequência de *links* λ comum a pelo menos um par de rotas distintas de R é armazenada em Λ_n. Já em π_λ são armazenados os conjuntos de rotas que tem em comum cada sequência λ. No próximo laço, linhas 11 a 16, são consideradas apenas as sequências λ que sejam comuns a pelo menos 2 pares de rotas distintos (linha 12). Para cada uma destas sequências, é formado um sistema de equações utilizando todos os conjuntos de rotas em π_λ. Se este sistema de equações não tiver solução (linha 12), então λ é uma sequência de *links* não neutra e é adicionada em $\Lambda_{\bar{n}}$ (linha 14). Por fim, retorna-se $\Lambda_{\bar{n}}$ contendo todas as sequências de *links* não neutras identificadas (linha 17).

Os autores afirmam que este algoritmo não gera falso-positivos, ou seja, nunca acusa erroneamente uma sequência de *links* como não neutra. A razão para isto é que medições que englobam uma sequência de *links* neutra sempre resultarão em um sistema de equações com solução. Já no caso de falso-negativos, os autores afirmam que ocorrem com pouca frequência. Nestes casos, o algoritmo considera como neutra uma sequência de *links* que na verdade não é neutra.

Algorithm 1 Algoritmo de inferência de NR.

Λ_n: conjunto de sequências de *links* a serem avaliadas
$\Lambda_{\bar{n}}$: conjunto de sequências de *links* não neutras
π_λ: conjunto de conjuntos de rotas que tem a sequência de *links* λ em comum
R: conjunto de todas as rotas medidas
$Links(r)$: sequência de *links* da rota r (atravessados pela medição)
$Sistema(\pi_\lambda)$: sistema de equações formado pelos conjuntos de rotas em π_λ

1: $\Lambda_n \leftarrow \emptyset$
2: $\Lambda_{\bar{n}} \leftarrow \emptyset$
3: **for** cada par de rotas $\{r_i, r_j\} : r_i, r_j \in R, r_i \neq r_j$ **do**
4: $\quad \lambda \leftarrow Links(r_i) \cap Links(r_j)$
5: $\quad$ **if** $\lambda \notin \Lambda_n$ **then**
6: $\quad\quad \Lambda_n \leftarrow \Lambda_n \cup \{\lambda\}$
7: $\quad\quad \pi_\lambda \leftarrow \emptyset$
8: $\quad$ **end if**
9: $\quad \pi_\lambda \leftarrow \pi_\lambda \cup \{\{r_i\}, \{r_j\}, \{r_i, r_j\}\}$
10: **end for**
11: **for** cada sequência $\lambda \in \Lambda_n$ **do**
12: $\quad$ **if** $|\pi_\lambda| > 4$ e $Sistema(\pi_\lambda)$ não tem solução **then**
13: $\quad\quad \Lambda_n \setminus \{\lambda\}$
14: $\quad\quad \Lambda_{\bar{n}} \cup \{\lambda\}$
15: $\quad$ **end if**
16: **end for**
17: retorna $\Lambda_{\bar{n}}$

Para avaliar o algoritmo, foram feitas duas séries de experimentos em ambientes emulados, com diferentes topologias. Primeiramente foi utilizada uma topologia com um único *link* discriminatório. Neste experimento, todas as medições foram feitas atravessando este *link*. Foram testados diferentes cenários, variando o comportamento do *link* discriminatório. Em todos os casos o algoritmo decidiu corretamente se o *link* era neutro ou não. Na segunda série de experimentos foi utilizada uma topologia com diversos *links* discriminatórios. Cada *link* destes teve um comportamento diferente. Assim como na primeira série, o algoritmo detectou corretamente os *links* não neutros em todos os experimentos.

Os autores também discutem os desafios para implementar a solução proposta em um ambiente real. A opção mais viável na prática, segundos os autores, é dispor de um conjunto de *hosts* finais que efetuam periodicamente medições das rotas entre eles e enviam estes dados para serem processados em um servidor central. Também é necessário o uso de alguma solução para descobrir a topologia da rede que conecta os *hosts* envolvidos nas medições, um requisito do algoritmo. Outro desafio é coletar medições a partir de uma quantidade suficiente de pontos de vista diferentes.

4.5.5. NANO

A NANO (*Network Access Neutrality Observatory*) [Tariq et al. 2009] é uma ferramenta cujo objetivo é inferir se um ISP está discriminando o tráfego de alguma aplicação específica. Isto é feito verificando-se se um ISP está causando degradação do desempenho de uma aplicação quando comparado ao desempenho da mesma aplicação em outros ISPs. Se o desempenho de uma aplicação medido na rede de um ISP é estatística e significati-

vamente menor que o desempenho da mesma aplicação medido na rede de outros ISPs, é possível que DT esteja sendo praticada. A NANO utiliza um modelo de inferência causal, tentando estabelecer uma relação entre a degradação de desempenho observada e as políticas de um ISP. As medições de desempenho na NANO são obtidas de forma passiva, ou seja, apenas são feitas medições do tráfego real das aplicações observadas.

As principais diferenças entre a NANO e outras soluções existentes na época de sua publicação para detecção de DT são, segundo os autores: (i) outras soluções detectam discriminação baseada em características específicas como, por exemplo, porta e conteúdo dos pacotes, enquanto a NANO tem um abordagem mais genérica, medindo o desempenho das aplicações independentemente dos mecanismos específicos de DT empregados pelos ISPs; (ii) outras soluções utilizam medições ativas das redes dos ISPs, enquanto a NANO captura suas métricas de forma passiva, o que torna mais difícil para os ISPs detectar e escapar da inferência da NANO; e (iii) as demais soluções comparam métricas de aplicações diferentes em um mesmo ISP, enquanto a NANO compara métricas de uma mesma aplicação em ISPs diferentes.

A estratégia de detecção de DT da NANO apresenta 3 grandes desafios: (i) o mecanismo de DT empregado pelo ISP pode não ser conhecido, assim a estratégia de detecção precisa ser genérica; (ii) o desempenho padrão de uma aplicação em um determinado ISP não é conhecido, dificultando a detecção de possíveis degradações, já que não há um valor base para comparação; e (iii) muitos fatores, além da DT, podem causar degradação no desempenho de aplicações, como sobrecarga, localização geográfica, *software*, *hardware* e outras particularidades da rede.

Os diferentes fatores, além da DT, que podem causar degradação no desempenho de uma aplicação, são representados, no modelo estatístico utilizado pela NANO, por variáveis de confusão [Sander Greenland 1999]. Assim, é necessário identificar quais são as variáveis de confusão e coletar dados não somente sobre o desempenho de aplicações, mas também sobre estas variáveis. A detecção de DT da NANO é feita, portanto, comparando-se o desempenho de uma mesma aplicação em ISPs diferentes, usando medições cujas variáveis de confusão são similares. Um exemplo de variável de confusão é o horário do dia: não se deve comparar medições obtidas em horários distintos, já que aplicações podem ter um desempenho diferente conforme o horário (devido a uma maior carga, por exemplo).

A NANO utiliza a técnica de estratificação para agrupar as medições de desempenho conforme o valor das respectivas variáveis de confusão. Esta técnica coloca cada medição em um estrato, de forma que as variáveis de confusão referentes a cada amostra em um mesmo estrato têm valores similares. São definidas três categorias de variáveis de confusão: (i) variáveis referentes ao cliente (exemplos incluem *softwares* que podem afetar o desempenho da aplicação medida, como sistema operacional ou um navegador Web específico); (ii) variáveis referentes à rede (como localização geográfica, por exemplo); e (iii) variáveis temporais (como o horário do dia, por exemplo, que podem afetar o desempenho da aplicação sendo medida).

Após a estratificação, a NANO estima, para cada estrato, quanto o desempenho de uma aplicação muda quando acessada através de um ISP, em relação ao desempenho obtido quando não se utiliza tal ISP – o desempenho médio (*baseline*). O desempenho

médio é a média do desempenho de todos os outros ISPs dentro do estrato, excluindo o ISP sendo avaliado. Estas estimativas representam uma quantificação da relação causal entre cada ISP e uma possível DT sendo praticada.

A partir das estimativas de cada estrato, o último passo na detecção de DT da NANO consiste em agregar as estimativas de todos os estratos e verificar se os valores obtidos são estatisticamente significativos. A ideia central é que se, na média, o desempenho de uma aplicação degradou-se significativamente ao utilizar-se um ISP específico, então tem-se uma relação causal entre o ISP e a prática de DT.

A implementação da NANO é dividida em duas partes: os agentes e um servidor. Um agente é executado em cada *host* cliente, sendo responsável por monitorar o desempenho da aplicação, medindo seu tráfego real a partir do *host* cliente. As métricas utilizadas são específicas de cada aplicação, conforme o que for mais adequado para cada um. Além dos dados de desempenho das aplicações, os agentes também coletam os dados referentes às variáveis de confusão. Todos os dados adquiridos pelos agentes são enviados periodicamente para o servidor. Os agentes são implementados como *sniffers* de rede, analisando todos os pacotes recebidos e enviados pelo *host*. Já o servidor da NANO recebe todos os dados coletados pelos agentes e é responsável por realizar a detecção de DT com base nestes dados.

A Figura 4.6 ilustra o funcionamento da NANO. Em (1) cada cliente executa um agente. Os agentes monitoram o tráfego real das aplicações sendo avaliadas. Os dados coletados são enviados para o servidor, que primeiramente os separa em estratos conforme as variáveis de confusão (2). Por fim, o servidor infere (3), seguindo o modelo causal, quais ISPs praticaram DT para cada aplicação medida.

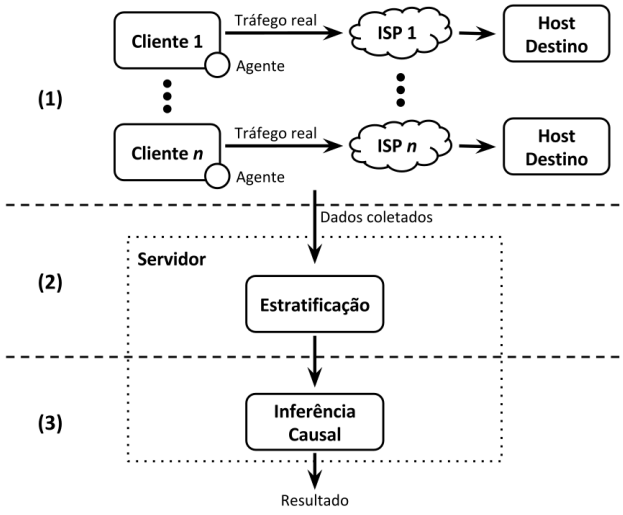


Figura 4.6: Funcionamento da ferramenta NANO.

Para avaliar a NANO, os autores conduziram experimentos em um ambiente controlado, utilizando os *testbeds* PlanetLab e Emulab. Foram selecionados nodos do PlanetLab geograficamente distribuídos. Estes nodos foram utilizados como servidores das aplicações a serem avaliadas. Um conjunto de ISPs foi criado no Emulab, cada um com um conjunto diferente de clientes. Cada ISP fornecia conectividade à Internet para os

seus clientes. Assim, todo acesso dos clientes às aplicações hospedados nos nodos do PlanetLab passava por estes ISPs, permitindo a emulação de diferentes práticas de DT e de diferentes variáveis de confusão.

Os resultados dos experimentos mostraram que a NANO é capaz de detectar DT praticada de diferentes formas e para diferentes tipos de aplicação, desde que todos os fatores que possam confundir significativamente a relação entre um ISP e o desempenho observado de uma aplicação – as variáveis de confusão – sejam conhecidos e medidos. A estratégia de detecção da NANO mostrou-se genérica o suficiente, detectando a discriminação de tráfego mesmo sem conhecer quais as políticas de DT empregadas pelos ISPs.

Porém, se a NANO não considerar todas as variáveis de confusão, a relação causal entre o ISP e uma possível DT pode ser erroneamente calculada – resultando em falso-negativos e falso-positivos. Não há formas automatizadas para enumerar todas as variáveis de confusão pertinentes ou concluir se um conjunto de variáveis de confusão é suficiente, o que pode inviabilizar a aplicação da NANO em um ambiente real.

4.5.6. Gnutella RSP

Os autores em [Beverly et al. 2007] apresentam uma estratégia para quantificar as práticas de bloqueio de portas efetuadas por ISPs na Internet utilizando a rede Gnutella[2]. A estratégia explora o procedimento de ingresso de novos clientes na rede para efetuar medições, aproveitando a infraestrutura já existente da rede Gnutella. Foi um dos primeiros trabalhos publicados sobre medições relacionadas à NR.

Gnutella é uma rede P2P totalmente descentralizada. Os *hosts* participantes da rede são de 2 tipos: os *superpeers* e as folhas (clientes). Cada *superpeer* é conectado com outros *superpeers* e tem um conjunto de folhas conectadas a ele. Para uma nova folha ingressar na rede, é necessário conectar-se a um *superpeer*. O *superpeer* pode aceitar a nova folha, mantendo-a ligada a ele, ou pode informar à folha que está ocupado – caso já tenha muitas folhas, por exemplo. Caso o *superpeer* rejeite a folha, ele indica outro *superpeer* ao qual a folha deve conectar-se para ingressar na rede. Esta indicação contém o endereço IP e a porta TCP do outro *superpeer*.

A estratégia para medição de bloqueio de portas apresentada utiliza 2 *hosts* diferentes: um *host* de medição e um *superpeer* chamado de RSP (*Rogue SuperPeer*). Quando um cliente conecta-se ao RSP para ingressar na rede, o RSP envia uma resposta informando que está ocupado e indica o *host* de medição. O cliente pode então seguir esta indicação e iniciar uma conexão com o *host* de medição na porta indicada. Caso o faça com sucesso, sabe-se então que tal porta não é bloqueada. A ideia central desta estratégia é, portanto, induzir os clientes Gnutella a se conectarem ao *host* de medição para verificar se esta conexão é permitida ou não. Os autores concluíram empiricamente que a probabilidade de um cliente Gnutella não seguir a indicação do RSP, ou seja, não conectar-se ao *host* de medição, é de 80%.

O RSP e o *host* de medição ambos registram as conexões vindas dos clientes Gnutella em um servidor centralizado. Este servidor também é responsável por informar

[2]`http://www.gnutellaforums.com`

qual porta será indicada aos clientes pelo RSP e qual porta o *host* de medição deverá escutar. O servidor altera esta porta a cada 5 minutos, visando obter dados suficientes sobre todas as portas a serem observadas.

Com base nos dados coletados pelo servidor central é feita uma inferência probabilística para determinar quais portas foram bloqueadas. Note que a estratégia proposta considera que o bloqueio de portas pode acontecer em qualquer ponto entre o cliente e o *host* de medição, não sendo capaz de identificar em que parte do caminho o bloqueio aconteceu. Determinar se uma porta não foi bloqueada é trivial: basta que o *host* de medição tenha recebido pelo menos uma conexão de um cliente Gnutella após a indicação do RSP. Porém, caso nenhuma conexão tenha sido feita com o *host* de medição em uma dada porta, isto não implica que tal porta foi bloqueada. É possível que todos os clientes redirecionados para tal porta não tenham seguido a indicação. Assim, segundo os autores, são necessárias pelo menos 50 indicações para concluir, com probabilidade de 99.5%, que uma porta foi bloqueada.

A Figura 4.7 ilustra o funcionamento da estratégia. Em (1) um cliente Gnutella conecta-se no RSP a fim de ingressar na rede. O RSP responde o cliente informando que está ocupado, e indica outro *host* (endereço IP e porta) com o qual o cliente deve tentar conectar-se. O *host* indicado pelo RSP é o *host* de medição. O cliente então inicia uma conexão com o *host* de medição (2). O *host* de medição e o RSP ambos registram as conexões vindas do cliente, informação utilizada para inferir probabilisticamente quais portas foram bloqueadas.

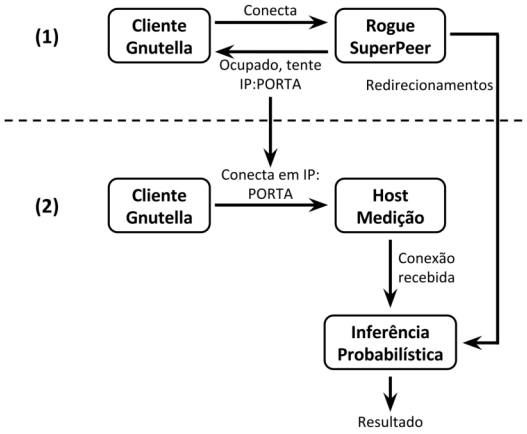


Figura 4.7: Funcionamento da estratégia RSP.

A estratégia do RSP foi executada durante 2 meses. Neste período, o RSP enviou aproximadamente 150 mil indicações para cerca de 72 mil clientes Gnutella distintos, distribuídos em aproximadamente 31 mil prefixos diferentes – uma fração significativa da Internet. Os resultados mostraram que dos 31 mil prefixos, em 256 houve bloqueio de pelo menos uma porta. A porta bloqueada com maior frequência foi a 136 e as bloqueadas com menor frequência foram a 80 (HTTP), 6346 (Gnutella) e 6969 (observada apenas para comparação). Algumas portas referentes a serviços de *e-mail* (25, 110 e 143) foram bloqueadas com frequência cerca de duas vezes maior do que a porta 6969. Depois da 136, as outras portas mais frequentemente bloqueadas foram as referentes ao serviços

FTP, SSH, Bittorrent e VPNs. Os autores também relatam que algumas universidades e ISPs bloquearam portas de serviços P2P (1214, 4662, 6346, 6881) e alguns ISPs no Canadá, E.U.A. e Polônia bloquearam portas do Skype.

4.5.7. Packsen

Packsen [Weinsberg et al. 2011] é um *framework* para geração de fluxos de dados utilizado com o objetivo de detectar se um ISP está praticando DT por meio de engenharia de tráfego. O Packsen também é capaz de inferir qual o tipo de escalonador está sendo empregado e seus parâmetros. A inferência do Packsen é baseada em uma comparação estocástica entre os tempos de chegada dos pacotes de dois fluxos de dados – um fluxo base e um fluxo de medição.

O Packsen considera que um modelador de tráfego (*traffic shaper*) mantém múltiplas filas de pacotes, correspondentes a diferentes classes de tráfego. Cada fluxo de dados é classificado em uma das classes, determinando em qual fila os pacotes do fluxo serão inseridos. Esta classificação pode basear-se em diferentes parâmetros, como protocolo de aplicação, porta, hora do dia, origem, destino, entre outros.

O Packsen utiliza dois fluxos de dados: um fluxo de medição, referente a aplicações específicas, e um fluxo base, o qual se assume não sofrer nenhuma discriminação. Estes fluxos são enviados entre *hosts* finais de forma intercalada e mantendo a mesma largura de banda para os dois fluxos. Se no recebimento dos fluxos for observada uma diferença significativa entre a largura de banda de cada um, então houve DT no caminho entre os dois *hosts*. A métrica utilizada pelo Packsen é, portanto, os tempos de chegada dos pacotes de cada fluxo.

A detecção do Packsen utiliza três métodos. O primeiro método apenas detecta se houve discriminação de um fluxo em relação a outro. Esta detecção é feita comparando-se as distribuições dos tempos de chegada dos pacotes dos dois fluxos. Se a diferença entre as distribuições for estocasticamente significativa, então houve DT. A comparação é feita utilizando o teste U de Mann-Whitney [H. B. Mann 1947]. O segundo método infere qual o tipo de manipulação de tráfego utilizada e quais os parâmetros empregados, como o peso atribuído a cada fluxo, por exemplo. Esta inferência é feita comparando-se a largura de banda dos fluxos no envio com a largura de banda observada no recebimento. Segundo os autores, este método não é robusto na presença de tráfego de fundo (*cross-traffic*). Quando outras aplicações estão gerando uma quantidade significativa de tráfego simultaneamente aos fluxos do Packsen, as larguras de banda dos fluxos podem ser alteradas de forma diferente. É possível que um processador de tráfego classifique o tráfego de fundo e o fluxo de medição como pertencentes à mesma classe, priorizando o fluxo base. Assim, o tráfego de fundo pode influenciar apenas o fluxo de medição e não o fluxo base. O terceiro método trata o tráfego de fundo, sendo capaz de medi-lo, ajustando o monitoramento efetuado. Neste método, as medições do Packsen precisam ser repetidas até que a variação dos resultados seja significativamente baixa.

A implementação do Packsen é dividida em três partes, ilustradas na Figura 4.8: o cliente, o servidor de experimentos e os servidores de medição. O cliente conecta-se ao servidor de experimentos, solicitando um experimento para ser executado (1). O servidor de experimentos escolhe um experimento em seu repositório e retorna-o ao cliente. O

cliente então escolhe um servidor de medição disponível (com baixa carga), informando o experimento que deve ser executado (2). O servidor de medição executa então o experimento, coletando os dados dos fluxos gerados. Os dados são enviados para o servidor de experimento que os armazena para posterior análise (3).

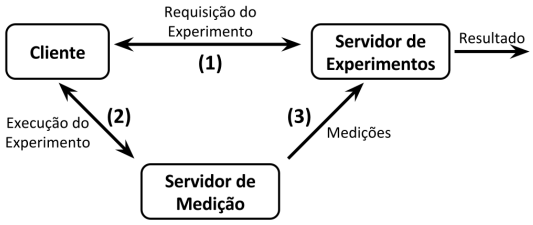


Figura 4.8: Funcionamento do Packsen.

Os autores avaliaram o Packsen primeiramente em um ambiente controlado, utilizando um *testbed* local. Este ambiente permitiu a emulação de diversos tipos de processadores de tráfego, com diferentes parâmetros, assim como diferentes combinações de tráfego de fundo. Após os experimentos no *testbed* local, foram conduzidos experimentos em cerca de 1000 *hosts* do PlanetLab, com o objetivo de melhor avaliar o Packsen em um ambiente real e de maior escala.

Os resultados obtidos no *testbed* local mostraram que o Packsen detectou, com baixa margem de erro, tanto a ocorrência de DT, quanto os parâmetros empregados nos processadores de tráfego, mesmo na presença de tráfego de fundo. Apenas um falso-negativo foi registrado nesses experimentos, no qual houve DT mas o Packsen não a detectou. Já nos experimentos conduzidos no PlanetLab, foi detectada DT em apenas 0.7% dos pares de *hosts* testados (4 de 518).

4.5.8. ChkDiff

ChkDiff [Ravaioli et al. 2012, Ravaioli et al. 2015] é uma ferramenta para a detecção de DT praticada por ISPs que atendem o mercado doméstico (*Tier* 3). O funcionamento da ferramenta consiste em reproduzir o tráfego real do usuário (previamente capturado e preparado) de forma que este tráfego atinja apenas os roteadores a poucos *hops* de distância – o ISP do cliente. São efetuadas medições de atraso e perda de pacotes para cada fluxo de dados presente neste tráfego. A partir destas medições, a ChkDiff é capaz de inferir se houve DT e também identificar a partir de qual roteador a DT aconteceu. Os autores afirmam que a estratégia de medição e detecção da ChkDiff é independente de aplicações específicas e dos mecanismos de DT utilizados pelo ISP. Quaisquer que sejam as aplicações discriminadas ou as técnicas utilizadas para tal, uma DT tipicamente resultará, para o *host* cliente, em maiores atrasos e perdas de pacotes.

O tráfego de dados utilizado pela ChkDiff é obtido capturando-se o tráfego real de um usuário durante uma sessão normal de uso (*trace*). Assim, os resultados produzidos pela ferramenta serão referentes ao conjunto de aplicações executadas pelo usuário durante a captura do *trace*. O *trace* capturado é utilizado com o mínimo de alterações: isto garante que os processadores de tráfego (*traffic shapers*) atravessados pelo *trace* terão o mesmo comportamento que teriam caso os pacotes estivessem sendo gerados pelas res-

pectivas aplicações. As únicas modificações feitas nos pacotes de um *trace* são no campo TTL, para alcançar apenas o *hop* desejado, e nos dados da aplicação, para que todos os pacotes tenham o mesmo tamanho, evitando assim diferentes tempos de transmissão.

A ChkDiff efetua suas medições reproduzindo o *trace* capturado diversas vezes, a partir do *host* cliente. Utiliza-se um valor incremental para o TTL dos pacotes, de forma que cada reprodução do *trace* alcance o roteador seguinte à reprodução anterior. Quando um pacote chega ao roteador ao qual foi destinado (TTL decrementado para zero), o roteador envia uma mensagem ICMP de tempo excedido (*ICMP Time Exceeded*) de volta ao *host* cliente. As medições de atraso e perda de pacotes utilizadas pela ChkDiff são referentes a estas respostas ICMP: o atraso é o RTT entre o envio do pacote e o recebimento da resposta ICMP. A perda de pacotes corresponde à taxa de respostas ICMP não recebidas. O objetivo é avaliar apenas os primeiros roteadores após o *host* do cliente, identificando a partir de qual roteador a DT acontece: assume-se a existência de um processador de tráfego logo antes deste roteador.

A ChkDiff faz uma análise estatística para inferir se um fluxo de dados sofreu DT ou não até o roteador destino (para o qual as medições foram obtidas). Compara-se o atraso e perda de pacotes medidos para este fluxo/roteador com os medidos para todo o resto do tráfego até o mesmo roteador. Se estas medições forem significativamente maiores do que as medições do resto do tráfego, então o fluxo sofreu discriminação a partir daquele roteador. Assim, a base para comparação (*baseline*) utilizada pela ChkDiff é o tráfego todo: a NR estabelece que um fluxo não discriminado é tratado da mesma forma que todo o resto do tráfego, ou seja, as medições obtidas para um fluxo discriminado irão se sobressair em relação ao resto do tráfego. Em um exemplo simplificado, caso a perda de pacotes medida para um fluxo for em torno de 50%, enquanto a perda medida para os demais fluxos for em torno de 10%, é possível que o ISP esteja violando a NR.

O funcionamento da ChkDiff pode ser dividido em 4 etapas, ilustradas na Figura 4.9. Na primeira etapa (1) o tráfego real do usuário é capturado resultando em um *trace*. Na segunda etapa (2) este *trace* é pré-processado, gerando-se um conjunto de *traces* modificados. Na terceira etapa (3), o conjunto de *traces* modificados é reproduzido e as medições são obtidas. Na quarta etapa (4) é feita a análise estatística para inferir se houve discriminação de algum fluxo e identificar a posição do processador de tráfego, relativa ao *host* cliente. Cada etapa é descrita em maiores detalhes abaixo.

Na primeira etapa, a ferramenta captura o tráfego real do *host* cliente. Esta captura é feita durante a atividade regular do usuário na Internet. Como a ChkDiff utiliza o tráfego de envio do usuário (*upstream*), espera-se que durante a captura sejam utilizadas aplicações com envio intenso de dados, como compartilhamento de arquivos, VoIP e mensagens instantâneas, por exemplo.

Na segunda etapa, a ChkDiff processa o *trace* capturado. Este pré-processamento gera um conjunto de *traces* que serão reproduzidos na etapa seguinte. O *trace* é separado em fluxos, agrupando os pacotes segundo 5 itens: endereço de origem e destino, porta de origem e destino e protocolo de transporte. O tamanho de todos os pacotes é padronizado, para evitar que os pacotes tenham tempos de transmissão diferentes, o que geraria erro na análise, já que os atrasos medidos para os pacotes devem ser comparáveis entre si. Com os pacotes padronizados em tamanhos iguais e separados em fluxos, são gerados diversos

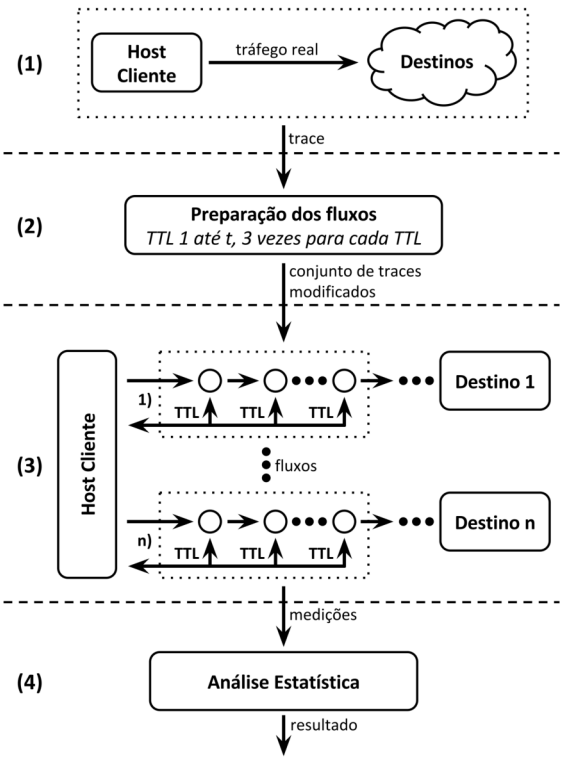


Figura 4.9: Funcionamento da ferramenta ChkDiff.

novos *traces*. Em cada um destes novos *traces*, os pacotes são reordenados e um valor específico de TTL é atribuído. O TTL varia de 1 a t e são criados 3 *traces* para cada valor de TTL. Assim, tem-se um conjunto de $3t$ *traces*, contendo os mesmos pacotes mas em ordem diferente e com valores diferentes de TTL. Os autores afirmam que um valor de 3 ou 4 para t deve ser suficiente para atravessar os roteadores do ISP do cliente.

A reordenação dos pacotes em cada *trace* criado na etapa 2 é feita de forma aleatória, mas sempre mantendo a ordem global dos pacotes do mesmo fluxo. Esta reordenação é necessária para evitar que os fluxos sejam afetados por possíveis vícios (*bias*) nas condições da rede. Segundo os autores, esta técnica também é útil para minimizar problemas como tráfego de fundo e limitação na taxa máxima de respostas ICMP de alguns roteadores. Ao final desta etapa, tem-se então um conjunto de *traces* modificados, prontos para serem reproduzidos.

Na terceira etapa, cada *trace* do conjunto resultante da etapa anterior é reproduzido. Seja h o TTL dos pacotes de um destes *traces*. Cada pacote do *trace* sendo reproduzido é enviado para seu destino e porta originais. Quando um destes pacotes alcança o h-ésimo *hop* a partir do *host* cliente, o roteador presente neste *hop* envia uma mensagem ICMP de tempo excedido para o *host* cliente. Duas medições são efetuadas pela ChkDiff baseadas nestas respostas ICMP: atraso e perda de pacotes, descritas acima. Esta estratégia de medição pode ser prejudicada caso os roteadores do ISP tenham uma taxa limitada de respostas ICMP. Os autores afirmam que, nestes casos, a reordenação dos pacotes em cada *trace* gerado faz com que as respostas ICMP não recebidas fiquem

razoavelmente bem distribuídas entre todos os fluxos do *trace*.

A quarta etapa consiste na análise estatística para inferir se houve discriminação de algum fluxo e identificar a posição do processador de tráfego, relativa ao *host* cliente. A ChkDiff considera nesta análise apenas os fluxos que tiveram pelo menos 20 respostas ICMP recebidas. Como descrito anteriormente, para cada valor h de TTL são reproduzidos e medidos 3 *traces*. Estas 3 repetições diminuem consideravelmente os falso-positivos, conforme os autores concluíram nos experimentos, descritos abaixo. Assim, se para os 3 *traces* o mesmo fluxo falhar no teste estatístico, é considerado que este fluxo sofreu discriminação em relação aos demais. Para as medições de atraso, a ChkDiff compara a distribuição de atraso de cada fluxo com a distribuição de atraso do restante do *trace*. O teste de hipótese utilizado é o Kolmogorov-Smirnov. No caso de uma rede neutra, espera-se que este teste indique que as duas distribuições são iguais. Assim, se um fluxo apresentou atrasos maiores do que o resto do trace, o teste para este fluxo falhou. Para as medições de perda de pacotes, a ChkDiff verifica se a perda de pacotes de cada fluxo é significativamente diferente da perda de pacotes do restante do *trace*. Utiliza-se um teste probabilístico inspirado em uma distribuição binomial. Se um fluxo teve uma perda de pacotes maior do que deveria, o teste probabilístico para este fluxo falhou, isto é a hipótese é falsa. Quando uma DT é detectada para um *hop* h, esta mesma DT será observável para todos os *hops* depois de h. Assim, se a DT não foi detectada para o *hop* anterior, a ChkDiff assume a existência de um processador de tráfego entre o *hop* $h-1$ e h.

A ChkDiff foi avaliada primeiramente em um ambiente neutro, sem nenhuma DT, e posteriormente em um ambiente não-neutro. Em ambos, o *trace* do usuário foi capturado durante um período de 3 minutos de uso típico da Internet. Durante este período foram feitos: envios de imagens em uma rede social, navegação em páginas de notícias e envio de mensagens em aplicativos de *chat*.

No ambiente neutro, a ChkDiff foi executada 100 vezes em uma configuração de rede controlada, em que o roteador no segundo *hop* garantidamente não discriminava nenhum dos fluxos presentes no *trace*. Analisando os resultados com apenas 1 reprodução para cada valor de TTL, cerca de 30% das execuções apresentaram de 1 a 3 falso-positivos. Os autores refizeram o experimento, mas com dois *traces* para cada valor de TTL: não houve nenhum falso-positivo. Com base nesta avaliação preliminar, os autores fixaram em 3 a quantidade de *traces* gerados para cada valor de TTL, como descrito anteriormente.

A avaliação em ambiente não-neutro foi feita primeiramente com apenas um fluxo discriminado. Posteriormente foram utilizados múltiplos fluxos discriminados, com diferentes frações do *trace* contendo fluxos discriminados. Em ambas as avaliações foi utilizada uma configuração de rede controlada. O *host* do usuário foi conectado a um *host* intermediário (*middlebox*), o qual fornecia acesso à Internet para o *host* cliente e também operava como um processador de tráfego (*traffic shaper*). Este *host* intermediário foi conectado a um roteador, no qual o TTL dos pacotes expirava. Foi utilizada a ferramenta Dummynet [Carbone and Rizzo 2010] no *host* intermediário para emular as práticas de DT. A DT foi implementada de duas formas diferentes: limitando a largura de banda dos fluxos selecionados e descartando pacotes dos fluxos selecionados de forma mais frequente.

Nos experimentos com apenas um fluxo discriminado, a ChkDiff foi capaz de detectar corretamente 100% dos fluxos discriminados por limitação de banda. Já quando a DT foi por descarte de pacotes, foram observados alguns falso-negativos (uma discriminação ocorreu mas não foi detectada). Nos experimentos com múltiplos fluxos discriminados, a análise estatística da ChkDiff deixou de funcionar corretamente quando a fração de fluxos discriminados é grande (cerca de 80% ou mais). A ChkDiff também foi avaliada na presença de um limite na taxa de envio de respostas ICMP do roteador. Os resultados mostraram que a ChkDiff manteve bons resultados na presença de tal limitação.

4.5.9. POPI

A POPI [Lu et al. 2010] é uma ferramenta que utiliza medições fim-a-fim para inferir se existe priorização no encaminhamento de pacotes de tipos diferentes (*packet forwarding prioritization*). A métrica utilizada pela POPI é a taxa de perda de pacotes, medida para tráfegos e diferentes tipos. O POPI considera que a priorização no encaminhamento de pacotes é feita seguindo alguma das estratégias de escalonamento da engenharia de tráfego, descritos anteriormente. A ideia central da POPI é que em uma rede neutra, todos os pacotes são encaminhados pelos roteadores conforme a ordem de chegada. Assim, caso a rede esteja congestionada e o descarte de pacotes seja necessário, tráfegos de tipos diferentes sofrerão uma taxa de perda similar. Porém, caso os pacotes de um tipo sejam encaminhados pelos roteadores com uma maior prioridade em relação aos pacotes de outros tipos, a taxa de perda de pacotes será diferente, configurando assim uma DT.

O funcionamento da POPI consiste em 3 etapas, ilustradas na Figura 4.10. A primeira etapa (1) consiste nas medições, que são obtidas em uma série de rajadas de mensagens. Na segunda etapa (2) as medições são computadas a fim de obter-se uma ordenação dos tipos de tráfego que sofreram maiores perdas de pacotes, para cada rajada. A terceira etapa (3) consiste em uma análise estatística para verificar se houve priorização de tipos específicos de tráfego ao longo das rajadas. Cada uma destas etapas é descrita abaixo.

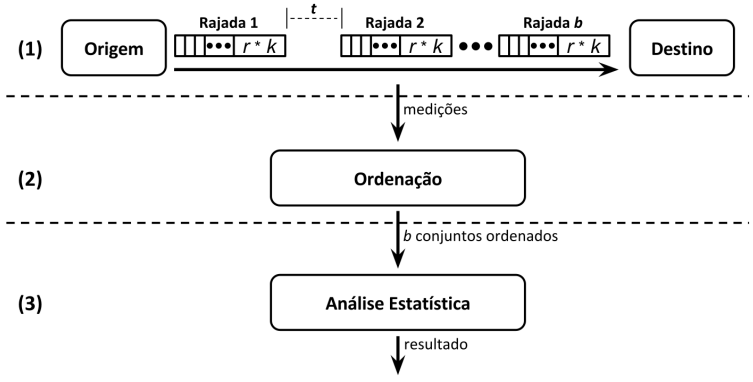


Figura 4.10: Funcionamento da ferramenta POPI.

Na primeira etapa (1), a POPI efetua as medições de perda de pacotes para k tipos de tráfego. As medições são feitas em b rajadas de pacotes, enviados entre um *host* origem e um *host* destino. As rajadas são separadas por intervalos de t segundos. Cada rajada é composta de r rodadas e em cada rodada são enviados k pacotes – um para cada tipo

de tráfego sendo avaliado, em ordem aleatória. Assim, em cada rajada são enviados $r * k$ pacotes em sequência. Segundo os autores, o valor de t não pode ser muito baixo, para que uma rajada não interfira na próxima, mas também não pode ser muito alto para que a medição toda termine dentro de um período curto de tempo, tornando-a menos suscetível a flutuações no tráfego de fundo. Os autores afirmam também que é necessário enviar uma quantidade grande de mensagens para garantir que os roteadores entre a origem e o destino fiquem congestionados e comecem a descartar pacotes, não dependendo assim do tráfego de fundo para tal.

Na segunda etapa (2), as taxas de perda de pacotes para todos os tipos de tráfego em cada rajada são computadas e ordenadas. Nesta ordenação, o tipo de tráfego com a maior taxa de perda de pacotes em uma dada rajada, por exemplo, fica na primeira posição, o tipo com a segunda maior taxa fica na segunda posição e assim por diante. Segundo os autores, se os pacotes de todos os tipos forem tratados de forma neutra, as posições de tipos diferentes formarão arranjos aleatórios ao longo de todas as rajadas, já que a ordem de envio dos pacotes de tipos diferentes em cada rodada é aleatória. Porém, caso alguns tipos de tráfego tenham baixa prioridade, estes tipos estarão sempre nas primeiras posições da ordenação. Assim, ao fim desta etapa, tem-se b conjuntos de tipos de tráfego – um para cada rajada – ordenados conforme as taxas de perda de pacotes observadas em cada rajada.

Na terceira etapa (3), é feita uma análise estatística para verificar se houve priorização de tipos específicos de tráfego ao longo das rajadas. Segundos os autores, se a POPI comparasse apenas dois tipos de tráfego diferentes, bastaria determinar se estes tipos foram tratados de forma diferente comparando as medições obtidas para cada um. Porém, para analisar mais de dois tipos de tráfegos é necessário agrupá-los conforme suas prioridades. Verificar se as posições relativas de k valores repetem-se de forma consistente ao longo de b observações é um problema estatístico conhecido como *"Problem of N Rankings"* [Noether 2012]. A solução adotada na POPI foi calcular uma média das posições de cada tipo de tráfego em todas as rajadas (*Average Normalized Ranks*) e agrupar os tipos cujas médias não apresentam diferença significativa. Este agrupamento é feito utilizando um método hierárquico divisivo. Ao fim deste processo, tem-se grupos de tipos de tráfego ordenados conforme suas prioridades. Em uma rede neutra, esta análise resulta em apenas um grupo, já que todos os tipos de tráfego apresentam a mesma taxa média de perda de pacotes e, portanto, têm a mesma prioridade.

Para avaliar a POPI, os autores primeiramente realizaram simulações utilizando o simulador de rede NS2. Nestas simulações foram utilizados dois pares de *hosts* origem/destino. Um destes pares foi responsável por simular o tráfego de fundo, enquanto o outro par simulou a execução da POPI. Na topologia usada nas simulações, a comunicação entre ambos os pares atravessa os mesmos dois roteadores, responsáveis por simular a priorização de determinados tipos de tráfego, com uma largura de banda máxima de 100 Mbps. O valor utilizado nas simulações para k e b foi 32, ou seja, 32 tipos de tráfego e 32 rajadas. Foram utilizados valores incrementais para r – a quantidade de rodadas por rajada. A taxa de envio do tráfego de fundo também variou de 10 a 90 Mbps. Os resultados obtidos nestas simulações mostraram que a POPI foi capaz de obter bons resultados mesmo na presença de uma grande quantidade de tráfego de fundo: os pacotes de baixa prioridade foram sempre descartados antes dos de alta prioridade. Outro resultado obtido

foi quanto ao valor de r. Para $r < 18$ o tráfego de medição não foi capaz de congestionar as filas dos roteadores, não gerando nenhuma perda de pacotes, impossibilitando assim a inferência. Conforme o valor de r aumenta, a perda de pacotes começa a ser observada de forma mais frequente para os tipos de tráfego de baixa prioridade. Com base nos resultados, os autores afirmam que $r > 30$ é suficiente para obter resultados confiáveis. Assim, foi utilizado $r = 40$ nos experimentos conduzidos no PlanetLab, descritos abaixo.

Foram conduzidos experimentos no PlanetLab para avaliar a POPI em um ambiente real e encontrar possíveis casos reais de priorização. Nestes experimentos foram utilizados 162 nodos do *testbed*, espalhados em diversos continentes. A POPI foi executada em todos os pares de nodos e em ambos os sentidos para cada par. Os valores utilizados das variáveis foram: $k = 26$, $b = 32$, $r = 40$ e $t = 10s$. O tamanho dos pacotes enviados foi de 1500 *bytes* cada, o que gerou um consumo de banda médio de 1.04 Mbps. Os resultados obtidos indicaram que houve algum tipo de priorização de pacotes para 15 pares de nodos. Os autores também executaram a POPI utilizando outras métricas com menor sobrecarga de medição, descritas acima. Os resultados para estas outras métricas mostraram que estas não foram capazes de detectar muitos dos casos de priorização detectados nos experimentos que utilizaram a taxa de perda de pacotes como métrica.

4.5.10. Resumo Comparativo das Soluções

Um resumo comparativo das soluções é mostrado na Tabela 4.1. Para cada solução, a Tabela mostra a topologia de medição utilizada, a(s) métrica(s) empregada(s), que tipo de comparação é feita (tráfegos de tipos diferentes, por exemplo), resultados obtidos e observações específicas sobre a solução e suas limitações.

4.5.11. Outras Soluções Relacionadas

Esta subseção descreve outros trabalhos relacionados à NR que não se referem diretamente à detecção de DT. Estes outros trabalhos tratam de outras práticas que também podem ser consideradas como violações da NR, por exemplo censura ou qualidade de serviço inferior à contratada, além de outras soluções que podem ser utilizadas no contexto da NR.

Soluções que medem a **qualidade do serviço de ISPs** efetivamente entregue ao usuário e/ou monitoram a conformidade do serviço fornecido com acordos de nível de serviço (SLA – *Service-Level Agreement*), também têm relação com a NR, já que a normatização da NR de alguns países incluem estes temas. É importante destacar que algumas destas soluções surgiram devido à preocupação de governos em garantir o cumprimento da NR em relação a estes quesitos. Existem diversos trabalhos [Bischof et al. 2012, Sánchez et al. 2011, Aida et al. 2003, SamKnows , Ookla , TestMy.net , Broadband Speed Checker , NIC.br , Sommers et al. 2007, Sommers et al. 2010, Ta and Mao 2006, Serral-Gracia et al. 2009, Qiu et al. 2008, Serral-Gracià et al. 2010, Yuksel et al. 2010, Hourton et al. 2012] focados em resolver estas questões, não necessariamente motivados pelo debate da NR.

A HAKOMetar [Weber et al. 2013] é uma ferramenta que permite a um usuário final verificar a qualidade de serviço que seu ISP está lhe fornecendo. Esta ferramenta foi desenvolvida pela HAKOM, a agência reguladora das telecomunicações da Croá-

Tabela 4.1: Comparação das características principais das soluções.

Solução	Medição	Métrica	Comparação	Resultados	Observações
Glasnost	Entre um *host* final e um servidor de medição	Taxa de transferência	Aplicação **X** Dados aleatórios	10% dos usuários sofreram DT de BitTorrent e 6% relataram possíveis falso-negativos	Detecta apenas DT baseada em porta e protocolo de aplicação
NetPolice	A partir de diversos *hosts* finais	Perda de pacotes (das respostas ICMP)	Diversos protocolos **X** HTTP	4 dos 18 ISPs avaliados realizaram DT em 4 aplicações e 10 ISPs realizaram DT baseada no AS anterior dos pacotes	Requer acesso a múltiplos *hosts* e a medição depende de respostas ICMP, que nem sempre são suportadas
DiffProbe	Entre um *host* final e um servidor de medição	Atraso e perda de pacotes	Skype/Vonage **X** Dados aleatórios	Simulações e experimentos em ambiente emulado mostraram que a detecção foi precisa.	Primeiramente congestiona a rede, inserindo uma grande quantidade de tráfego artifical na rede
Tomografia	Fim-a-fim entre diversos pares de *hosts*	Qualquer métrica aditiva	Técnica de tomografia de redes	Em ambiente emulado, o algoritmo inferiu corretamente a presença e localização da DT	É necessário conhecer a topologia da rede e ter acesso a uma grande quantidade de *hosts*
NANO	Captura passivamente o tráfego real de aplicações	Depende da aplicação	Mesma aplicação em ISPs diferentes	Em ambiente emulado, a NANO foi capaz de detectar DT praticada de diferentes formas e para diferentes tipos de aplicação	Caso as variáveis de confusão não forem todas conhecidas, a inferência pode ser efetuada erroneamente.
Gnutella RSP	Induz clientes Gnutella a tentarem se conectar a um servidor de medição	Bloqueio de portas	Diversos protocolos (número das portas)	Houve bloqueio de pelo menos 1 porta em 256 de 31 mil prefixos e as portas mais bloqueadas foram a 136 e as referentes aos serviços FTP, SSH, Bittorrent e VPNs	Explora o procedimento de ingresso de novos clientes na rede Gnutella, detectando DT baseada em bloqueio de portas
Packsen	Entre um *host* final e um servidor de medição	Tempos de chegada dos pacotes	Aplicação **X** Não-discriminado	Em um *testbed* local, a Packsen detectou com baixa margem de erro tanto a presença e parâmetros da DT, mesmo na presença de tráfego de fundo	Assume a presença de um modelador de tráfego e envia uma grande quantidade de dados para forçar este modelador a enfileirar os pacotes
ChkDiff	Reproduz tráfego real (previamente capturado) a partir de um *host* final	Atraso e perda de pacotes (das respostas ICMP)	Cada fluxo **X** Restante do tráfego	Em ambiente emulado, a ChkDiff detectou com baixa margem de erro os casos de DT por limitação de banda e por descarte de pacotes	A medição depende de respostas ICMP, que nem sempre são suportadas
POPI	Entre dois *hosts* finais	Perda de pacotes	Diversas aplicações são agrupadas conforme similaridade das medições	Experimentos com 162 nodos espalhados em diversos continentes mostraram que houve DT em 15 pares de nodos.	Congestiona a rede antes de efetuar medições enviando grande quantidade de dados

cia. A estratégia da agência em relação à NR é utilizar a HAKOMetar para aumentar a transparência e competitividade no mercado de banda larga do país. O desenvolvimento da ferramenta baseou-se em resultados anteriores acerca de práticas de gestão de tráfego, obtidos em experimentos conduzidos em ambientes de teste na Croácia [Jukic et al. 2011]. A medição da HAKOMetar é feita em três etapas. Na primeira etapa são recolhidos dados sobre o *host* em que a HAKOMetar está sendo executado e sobre sua rede local. Na segunda etapa, mede-se a taxa de envio e recebimento de dados entre o cliente e o servidor, assim como outras propriedades como latência. Na última etapa, a HAKOMetar cria uma grande quantidade de conexões em paralelo com diversos destinos

diferentes, transferindo uma grande quantidade de dados entre o cliente e estes destinos (usando HTTP e/ou FTP). A partir dos resultados da medição, o usuário pode comparar se a largura de banda medida é a mesma que a contratada. Segundo os autores, os resultados reais obtidos com a HAKOMetar indicam que a ferramenta efetivamente aumentou a transparência do mercado de banda larga na Croácia, já que os consumidores passaram a poder verificar se a qualidade do serviço fornecido pelos ISPs estava de acordo com a contratada. Os autores também afirmam que ainda é necessário incluir mais medições na ferramenta para que ela possa ser usada para detectar violações da NR.

A Adkintun [Bustos-Jiménez et al. 2013] é uma solução para monitoramento da qualidade do serviço de banda larga no Chile. A solução foi desenvolvida pelo *NIC Chile Research Labs* à pedido da Secretaria Nacional de Telecomunicações do Chile (SUBTEL), com o objetivo de monitorar o cumprimento da Lei de NR vigente no país. As medições da Adkintun são efetuadas periodicamente por um *software* cliente instalado nos *hosts* dos usuários finais ou embarcado em roteadores residenciais fornecidos a alguns usuários selecionados. Um servidor central informa periodicamente aos *softwares* cliente quais medições devem ser feitas e quais *hosts* destino devem ser utilizados nestas medições. As medições da Adkintun incluem disponibilidade, taxa de transferência, latência, perda de pacotes, bloqueio de portas, entre outras. Os *hosts* destino utilizados nas medições podem estar localizados dentro da infraestrutura do ISP do cliente, em outro ISP chileno ou ainda em uma localização internacional, dependendo da medição. Todas as medições são coletadas por um servidor central, que disponibiliza todos os resultados publicamente em uma página Web. Assim, é possível consultar dados históricos da qualidade de serviço de cada ISP chileno monitorado pela Adkintun. Os autores afirmam que a Adkintun tem proporcionado aos cidadãos meios para proteger seus direitos, já que os resultados obtidos pela ferramenta estão sendo utilizados como base para reclamações de usuários contra a má qualidade dos serviços prestados por ISPs e até mesmo como evidência em processos judiciais envolvendo a SUBTEL e ISPs. Também foi desenvolvida uma versão da ferramenta para redes móveis, a Adkintun Mobile [Bustos-Jiménez et al. 2013, Lalanne et al. 2015]. Esta versão utiliza uma combinação de medições passivas com algumas medições ativas efetuadas em dispositivos móveis, com o objetivo de monitorar a qualidade do serviço de Internet móvel no Chile.

Além da qualidade de serviço, a liberdade de escolha dos usuários quanto ao conteúdo que desejam acessar também está inserida no contexto da NR. Assim, trabalhos sobre **detecção de censura na Internet** também têm relação com o tema. A censura na Internet ocorre, por exemplo, quando usuários têm seu acesso bloqueado a determinadas páginas Web ou serviços. Existem diversas soluções para detecção de censura [Sfakianakis et al. 2011, Net Neutrality Monitor , Hwang 2007, Network of Excellence in InterNet Science , Filasto and Appelbaum 2012]. Estas soluções monitoram a rede efetuando medições periodicamente, criando assim um "censo" sobre assuntos, serviços e páginas Web bloqueados e/ou filtrados. Um *survey* bastante completo sobre detecção de censura na Internet foi publicado recentemente [Aceto and Pescapé 2015].

Um tema relacionado à censura é a **modificação de conteúdo**. Exemplos desta prática incluem: alterar o conteúdo de uma página Web (inserindo anúncios, por exemplo), injetar pacotes forjados em um fluxo de comunicação ou ainda modificar o conteúdo dos pacotes (prejudicando a integridade dos dados transferidos por BitTorrent, por exem-

plo). Existem algumas soluções para detectar estes tipos de práticas. A Switzerland [Electronic Frontier Foundation] é uma ferramenta para a detecção de modificação e injeção de pacotes de dados trafegando na Internet. Já em [Reis et al. 2008] os autores apresentam uma solução para detectar modificações feitas em páginas Web no caminho entre o servidor e o usuário final, como inserção de anúncios e códigos maliciosos, por exemplo.

Diversos trabalhos sobre **medições de rede** podem ser utilizados no contexto de NR. Os dados coletados por plataformas e serviços de medição [Dhawan et al. 2012, Dischinger et al. 2007, Mahajan et al. 2008, Bischof et al. 2011, Sánchez et al. 2013, Dovrolis et al. 2010, Trestian et al. 2009, Antoniades et al. 2010, Miorandi et al. 2013, Molavi Kakhki et al. 2015] podem ser utilizados para detecção de violações da NR. Estas soluções monitoram continuamente diversas propriedades da rede de diversos ISPs, possibilitando também uma comparação de desempenho entre ISPs distintos. Um *survey* completo sobre plataformas de medição na Internet foi publicado recentemente [Bajpai and Schönwälder 2015]. Diversas técnicas para medição de rede e geração de tráfego [Vishwanath and Vahdat 2009, Michaut and Lepage 2005, Basso et al. 2013, Kanuparthy and Dovrolis 2011, Cheng et al. 2004, Botta et al. 2012, Detal et al. 2013] também podem ser utilizadas na detecção de DT (empregando tipos diferentes de tráfego) e para medir qualidade de serviço de ISPs. São descritos abaixo alguns trabalhos sobre medição de rede diretamente voltados para a obtenção de dados que podem ser utilizados na detecção de algum tipo de violação da NR.

A Neubot (*Network Neutrality Bot*) [Martin and Glorioso 2008, Basso et al. 2011] é uma plataforma de *software* para a obtenção contínua de medições distribuídas na Internet. A Neubot permite a implementação de ferramentas e estratégias para verificar a qualidade do serviço oferecido por ISPs, conforme diferentes protocolos e/ou aplicações são utilizados nas medições. As medições implementadas na Neubot são executadas periodicamente em *hosts* finais e todos os dados obtidos são disponibilizados publicamente. As medições de rede já implementadas pela Neubot incluem os protocolos HTTP, BitTorrent, RTP, VoIP, entre outros. Destaca-se que a Neubot não implementa a detecção de violações da NR propriamente dita, servindo para obtenção de medições que poderão ser utilizadas para tal. Desde fevereiro de 2012 a Neubot efetua suas medições dentro da plataforma de medição Measurement Lab [Dovrolis et al. 2010], utilizando os diversos servidores de medição disponibilizados pela plataforma. Os autores afirmam que a grande quantidade de dados de medição referentes à qualidade da Internet de usuário finais permite uma análise sistemática dos serviços de Internet sendo oferecidos pelos ISPs. Os autores afirmam ainda que os os dados coletados pela Neubot podem trazer um melhor entendimento da NR baseado em dados reais, contribuindo assim com o atual debate mundial.

O Netalyzr [Kreibich et al. 2010] é um serviço de medição de rede proposto no contexto da NR. O objetivo deste serviço é avaliar a conexão de Internet de usuários finais, coletando dados que podem ser utilizados para identificar violações da NR e problemas de rede. O projeto do Netalyzr visa abranger a maior quantidade possível de métricas e ser de fácil utilização por usuários sem conhecimento técnico. O Netalyzr é implementado como um *applet* Java (executado em um navegador) que se comunica com diversos servidores de medição. São efetuadas diversas medições referentes a diversos protocolos (como TCP,

UDP, HTTP e DNS), à rede local do usuário (como NAT e *buffers*), ao ISP de acesso (como suporte IPv6, modificação de conteúdo, filtragem de portas, largura de banda e latência), entre outras. Além de informar as medições ao usuário, a ferramenta também funciona como um serviço de monitoramento contínuo de *hosts* na borda da Internet, já que armazena todas as medições feitas pelos diversos usuários finais na Internet. Assim, tem-se uma grande base de dados que pode ser utilizada para a detecção de diversas características da rede, inclusive para estudos referentes à NR. Os autores apresentam uma análise sobre 130.000 medições registradas pelo Netalyzr, as quais foram disponibilizadas publicamente.

A NNMA (*NNSquad Network Measurement Agent*) [Network Neutrality Squad] é uma ferramenta que monitora a atividade de rede dos *hosts* em que está instalada. Este monitoramento tem como objetivo obter diversas medições de rede que possam posteriormente auxiliar na detecção de violações da NR e problemas na rede. No contexto de NR, a principal medição feita pela NNMA é a identificação de pacotes forjados do tipo RST (*reset*) do protocolo TCP. Um pacote RST indica que um dos *hosts* encerrou a conexão e não irá mais enviar ou receber pacotes. ISPs podem injetar pacotes RST forjados para encerrar conexões referentes à algum tipo de tráfego específico [Weaver et al. 2009]. A NNMA não efetua nenhuma comparação sobre a quantidade de pacotes RST forjados para o tráfego de diferentes aplicações. Porém, é uma métrica que pode ser utilizada para inferir se um ISP está efetuando DT utilizando tal prática.

4.6. Conclusão

A Neutralidade da Rede é um tema em crescente discussão ao redor do mundo. Conforme crescem a quantidade de usuários e os serviços oferecidos na Internet, práticas de DT tornam-se cada vez mais comuns. É mais barato para os ISPs bloquearem ou degradarem conteúdos que demandem alto desempenho da rede do que investir em melhorias de infraestrutura ou em soluções não discriminatórias. Além do aumento da demanda, muitos ISPs empregam a DT por interesses comerciais, priorizando seus próprios serviços em detrimento dos serviços concorrentes, por exemplo.

Apesar de diversos países já terem criado regulamentações, Leis, entre outros, que exigem redes neutras por parte dos ISPs, fiscalizar se a NR está sendo cumprida pelas operadoras ainda é um desafio. Este minicurso apresentou diversas soluções para a detecção de DT. As soluções apresentadas baseiam-se, em geral, em medições de rede e métodos estatísticos para inferir se um ISP está discriminando um tipo de tráfego de dados em relação a outros. Estas soluções diferem, principalmente, na topologia utilizada para medição, nas métricas empregadas, no tipo de comparação estatística realizada e nos requisitos e limitações das estratégias adotadas.

Também foi apresentada neste minicurso uma introdução sobre o debate da NR e os conceitos relacionados, assim como um panorama da normatização da NR ao redor do mundo. Uma linha do tempo com diversos incidentes relacionados à NR que ocorreram durante o período do debate também foi apresentada, destacando a importância e atualidade do tema.

Espera-se que este minicurso fomente o debate sobre a NR no Brasil, em especial na comunidade de pesquisa em redes de computadores e sistemas distribuídos.

Referências

[Aceto and Pescapé 2015] Aceto, G. and Pescapé, A. (2015). Internet Censorship detection: A survey. *Computer Networks*, 83.

[Aida et al. 2003] Aida, M., Miyoshi, N., and Ishibashi, K. (2003). A scalable and lightweight QoS monitoring technique combining passive and active approaches. In *IEEE INFOCOM*, volume 1.

[American Cable Association 2016] American Cable Association (2016). ACA Statement On Netflix's Throttling Of Wireless Video Streaming Traffic. `http://www.americancable.org/node/5668`. Acessado em 25/01/2017.

[Anderson 2013] Anderson, C. (2013). Dimming the Internet: Detecting Throttling as a Mechanism of Censorship in Iran. `http://arxiv.org/abs/1306.4361`. Acessado em 25/01/2017.

[Antoniades et al. 2010] Antoniades, D., Markatos, E. P., and Dovrolis, C. (2010). *MOR: Monitoring and Measurements through the Onion Router*, pages 131–140. Springer Berlin Heidelberg.

[Austen 2005] Austen, I. (2005). A Canadian Telecom's Labor Dispute Leads to Blocked Web Sites and Questions of Censorship. `http://www.nytimes.com/2005/08/01/business/worldbusiness/a-canadian-telecoms-labor-dispute-leads-to-blocked.html`. Acessado em 25/01/2017.

[Bajpai and Schönwälder 2015] Bajpai, V. and Schönwälder, J. (2015). A Survey on Internet Performance Measurement Platforms and Related Standardization Efforts. *IEEE Communications Surveys Tutorials*, 17(3).

[Bashko et al. 2013] Bashko, V., Melnikov, N., Sehgal, A., and Schönwälder, J. (2013). Bonafide: A traffic shaping detection tool for mobile networks. In *IFIP/IEEE International Symposium on Integrated Network Management (IM)*.

[Basso et al. 2013] Basso, S., Meo, M., and De Martin, J. C. (2013). Strengthening Measurements from the Edges: Application-level Packet Loss Rate Estimation. *SIGCOMM Computer Communication Review*, 43(3).

[Basso et al. 2011] Basso, S., Servetti, A., and Martin, J. C. D. (2011). The network neutrality bot architecture: A preliminary approach for self-monitoring of Internet access QoS. In *Computers and Communications (ISCC), 2011 IEEE Symposium on*.

[Berners-Lee 2010] Berners-Lee, T. (2010). Long Live the Web. *Scientific American*, 303(6).

[Beverly et al. 2007] Beverly, R., Bauer, S., and Berger, A. (2007). The Internet is Not a Big Truck: Toward Quantifying Network Neutrality. In *International Conference on Passive and Active Network Measurement (PAM)*. Springer-Verlag.

[Bischof et al. 2012] Bischof, Z. S., Otto, J. S., and Bustamante, F. E. (2012). Up, Down and Around the Stack: ISP Characterization from Network Intensive Applications. *SIGCOMM Computer Communication Review*, 42(4).

[Bischof et al. 2011] Bischof, Z. S., Otto, J. S., Sánchez, M. A., Rula, J. P., Choffnes, D. R., and Bustamante, F. E. (2011). Crowdsourcing ISP Characterization to the Network Edge. In *SIGCOMM Workshop on Measurements Up the Stack (W-MUST)*. ACM.

[Body of European Regulators for Electronic Communications] Body of European Regulators for Electronic Communications. All you need to know about Net Neutrality rules in the EU. `http://berec.europa.eu/eng/net/introduction`. Acessado em 25/01/2017.

[Body of European Regulators for Electronic Communications 2012a] Body of European Regulators for Electronic Communications (2012a). BEREC Guidelines for quality of service in the scope of net neutrality. `http://berec.europa.eu/eng/document_register/subject_matter/berec/regulatory_best_practices/guidelines/1101-berec-guidelines-for-quality-of-service-in-the-scope-of-net-neutrality`. Acessado em 25/01/2017.

[Body of European Regulators for Electronic Communications 2012b] Body of European Regulators for Electronic Communications (2012b). Summary of BEREC positions on net neutrality. http://berec.europa.eu/eng/document_register/subject_matter/berec/opinions/1128-summary-of-berec-positions-on-net-neutrality. Acessado em 25/01/2017.

[Botta et al. 2012] Botta, A., Dainotti, A., and Pescapé, A. (2012). A Tool for the Generation of Realistic Network Workload for Emerging Networking Scenarios. *Computer Networks*, 56(15).

[Broadband Speed Checker] Broadband Speed Checker. The UK's No.1 Broadband Speed Test. http://www.broadbandspeedchecker.co.uk. Acessado em 25/01/2017.

[Brodkin 2014] Brodkin, J. (2014). Netflix performance on Verizon and Comcast has been dropping for months. *http://arstechnica.com/information-technology/2014/02/netflix-performance-on-verizon-and-comcast-has-been-dropping-for-months*. Accessed in October 19, 2016.

[Bustos-Jiménez et al. 2013] Bustos-Jiménez, J., Del Canto, G., Pereira, S., Lalanne, F., Piquer, J., Hourton, G., Cádiz, A., and Ramiro, V. (2013). How AdkintunMobile Measured the World. In *ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication (UbiComp)*. ACM.

[Bustos-Jiménez and Fuenzalida 2014] Bustos-Jiménez, J. and Fuenzalida, C. (2014). All Packets Are Equal, but Some Are More Equal Than Others. In *Latin America Networking Conference (LANC)*. ACM.

[Bustos-Jiménez et al. 2013] Bustos-Jiménez, J., Ramiro, V., Lalanne, F., and Barros, T. (2013). Adkintun: SLA Monitoring of ISP Broadband Offerings. In *International Conference on Advanced Information Networking and Applications Workshops (WAINA)*.

[Campbell 2016] Campbell, P. S. (2016). Public Interest Groups Urge FCC Action Against Zero-Rating. http://www.lexology.com/library/detail.aspx?g=e4fbf6ad-03f4-4a04-83c9-f4220c6dea26. Acessado em 25/01/2017.

[Canadian Radio-television and Telecommunications Commission 2009] Canadian Radio-television and Telecommunications Commission (2009). Review of the Internet traffic management practices of Internet service providers. http://www.crtc.gc.ca/eng/archive/2009/2009-657.htm. Acessado em 25/01/2017.

[Carbone and Rizzo 2010] Carbone, M. and Rizzo, L. (2010). Dummynet Revisited. *SIGCOMM Computer Communication Review*, 40(2).

[Cellan-Jones 2009] Cellan-Jones, R. (2009). BT accused of iPlayer throttling. http://news.bbc.co.uk/2/hi/technology/8077839.stm. Acessado em 25/01/2017.

[Cheng et al. 2004] Cheng, Y.-C., Hölzle, U., Cardwell, N., Savage, S., and Voelker, G. M. (2004). Monkey See, Monkey Do: A Tool for TCP Tracing and Replaying. In *USENIX Annual Technical Conference (ATEC)*.

[Coates et al. 2002] Coates, A., III, A. O. H., Nowak, R., and Yu, B. (2002). Internet tomography. *IEEE Signal Processing Magazine*, 19(3).

[Comisión de Regulación de Comunicaciones 2011] Comisión de Regulación de Comunicaciones (2011). Condiciones regulatorias relativas a la neutralidad en Internet. http://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=45061. Acessado em 25/01/2017.

[Comitê Gestor da Internet no Brasil 2009] Comitê Gestor da Internet no Brasil (2009). Princípios para a Governança e Uso da Internet no Brasil. http://www.cgi.br/resolucoes/documento/2009/003. Acessado em 25/01/2017.

[Cooper and Brown 2015] Cooper, A. and Brown, I. (2015). Net Neutrality: Discrimination, Competition, and Innovation in the UK and US. *ACM Transactions on Internet Technology*, 15(1).

[Crowcroft 2007] Crowcroft, J. (2007). Net Neutrality: The Technical Side of the Debate: a White Paper. *SIGCOMM Computer Communication Review*, 37(1).

[Detal et al. 2013] Detal, G., Hesmans, B., Bonaventure, O., Vanaubel, Y., and Donnet, B. (2013). Revealing Middlebox Interference with Tracebox. In *Internet Measurement Conference*, pages 1–8. ACM.

[Dhawan et al. 2012] Dhawan, M., Samuel, J., Teixeira, R., Kreibich, C., Allman, M., Weaver, N., and Paxson, V. (2012). Fathom: A Browser-based Network Measurement Platform. In *ACM Conference on Internet Measurement Conference (IMC)*. ACM.

[Dischinger et al. 2007] Dischinger, M., Haeberlen, A., Gummadi, K. P., and Saroiu, S. (2007). Characterizing Residential Broadband Networks. In *SIGCOMM Conference on Internet Measurement (IMC)*. ACM.

[Dischinger et al. 2010] Dischinger, M., Marcon, M., Guha, S., Gummadi, K. P., Mahajan, R., and Saroiu, S. (2010). Glasnost: Enabling End Users to Detect Traffic Differentiation. In *USENIX Conference on Networked Systems Design and Implementation (NSDI)*.

[Dischinger et al. 2008] Dischinger, M., Mislove, A., Haeberlen, A., and Gummadi, K. P. (2008). Detecting Bittorrent Blocking. In *SIGCOMM Conference on Internet Measurement*. ACM.

[Dovrolis et al. 2010] Dovrolis, C., Gummadi, K., Kuzmanovic, A., and Meinrath, S. D. (2010). Measurement Lab: Overview and an Invitation to the Research Community. *SIGCOMM Computer Communication Review*, 40(3).

[Dreier 2016] Dreier, T. (2016). Comcast Hit With FCC Complaint Over Net Neutrality Violations. `http://www.streamingmedia.com/Articles/News/Online-Video-News/Comcast-Hit-With-FCC-Complaint-Over-Net-Neutrality-Violations-109609.aspx`. Acessado em 25/01/2017.

[El Congreso de Colombia 2011] El Congreso de Colombia (2011). Plan Nacional de Desarrollo, 2010-2014. `http://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=43101`. Acessado em 25/01/2017.

[Electronic Frontier Foundation] Electronic Frontier Foundation. Switzerland Network Testing Tool. `https://www.eff.org/pages/switzerland-network-testing-tool`. Acessado em 25/01/2017.

[Esnaashari 2014] Esnaashari, S. (2014). Invisible Barriers: Identifying restrictions affecting New Zealanders' access to the Internet. Master's thesis, Victoria University of Wellington. `http://researcharchive.vuw.ac.nz/xmlui/bitstream/handle/10063/3263/thesis.pdf`.

[European Commission 2009] European Commission (2009). EU Telecoms Reform: 12 reforms to pave way for stronger consumer rights, an open internet, a single European telecoms market and high-speed internet connections for all citizens. `http://europa.eu/rapid/press-release_MEMO-09-513_en.htm`. Acessado em 25/01/2017.

[European Commission 2010] European Commission (2010). Digital Agenda: Commission launches consultation on net neutrality. `http://europa.eu/rapid/press-release_IP-10-860_en.htm`. Acessado em 25/01/2017.

[European Commission 2014] European Commission (2014). 2014 Report on Implementation of the EU regulatory framework for electronic communications. `https://ec.europa.eu/digital-single-market/en/news/2014-report-implementation-eu-regulatory-framework-electronic-communications`. Acessado em 25/01/2017.

[European Commission 2015] European Commission (2015). Commission welcomes agreement to end roaming charges and to guarantee an open Internet. `http://europa.eu/rapid/press-release_IP-15-5265_en.htm`. Acessado em 25/01/2017.

[European Parliament and Council of the European Union 2009] European Parliament and Council of the European Union (2009). Regulation 1211/2009 establishing the Body of European Regulators for Electronic Communications. `http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2009.337.01.0001.01.ENG`. Acessado em 25/01/2017.

[Federal Antimonopoly Service 2016] Federal Antimonopoly Service (2016). Creating equal conditions on the market of Internet services. http://en.fas.gov.ru/press-center/news/detail.html?id=44823. Acessado em 25/01/2017.

[Federal Communications Commission 2002] Federal Communications Commission (2002). FCC Classifies Cable Modem Service as "Information Service": Initiates Proceeding to Promote Broadband Deployment and Examine Regulatory Implications of Classification. http://transition.fcc.gov/Bureaus/Cable/News_Releases/2002/nrcb0201.html. Acessado em 25/01/2017.

[Federal Communications Commission 2005] Federal Communications Commission (2005). FCC 05-150. Report and order and notice of proposed rulemaking. https://apps.fcc.gov/edocs_public/attachmatch/FCC-05-150A1.pdf. Acessado em 25/01/2017.

[Federal Communications Commission 2010] Federal Communications Commission (2010). FCC 10-201. Report And Order. https://apps.fcc.gov/edocs_public/attachmatch/FCC-10-201A1_Rcd.pdf. Acessado em 25/01/2017.

[Federal Communications Commission 2014] Federal Communications Commission (2014). Public Notice.DA 14-211.New docket established to address open internet remand GN Docket No. 14-28. https://apps.fcc.gov/edocs_public/attachmatch/DA-14-211A1.pdf. Acessado em 25/01/2017.

[Federal Communications Commission 2015] Federal Communications Commission (2015). Open Internet. https://www.fcc.gov/general/open-internet. Acessado em 25/01/2017.

[Federal Communications Commission 2016] Federal Communications Commission (2016). Remarks of FCC Chairman Tom Wheeler. http://transition.fcc.gov/Daily_Releases/Daily_Business/2016/db0411/DOC-338806A1.pdf. Acessado em 25/01/2017.

[Filasto and Appelbaum 2012] Filasto, A. and Appelbaum, J. (2012). OONI: Open Observatory of Network Interference. In *USENIX Workshop on Free and Open Communications on the Internet*.

[Flach et al. 2016] Flach, T., Papageorge, P., Terzis, A., Pedrosa, L., Cheng, Y., Karim, T., Katz-Bassett, E., and Govindan, R. (2016). An Internet-Wide Analysis of Traffic Policing. In *ACM SIGCOMM*. ACM.

[Ganley and Allgrove 2006] Ganley, P. and Allgrove, B. (2006). Net neutrality: A user's guide. *Computer Law & Security Review*, 22(6).

[GreatFire.org] GreatFire.org. Expanding Online Freedom of Speech in China and Beyond. https://en.greatfire.org. Acessado em 25/01/2017.

[Guo and Easley 2016] Guo, H. and Easley, R. F. (2016). Network Neutrality Versus Paid Prioritization: Analyzing the Impact on Content Innovation. *Production and Operations Management*, 25(7):1261–1273.

[H. B. Mann 1947] H. B. Mann, D. R. W. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, 18(1).

[Habibi Gharakheili et al. 2016] Habibi Gharakheili, H., Vishwanath, A., and Sivaraman, V. (2016). Perspectives on Net Neutrality and Internet Fast-Lanes. *SIGCOMM Computer Communication Review*, 46(1):64–69.

[Hahn and Wallsten 2006] Hahn, R. W. and Wallsten, S. (2006). The Economics of Net Neutrality. *The Economists' Voice*, 3(6).

[Hourton et al. 2012] Hourton, G., Canto, G. D., Bustos, J., and Lalanne, F. (2012). Crowd-measuring: Assessing the quality of mobile Internet from end-terminals. In *International Conference on Network Games, Control and Optimization (NetGCooP)*, pages 145–148.

[Hwang 2007] Hwang, T. (2007). Threat Modeling: Herdict: A Distributed Model for Threats Online. *Network Security*, 2007(8).

[Internet Society] Internet Society. Net Neutrality. http://www.internetsociety.org/net-neutrality. Acessado em 25/01/2017.

[InternetNZ 2015] InternetNZ (2015). Network Neutrality. https://internetnz.nz/content/network-neutrality-discussion-document. Acessado em 25/01/2017.

[Joch 2009] Joch, A. (2009). Debating net neutrality. *Communications of the ACM*, 52(10):14–15.

[Jordan 2009a] Jordan, S. (2009a). Four questions that determine whether traffic management is reasonable. In *IFIP/IEEE International Symposium on Integrated Network Management*.

[Jordan 2009b] Jordan, S. (2009b). Some Traffic Management Practices Are Unreasonable. In *International Conference on Computer Communications and Networks (ICCCN)*.

[Jukic et al. 2011] Jukic, Z., Weber, M., Svedek, V., Vukovic, M., Katusic, D., and Jezic, G. (2011). Technical aspects of network neutrality. In *International Conference on Telecommunications (ConTEL)*.

[Kanuparthy and Dovrolis 2010] Kanuparthy, P. and Dovrolis, C. (2010). DiffProbe: Detecting ISP Service Discrimination. In *IEEE INFOCOM*.

[Kanuparthy and Dovrolis 2011] Kanuparthy, P. and Dovrolis, C. (2011). ShaperProbe: End-to-end Detection of ISP Traffic Shaping Using Active Methods. In *SIGCOMM Conference on Internet Measurement Conference (IM)*. ACM.

[Kendrick 2009] Kendrick, J. (2009). T-Mobile Germany Blocks iPhone Skype Over 3G and WiFi. https://gigaom.com/2009/04/06/t-mobile-germany-blocks-iphone-skype-over-3g-too. Acessado em 25/01/2017.

[K.G. Coffman 2002] K.G. Coffman, A. O. (2002). *Internet Growth: Is There a "Moore's Law" for Data Traffic?* Springer US.

[Kinzinger 2016] Kinzinger, A. (2016). H.R. 2666 - No Rate Regulation of Broadband Internet Access Act. http://www.gop.gov/bill/h-r-2666-no-rate-regulation-of-broadband-internet-access-act. Acessado em 25/01/2017.

[Knutson and Ramachandran 2016] Knutson, R. and Ramachandran, S. (2016). Netflix Throttles Its Videos on AT&T, Verizon Networks. http://www.wsj.com/articles/netflix-throttles-its-videos-on-at-t-verizon-phones-1458857424. Accessed in October 19, 2016.

[Korea Communications Commission 2012] Korea Communications Commission (2012). Annual Report 2011. http://eng.kcc.go.kr/download.do?fileSeq=35215. Acessado em 25/01/2017.

[Kreibich et al. 2010] Kreibich, C., Weaver, N., Nechaev, B., and Paxson, V. (2010). Netalyzr: Illuminating the Edge Network. In *SIGCOMM Conference on Internet Measurement (IMC)*. ACM.

[Krämer et al. 2013] Krämer, J., Wiewiorra, L., and Weinhardt, C. (2013). Net neutrality: A progress report. *Telecommunications Policy*, 37(9).

[Lalanne et al. 2015] Lalanne, F., Aguilera, N., Graves, A., and Bustos, J. (2015). Adkintun Mobile: Towards using personal and device context in assessing mobile QoS. In *International Wireless Communications and Mobile Computing Conference (IWCMC)*.

[Lee 2016] Lee, M. (2016). S.2602 - Restoring Internet Freedom Act. https://www.congress.gov/bill/114th-congress/senate-bill/2602/text. Acessado em 25/01/2017.

[Lessig 2001] Lessig, L. (2001). *The Future of Ideas: The Fate of the Commons in a Connected World*. The Future of Ideas: The Fate of the Commons in a Connected World. Random House.

[Ling et al. 2010] Ling, F.-Y., Tang, S.-L., Wu, M., Li, Y.-X., and Du, H.-Y. (2010). Research on the net neutrality: The case of Comcast blocking. In *International Conference on Advanced Computer Theory and Engineering (ICACTE)*, volume 5.

[Lomas 2016] Lomas, N. (2016). Verizon Accused Of Net Neutrality Foul By Zero-Rating Its Go90 Mobile Video Service. https://techcrunch.com/2016/02/07/verizon-accused-of-net-neutrality-foul-by-zero-rating-its-go90-mobile-video-service. Acessado em 25/01/2017.

[Lu et al. 2010] Lu, G., Chen, Y., Birrer, S., Bustamante, F. E., and Li, X. (2010). POPI: A User-Level Tool for Inferring Router Packet Forwarding Priority. *IEEE/ACM Transactions on Networking (TON)*, 18(1).

[Mahajan et al. 2008] Mahajan, R., Zhang, M., Poole, L., and Pai, V. (2008). Uncovering Performance Differences Among Backbone ISPs with Netdiff. In *USENIX Symposium on Networked Systems Design and Implementation (NSDI)*.

[Maille et al. 2016] Maille, P., Simon, G., and Tuffin, B. (2016). Toward a net neutrality debate that conforms to the 2010s. *IEEE Communications Magazine*, 54(3):94–99.

[Martin and Glorioso 2008] Martin, J. C. D. and Glorioso, A. (2008). The Neubot project: A collaborative approach to measuring internet neutrality. In *IEEE International Symposium on Technology and Society*.

[Michaut and Lepage 2005] Michaut, F. and Lepage, F. (2005). Application-oriented network metrology: metrics and active measurement tools. *IEEE Communications Surveys Tutorials*, 7(2).

[Ministry of Internal Affairs and Communications 2006] Ministry of Internal Affairs and Communications (2006). New Competition Promotion Program 2010. `http://www.soumu.go.jp/main_sosiki/joho_tsusin/eng/pdf/060928_1.pdf`. Acessado em 25/01/2017.

[Ministry of Internal Affairs and Communications 2007] Ministry of Internal Affairs and Communications (2007). Report on Network Neutrality. `http://www.soumu.go.jp/main_sosiki/joho_tsusin/eng/pdf/070900_1.pdf`. Acessado em 25/01/2017.

[Ministry of Internal Affairs and Communications 2008] Ministry of Internal Affairs and Communications (2008). Report from Panel on Neutrality of Networks. `http://www.soumu.go.jp/main_sosiki/joho_tsusin/eng/Releases/NewsLetter/Vol18/Vol18_23/Vol18_23.html`. Acessado em 25/01/2017.

[Ministério da Justiça] Ministério da Justiça. Marco Civil da Internet. `http://pensando.mj.gov.br/marcocivil/`. Acessado em 25/01/2017.

[Miorandi et al. 2013] Miorandi, D., Carreras, I., Gregori, E., Graham, I., and Stewart, J. (2013). Measuring net neutrality in mobile Internet: Towards a crowdsensing-based citizen observatory. In *IEEE International Conference on Communications Workshops (ICC)*, pages 199–203.

[Molavi Kakhki et al. 2015] Molavi Kakhki, A., Razaghpanah, A., Li, A., Koo, H., Golani, R., Choffnes, D., Gill, P., and Mislove, A. (2015). Identifying Traffic Differentiation in Mobile Networks. In *ACM Conference on Internet Measurement Conference*, pages 239–251. ACM.

[Mr T. 2013] Mr T. (2013). Zambia, a country under Deep Packet Inspection. `https://ooni.torproject.org/post/zambia`. Acessado em 25/01/2017.

[Mueller and Asghari 2012] Mueller, M. L. and Asghari, H. (2012). Deep Packet Inspection and Bandwidth Management: Battles over BitTorrent in Canada and the United States. *Telecommunications Policy*, 36(6).

[Net Neutrality Monitor] Net Neutrality Monitor. `http://www.neumon.org`. Acessado em 25/01/2017.

[Network Neutrality Squad] Network Neutrality Squad. NNSquad Network Measurement Agent (NNMA). `https://www.nnsquad.org/agent.html`. Acessado em 25/01/2017.

[Network of Excellence in InterNet Science] Network of Excellence in InterNet Science. MorFEO: MOnitoRing network connections to assess Freedom of Expression Online. `http://www.internet-science.eu/open-call-projects/morfeo`. Acessado em 25/01/2017.

[NIC.br] NIC.br. Sistema de Medição de Tráfego Internet (SIMET). `http://simet.nic.br`. Acessado em 25/01/2017.

[Noether 2012] Noether, G. E. (2012). *Introduction to statistics: the nonparametric way*. Springer Science & Business Media.

[Norwegian Communications Authority a] Norwegian Communications Authority. Net neutrality. http://eng.nkom.no/technical/internet/net-neutrality/net-neutrality. Acessado em 25/01/2017.

[Norwegian Communications Authority b] Norwegian Communications Authority. The Norwegian model. http://eng.nkom.no/technical/internet/net-neutrality/the-norwegian-model. Acessado em 25/01/2017.

[Ookla] Ookla. The world standard in Internet metrics. https://www.ookla.com. Acessado em 25/01/2017.

[O'Rielly 2016] O'Rielly, M. (2016). Shining the Spotlight: How FCC Rules Impact Consumers and Industries. https://apps.fcc.gov/edocs_public/attachmatch/DOC-338600A1.pdf. Acessado em 25/01/2017.

[Poder Executivo 2011] Poder Executivo (2011). PL 2126/2011. http://www.camara.gov.br/proposicoesWeb/fichadetramitacao?idProposicao=517255. Acessado em 25/01/2017.

[Presidência da República 2014] Presidência da República (2014). Lei 12965. http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm. Acessado em 25/01/2017.

[Presidência da República 2016] Presidência da República (2016). Decreto 8771. http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2016/Decreto/D8771.htm. Acessado em 25/01/2017.

[Press Trust of India 2015] Press Trust of India (2015). Net Neutrality debate: TRAI aims to resolve some issues by early 2016. http://indianexpress.com/article/technology/tech-news-technology/trai-aims-to-resolve-some-net-neutrality-issues-by-early-2016. Acessado em 25/01/2017.

[Public Knowledge 2016] Public Knowledge (2016). Petition for the Federal Communications Commission to Enforce Merger Conditions and its Policies. https://ecfsapi.fcc.gov/file/60001526808.pdf. Acessado em 25/01/2017.

[Qazi et al. 2013] Qazi, Z. A., Lee, J., Jin, T., Bellala, G., Arndt, M., and Noubir, G. (2013). Application-awareness in SDN. *SIGCOMM Computer Communication Review*, 43(4).

[Qiu et al. 2008] Qiu, T., Ni, J., Wang, H., Hua, N., Yang, Y. R., and Xu, J. J. (2008). Packet Doppler: Network Monitoring Using Packet Shift Detection. In *ACM CoNEXT Conference*. ACM.

[Ravaioli et al. 2012] Ravaioli, R., Barakat, C., and Urvoy-Keller, G. (2012). Chkdiff: Checking Traffic Differentiation at Internet Access. In *ACM Conference on CoNEXT Student Workshop*. ACM.

[Ravaioli et al. 2015] Ravaioli, R., Urvoy-Keller, G., and Barakat, C. (2015). Towards a General Solution for Detecting Traffic Differentiation at the Internet Access. In *Teletraffic Congress (ITC)*.

[reddit 2014] reddit (2014). I just doubled my PIA VPN throughput that I am getting on my router by switching from UDP:1194 to TCP:443. https://www.reddit.com/r/VPN/comments/1xkbca/i_just_doubled_my_pia_vpn_throughput_that_i_am. Acessado em 25/01/2017.

[Reis et al. 2008] Reis, C., Gribble, S. D., Kohno, T., and Weaver, N. C. (2008). Detecting In-flight Page Changes with Web Tripwires. In *USENIX Symposium on Networked Systems Design and Implementation (NDSI)*.

[Respect My Net] Respect My Net. Report cases of Net Neutrality violations. https://respectmynet.eu. Acessado em 25/01/2017.

[SamKnows] SamKnows. The global platform for internet measurement. https://www.samknows.com. Acessado em 25/01/2017.

[Sánchez et al. 2011] Sánchez, M. A., Otto, J. S., Bischof, Z. S., and Bustamante, F. E. (2011). Dasu - ISP Characterization from the Edge: A BitTorrent Implementation. *SIGCOMM Computer Communication Review*, 41(4).

[Sánchez et al. 2013] Sánchez, M. A., Otto, J. S., Bischof, Z. S., Choffnes, D. R., Bustamante, F. E., Krishnamurthy, B., and Willinger, W. (2013). Dasu: Pushing Experiments to the Internet's Edge. In *USENIX Conference on Networked Systems Design and Implementation*, pages 487–500. USENIX Association.

[Sander Greenland 1999] Sander Greenland, James M. Robins, J. P. (1999). Confounding and Collapsibility in Causal Inference. *Statistical Science*, 14(1).

[Sandvine] Sandvine. Intelligent Broadband Networks. `https://www.sandvine.com`. Acessado em 25/01/2017.

[Scott 2014] Scott, M. (2014). Tim Berners-Lee, Web Creator, Defends Net Neutrality. `http://bits.blogs.nytimes.com/2014/10/08/tim-berners-lee-web-creator-defends-net-neutrality`. Acessado em 25/01/2017.

[Secretaria de Comunicaciones y Transportes 2014] Secretaria de Comunicaciones y Transportes (2014). Ley Federal de Telecomunicaciones y Radiodifusión. `http://www.sct.gob.mx/fileadmin/Comunicaciones/LFTR.pdf`. Acessado em 25/01/2017.

[Serral-Gracia et al. 2009] Serral-Gracia, R., Labit, Y., Domingo-Pascual, J., and Owezarski, P. (2009). Towards an Efficient Service Level Agreement Assessment. In *IEEE INFOCOM*.

[Serral-Gracià et al. 2010] Serral-Gracià, R., Yannuzzi, M., Labit, Y., Owezarski, P., and Masip-Bruin, X. (2010). An efficient and lightweight method for Service Level Agreement assessment. *Computer Networks*, 54(17).

[Sfakianakis et al. 2011] Sfakianakis, A., Athanasopoulos, E., and Ioannidis, S. (2011). CensMon: A Web censorship monitor. In *USENIX Workshop on Free and Open Communications on the Internet*.

[Shankesi 2013] Shankesi, R. (2013). *Friendsourcing to detect network manipulation*. PhD thesis, University of Illinois. `https://www.ideals.illinois.edu/bitstream/handle/2142/45321/Ravinder_Shankesi.pdf`.

[Sommers et al. 2007] Sommers, J., Barford, P., Duffield, N., and Ron, A. (2007). Accurate and Efficient SLA Compliance Monitoring. *SIGCOMM Computer Communication Review*, 37(4).

[Sommers et al. 2010] Sommers, J., Barford, P., Duffield, N., and Ron, A. (2010). Multiobjective Monitoring for SLA Compliance. *IEEE/ACM Transactions on Networking (TON)*, 18(2).

[Subsecretaría de Telecomunicaciones 2010] Subsecretaría de Telecomunicaciones (2010). Consagra el Principio de Neutralidad en la Red para los Consumidores y Usuarios de Internet. `http://www.leychile.cl/Navegar?idNorma=1016570`. Acessado em 25/01/2017.

[Subsecretaría de Telecomunicaciones 2011] Subsecretaría de Telecomunicaciones (2011). SUBTEL instruye y exige a empresas de internet mayor transparencia en planes de banda ancha por Ley de Neutralidad de Red. `http://www.subtel.gob.cl/subtel-instruye-y-exige-a-empresas-de-internet-mayor-transparencia-en-planes-de-banda-ancha-por-ley-de-neutralidad-de-red/`. Acessado em 25/01/2017.

[Subsecretaría de Telecomunicaciones 2014] Subsecretaría de Telecomunicaciones (2014). Ley de Neutralidad y Redes Sociales Gratis. `http://www.subtel.gob.cl/ley-de-neutralidad-y-redes-sociales-gratis/`. Acessado em 25/01/2017.

[Sørensen 2014] Sørensen, F. (2014). Net neutrality and charging models. `http://eng.nkom.no/topical-issues/news/net-neutrality-and-charging-models`. Acessado em 25/01/2017.

[Ta and Mao 2006] Ta, X. and Mao, G. (2006). Online End-to-End Quality of Service Monitoring for Service Level Agreement Verification. In *IEEE International Conference on Networks*, volume 2.

[Tariq et al. 2009] Tariq, M. B., Motiwala, M., Feamster, N., and Ammar, M. (2009). Detecting Network Neutrality Violations with Causal Inference. In *International Conference on Emerging Networking Experiments and Technologies (CoNEXT)*. ACM.

[Telecom Regulatory Authority of India 2016] Telecom Regulatory Authority of India (2016). Prohibition of Discriminatory Tariffs for Data Services. `http://www.trai.gov.in/WriteReadData/WhatsNew/Documents/Regulation_Data_Service.pdf`. Acessado em 25/01/2017.

[TestMy.net] TestMy.net. Broadband Internet Speed Test. `http://testmy.net`. Acessado em 25/01/2017.

[Topolski 2007] Topolski, R. (2007). Comcast is using Sandvine to manage P2P Connections. `http://www.dslreports.com/forum/r18323368-Comcast-is-using-Sandvine-to-manage-P2P-Connections`. Acessado em 25/01/2017.

[Trestian et al. 2009] Trestian, I., Potharaju, R., and Kuzmanovic, A. (2009). Closing the Loop: Feedback at Your Fingertips. `http://www.cs.northwestern.edu/~ict992/docs/draft.pdf`.

[van Schewick 2016] van Schewick, B. (2016). T-Mobile's Binge On Video Streaming Program. `https://prodnet.www.neca.org/publicationsdocs/wwpdf/2216she.pdf`. Acessado em 25/01/2017.

[van Schewick and Farber 2009] van Schewick, B. and Farber, D. (2009). Point/Counterpoint: Network Neutrality Nuances. *Communications of the ACM*, 52(2).

[Vishwanath and Vahdat 2009] Vishwanath, K. V. and Vahdat, A. (2009). Swing: Realistic and Responsive Network Traffic Generation. *IEEE/ACM Transactions on Networking*, 17(3).

[Weaver et al. 2009] Weaver, N., Sommer, R., and Paxson, V. (2009). Detecting Forged TCP Reset Packets. In *Network and Distributed System Security Symposium (NDSS)*.

[Weber et al. 2013] Weber, M., Svedek, V., Jukic, Z., Golub, I., and Zuljevic, T. (2013). Can HAKOMetar be used to increase transparency in the context of network neutrality? In *International Conference on Telecommunications (ConTEL)*.

[Weinsberg et al. 2011] Weinsberg, U., Soule, A., and Massoulie, L. (2011). Inferring traffic shaping and policy parameters using end host measurements. In *IEEE INFOCOM*.

[Weitzner 2008] Weitzner, D. J. (2008). Net Neutrality... Seriously this Time. *IEEE Internet Computing*, 12(3):86–89.

[Wu 2002] Wu, T. (2002). A Proposal for Network Neutrality. `http://www.timwu.org/OriginalNNProposal.pdf`. Acessado em 25/01/2017.

[Wu and Lessig 2003] Wu, T. and Lessig, L. (2003). Ex Parte Submission in CS Docket No. 02-52. `http://www.savetheinternet.com/sites/default/files/resources/wu_lessig_fcc.pdf`. Acessado em 25/01/2017.

[Yuksel et al. 2010] Yuksel, M., Ramakrishnan, K. K., Kalyanaraman, S., Houle, J. D., and Sadhvani, R. (2010). Quantifying Overprovisioning vs. Class-of-Service: Informing the Net Neutrality Debate. In *International Conference on Computer Communications and Networks*, pages 1–8.

[Zhang et al. 2009] Zhang, Y., Mao, Z. M., and Zhang, M. (2009). Detecting Traffic Differentiation in Backbone ISPs with NetPolice. In *SIGCOMM Conference on Internet Measurement Conference*. ACM.

[Zhang et al. 2008] Zhang, Y., Morley, Z., and Zhang, M. M. (2008). Ascertaining the Reality of Network Neutrality Violation in Backbone ISPs. In *ACM Workshop on Hot Topics in Networks*.

[Zhang et al. 2014] Zhang, Z., Mara, O., and Argyraki, K. (2014). Network Neutrality Inference. *SIGCOMM Computer Communication Review*, 44(4).

Capítulo 5

Mecanismos de autenticação e autorização para nuvens computacionais: definição, classificação e análise de soluções

Charles Christian Miers (UDESC), Guilherme Piêgas Koslovski (UDESC), Marcos Antonio Simplicio Jr.(USP), Tereza Cristina Melo de Brito Carvalho(USP), Fernando Frota Redígolo(USP), Marco Antonio Torrez Rojas(USP), Glauber Cassiano Batista (UDESC)

Abstract

The use of public and private cloud computing solutions has been growing in the last years. Several organizations identify in private clouds a way to consolidate the computational resources scattered around their infrastructure. Albeit promising, this consolidation implies on managing the (potentially large) set of users and services that need to access these clouds. In this scenario, it is highly advisable to employ authentication and authorization mechanisms that allow several systems to connect, reducing management efforts. The goal of this work is to identify and analyze the main cloud computing authentication models, focusing on single sign-on authentication mechanisms, to propose a consolidated taxonomy. As a case study, we explain the integration of OpenID Connect with OpenStack to provide authentication and authorization services using the Google Identity Provider.

Resumo

O uso de soluções de computação em nuvem vem aumentando constantemente nos últimos anos, tanto de nuvens públicas como privadas. Diversas organizações encontram em nuvens privadas uma forma de consolidar seus recursos computacionais dispersos pela sua infraestrutura. Embora promissora, essa consolidação implica em gerenciar o considerável conjunto de usuários e serviços que necessitam acessar essas nuvens. Nesse cenário, é fortemente recomendável fazer-se uso de mecanismos de autenticação e autorização que se relacionem com diversos sistemas, reduzindo-se esforços de gerenciamento desses sistemas. O objetivo do presente trabalho é identificar e analisar os principais modelos de autenticação utilizados em nuvens computacionais, com ênfase nos mecanismos de autenticação de logon único, propondo uma taxonomia consolidada. Como estudo de caso, é explicada a integração do OpenID Connect com o OpenStack para fornecer serviços de autenticação e autorização por meio do provedor de identidades do Google.

5.1. Autenticação e autorização: conceitos básicos

O uso de soluções de computação em nuvem continua em constante crescimento por parte de instituições diversas, incluindo empresas, governo e universidades [1]. A adoção de nuvens computacionais pode ocorrer pela contratação de um serviço externo (nuvem pública), pela ação conjunta entre várias organizações (nuvem comunitária), ou pela criação de uma nuvem para uso próprio (nuvem privada), ou uma combinação desses modelos (nuvem híbrida) [2]. A criação de uma nuvem privada ocorre, em geral, com o objetivo de fazer um uso mais racional dos recursos computacionais disponíveis, que podem ser oriundos de projetos descontinuados ou encerrados. O acesso a esses recursos na nuvem é controlado e restrito a pessoas autorizadas, que utilizam os serviços disponíveis após passarem por um processo de autenticação. Embora necessário, um problema geral existente em processos de autenticação e autorização é que eles costumam aumentar a carga sobre os usuários, que idealmente devem usar credenciais distintas para a autenticação em serviços diferentes. Em sistemas baseados em senhas, por exemplo, a tarefa de memorização de um grande número de sequências complexas de caracteres pode ser vista como uma barreira intransponível para diversos usuários [3].

Mecanismos de autenticação de logon único (*Single Sign-On* (SSO)) surgiram como uma solução para minimizar este problema. O processo de autenticação em sistemas SSO é realizado, basicamente, pela sincronização de senhas, cache seguro de credenciais e métodos de autenticação baseados em Infraestrutura de Chave Pública (ICP) ou baseados em *token* [4]. Algumas instituições utilizam sistemas federados para prover acesso aos serviços e recursos, porém a maioria desses sistemas são do tipo cliente-servidor (*e.g.,* Kerberos e LDAP) e necessitam de um mecanismo centralizado de autenticação [5] . Felizmente, há uma considerável variedade de soluções para autenticação e autorização disponíveis que podem ser aplicadas a nuvens computacionais [6]. Assim, para verificar sua adequação ao contexto de nuvens computacionais, é importante considerar suas principais características e eventuais limitações. Por exemplo, uma solução de SSO que se destaca no cenário atual é o OpenID, utilizado por grandes corporações (*e.g.,* Google e Microsoft). O OpenID é um sistema de autenticação descentralizado e com foco no usuário, envolvendo três entidades distintas: usuário, provedor de serviço (*Service Provider* (SP)) e provedor de identidades (*Identity Provider* (IdP)). A interação entre essas entidades no OpenID é tal que, assumindo-se uma relação de confiança entre o SP e o IdP, este último fica responsável por endossar as solicitações de autenticação e autorização dos usuários, permitindo ou negando acesso aos serviços do SP. Essa forma de autenticação descentralizada permite que múltiplos IdPs sejam utilizados para acesso a um mesmo SP, garantindo ao usuário o direito de escolha de um IdP do seu interesse. O mecanismo de autenticação OpenID pode ainda ser integrado com o mecanismo de autorização OAuth-v2 (um padrão aberto para delegação de *tokens* de autorização), como ocorre na solução conhecida como OpenID Connect, para proporcionar uma solução de controle de acesso padronizada para diversos serviços. A ampla utilização do OpenID Connect na web levou também o OpenStack, uma solução aberta de nuvem, a integrar essa ferramenta de autenticação ao seu mecanismo interno de gerenciamento de identidades.

Com o objetivo de discutir os principais benefícios e limitações do uso de ferramentas de SSO em sistemas em nuvem, bem como identificar as potenciais alternativas ao

OpenID Connect, o presente trabalho analisa os principais modelos de autenticação existentes, suas taxonomias, e seu uso no contexto de nuvens computacionais. Dessa forma, este trabalho revisa os principais conceitos de autenticação e autorização, aprofundando o estudo em mecanismos de autenticação descentralizada, em especial mecanismos de SSO. Além disso, é apresentada uma análise e classificação dos principais mecanismos de autenticação voltados para nuvens computacionais. Por fim, é descrito um estudo de caso prático, tendo como base um experimento no qual é integrado um mecanismo de autenticação de logon único em uma nuvem privada baseada em OpenStack, na sua versão Mitaka, de Abril/2016. Nesse experimento é explicado o funcionamento do componente de gerenciamento de identidades do OpenStack, o Keystone, e como ele se relaciona com as demais entidades em um mecanismo de autenticação de logon único, neste caso o OpenID Connect. O experimento envolve, além da integração do OpenID com um IdP da Google, uma análise do fluxo de mensagens entre as entidades.

5.1.1. Credenciais Digitais

Credenciais digitais foram propostas com o objetivo de implementar objetos de identificação de forma digital, sem comprometer a privacidade da entidade [7, 8]. Basicamente, uma credencial pode ser definida como uma declaração, emitida e assinada digitalmente por um emissor, sobre algumas propriedades da entidade. A entidade, por sua vez, utiliza essa credencial junto a uma terceira parte para provar algumas ou todas essas propriedades [9, 10]. Dessa forma, a entidade pode determinar quando, como e quais informações serão compartilhadas para confirmar sua identidade, potencialmente revelando apenas as informações mínimas necessárias com o objetivo de se autenticar de forma (quase) anônima. Para que seja possível de fato prover privacidade aos usuários, entretanto, é necessário que o mecanismo de uso da credencial inclua mecanismos que dificultem seu rastreamento e sua associação com a identidade de seu dono.

Um exemplo, descrito em [8], é considerar que um estudante de uma universidade deseja comprovar de forma eletrônica o seu estado de matrícula ativa, para pagar meia entrada em um cinema. Para isso, o estudante deve ser capaz de obter uma credencial da universidade que ateste o seu vínculo estudantil, mas não precisa revelar outros atributos (*e.g.,* nome, data de nascimento). Dessa forma, a universidade não seria capaz de rastrear o uso da credencial por parte do estudante, garantindo-lhe uma maior privacidade. Isso é possível se a credencial emitida para o estudante em questão for indistinguível de qualquer credencial emitida a outro estudante.

5.1.2. Autenticação

Protocolos de autenticação são a base de segurança em diversos sistemas na Internet. No contexto de comunicações seguras, o processo de autenticação consiste em verificar se uma entidade qualquer está cadastrada no sistema e tem direito de acessá-lo [11]. Assim, distinguem-se os seguintes conceitos no processo de autenticação [12]:

- Entidade: Um elemento do ecossistema em questão, seja ele humano ou um sistema. No contexto de nuvens computacionais, entidades podem ser definidas como uma abstração de alto nível (*e.g.,* serviços, consumidores e provedores) ou de baixo nível (*e.g.,* máquinas virtuais e processos);

- Atributo: Uma característica ou parte de uma informação que pertence a uma entidade ou a qualifica;
- Identidade: Uma representação única de uma entidade física ou lógica; e
- Credencial: Um conjunto de atributos relacionados a uma entidade em específico.

Assim, durante o processo de autenticação, a entidade que deseja acessar o sistema deve fornecer credenciais que permitam confirmar a identidade por ela alegada [13]. As credenciais comumente utilizadas em sistemas de autenticação podem ser de quatro tipos principais:

- Algo que a entidade sabe, como senhas, *tokens*, ou informações secretas diversas;
- Algo que a entidade possui, como cartões, crachás, entre outros;
- Algo que a entidade é, o que se traduz por informações estáticas como impressões digitais, padrão de veias das mãos, retina, ou outras informações biométricas; e
- Alguma informação transiente sobre a entidade, como sua localização física (*e.g.,* em um terminal dentro da rede interna) ou virtual (*e.g.,*, em uma sub-rede).

Para um nível de segurança mais elevado, essas credenciais podem ser combinadas, levando a um mecanismo de autenticação multi-fator, *Multi Factor Authentication* (MFA). Contudo, algumas credenciais são mais adequadas a alguns tipos de entidades do que a outras; por exemplo, embora informações biométricas possam ser utilizadas por humanos, sistemas computacionais são mais comumente autenticados usando mecanismos que envolvem chaves criptográficas. Além disso, os diferentes tipos de credenciais exigem mecanismos e ferramentas distintas para serem verificados, o que potencialmente afeta a usabilidade e/ou custo computacional do sistema de autenticação Nesses casos, mecanismos de SSO ganham ainda mais importância, pois um único processo de autenticação permite o acesso a vários serviços, amortizando os custos associados à apresentação e verificação de diversas credenciais [14, 15].

5.1.3. Autorização

O processo de autorização acontece após a autenticação, e consiste em verificar quais são as permissões de acesso dos usuários legítimos, assegurando-lhes direitos adequados ao interagir com as funcionalidades e serviços do sistema. Desta forma, as operações que um usuário autenticado pode realizar sobre determinados recursos são definidos pelo processo de autorização [16, 17].

Existem na literatura vários modelos para o controle de acesso de usuários, incluindo modelos simples como *Access Control List*s (ACLs) e capacidades (*capabilities*), e também modelos de mais alto nível como *Role-Based Access Control* (RBAC) e *Attribute-Based Access Control* (ABAC). Com o objetivo de dar uma visão geral sobre o tema, essas diferentes abordagens são brevemente discutidas neste trabalho.

Uma ACL consiste em uma tabela em que são listados os direitos de acesso de cada usuário a um objeto em particular (*e.g.,* um arquivo ou um serviço). Assim, cada objeto tem a ele associado uma lista explicitando as entidades do sistema e seu conjunto correspondente de direitos de acesso sobre aquele objeto [13, 18]. Um exemplo é o controle de acesso tradicional usado em sistemas UNIX, em que cada objeto (arquivos ou pastas) tem a ele associado direitos de leitura, escrita e execução para três classes de usuários (proprietário, grupo e outros). Embora simples, ACLs apresentam limitações em termos de gerenciamento, uma vez que alterar todos os direitos de uma entidade aos diversos objetos do sistema se torna um trabalho um tanto dispendioso e complexo. Também é possível que existam conflitos entre as regras de uma mesma ACL, fazendo-se necessário um mecanismo adicional para tratar conflitos e aplicar uma regra válida [18].

Sistemas baseados em capacidades, por sua vez, adotam uma abordagem complementar às ACLs: nesse caso, são as entidades que têm a elas associadas uma lista explicitando os objetos do sistema e os direitos de acesso correspondentes [13, 18]. Como resultado, a alteração dos direitos de acesso de uma entidade é uma tarefa razoavelmente simples. Entretanto, em sistemas com muitos usuários, o gerenciamento dos direitos de acesso a um dado objeto torna-se bastante trabalhoso, pois a lista de cada entidade deve ser verificada e atualizada.

Dada a complementariedade entre ACLs e mecanismos de autorização baseados em capacidades, os dois métodos podem ser combinados para tirar proveito dos aspectos positivos de cada um. Essa combinação é muito utilizada em sistemas distribuídos, nos quais o usuário pode se autenticar uma única vez, utilizar sua lista de capacidades para acessar recursos e serviços de vários servidores, e então cada servidor utiliza suas ACLs para prover um segundo nível de controle, mais rígido [18].

Já o modelo de autorização RBAC envolve a definição de "papéis", abstrações que normalmente se referem a diferentes funções ou cargos dentro do sistema (*e.g.,* administrador ou usuário). Assim, cada papel recebe um conjunto de direitos de acesso sobre recursos e serviços com base nas suas necessidades [19, 20]. Cada entidade é então associada a um ou mais papéis, recebendo, indiretamente, os direitos de acesso correspondentes. A organização dos papéis é comumente feita de forma hierárquica, permitindo que os níveis mais altos herdem as permissões dos níveis mais baixos [19]. Atualmente, esse modelo é amplamente utilizado em diversas organizações, por facilitar a adequação do controle de acesso condizente com funções internas dos usuários.

O ABAC, por sua vez, é um modelo em que o acesso a recursos e serviços é determinado com base nos atributos das entidades do sistema [21, 22]. Assim, as regras de autorização no sistema consideram os atributos associados à entidade e ao objeto alvo, as operações solicitadas e, em alguns casos, as condições ambientais (*e.g.,* localização física ou período do dia) [23]. Dessa forma, o ABAC pode ser visto como um modelo mais completo de autorização, permitindo a construção de políticas que levam em conta múltiplos fatores permanentes ou transitórios das entidades envolvidas em uma interação qualquer.

Apesar dessa diversidade de modelos de autorização, no contexto específico de nuvens computacionais o RBAC é provavelmente o modelo mais adotado, principalmente devido a sua estrutura hierárquica [22]. O sistemas de nuvem aberto OpenStack, por

exemplo, utiliza o modelo RBAC para controlar o acesso aos mais variados serviços, com base nos papéis dos usuários e projetos aos quais estão associados [15].

5.1.4. Computação em nuvem

De acordo com o *National Instutite of Standards and Technology* (NIST) a computação em nuvem é um modelo para habilitar o acesso ubíquo, conveniente e sob demanda, por rede, a um conjunto compartilhado de recursos de computação (*e.g.*, redes, servidores, armazenamento, aplicações e serviços) que possam ser rapidamente provisionados e liberados com o mínimo de esforço de gerenciamento ou interação com o provedor de serviços. Dependendo do modo como os recursos são utilizados e gerenciados pelo provedor de nuvem para fornecer tais características, podem-se definir os seguintes modelos de serviço [2]:

- *Software as a Service* **(SaaS)**: O recurso fornecido ao consumidor é o uso de aplicações do provedor executando em uma infraestrutura na nuvem. As aplicações podem ser acessadas por um navegador *web* ou via *Application Programming Interface*s (APIs). O consumidor não gerencia nem controla a infraestrutura na nuvem subjacente (*e.g.*, rede, armazenamento, sistema operacional);
- *Platform as a Service* **(PaaS)**: O recurso fornecido ao consumidor é a instalação de aplicações na infraestrutura da nuvem. Tais aplicações podem ser criadas ou adquiridos pelo consumidor, contando que sejam suportados pelo provedor de nuvem e compatíveis com seu ambiente computacional. O consumidor não gerencia nem controla a infraestrutura na nuvem subjacente, mas tem controle sobre as aplicações instaladas e algumas configurações do ambiente de hospedagem de aplicações; e
- *Infrastructure as a Service* **(IaaS)**: O recurso fornecido ao consumidor é provisionar processamento, armazenamento, comunicação de rede e outros recursos de computação fundamentais. O consumidor não gerencia nem controla a infraestrutura na nuvem subjacente mas tem controle sobre os sistemas operacionais, armazenamento, e aplicativos instalados, e possivelmente um controle limitado sobre alguns componentes de rede (*e.g.*, *firewalls*).

Adicionalmente, dependendo do tipo de instalação e da forma como os usuários são atendidos, definem-se diferentes modelos de implantação para nuvens computacionais [2]:

- **Nuvem Privada**: A infraestrutura da nuvem é provisionada para uso de uma única organização composta por vários consumidores. A propriedade, o gerenciamento e a operação pode ser da própria organização, de terceiros, ou uma combinação de ambos;
- **Nuvem Comunitária**: A infraestrutura é provisionada para uso exclusivo de uma comunidade de consumidores de organizações com interesses em comum. A propriedade, o gerenciamento e a operação pode ser de uma ou mais organizações da comunidade, de terceiros ou uma combinação de ambos;

- **Nuvem Pública**: A infraestrutura na nuvem é provisionada para uso aberto ao público em geral. A propriedade, o gerenciamento e a operação pode ser de uma ou mais organizações e fica localizada nas instalações do provedor; e
- **Nuvem Híbrida**: A infraestrutura na nuvem é uma composição de duas ou mais infraestruturas de nuvem (Privada, Comunitária ou Pública) que permanecem com entidades distintas, mas conectadas por uma tecnologia padronizada que permite a portabilidade e a comunicação de dados.

A computação em nuvem suporta a arquitetura orientada a serviços distribuídos e infraestrutura administrativa multiusuário e multidomínio [24]. Dessa forma, ela torna-se propensa a ameaças de segurança e vulnerabilidades associadas a tais cenários, nos quais as diferentes entidades que interagem como o sistema não são mutuamente confiáveis. Essa preocupação ganha importância prática principalmente ao considerar que o uso de serviços baseados na computação em nuvem tem crescido substancialmente nos últimos anos. Assim, os usuários utilizam os serviços de nuvem nas mais diversas modalidades de serviço, incluindo armazenamento e processamento de dados, devem ser protegidos contra o acesso e o uso indevido dos seus dados e quebras de privacidade [15, 25]. Para isso, a utilização de mecanismos de autenticação e autorização adequados é fundamental.

Motivados por essa preocupação, algumas nuvens computacionais já empregam mecanismos SSO para autenticar seus usuários [15, 26, 27]. Tal abordagem evita eventuais problemas advindos do uso de sistemas de autenticação distintos, com credenciais potencialmente incompatíveis. Por outro lado, ao mesmo tempo que a maior integração provida por mecanismos de SSO facilita o gerenciamento de identidades em nuvens computacionais, ela também pode acarretar em outros problemas de segurança e privacidade [5, 28]. Para entender como evitar tais armadilhas, faz-se necessário analisar mais profundamente os mecanismos de SSO disponíveis para uso em nuvens computacionais.

5.2. Autenticação Única

Conforme discutido na Subseção 5.1.2, o conceito de autenticação é baseada na verificação de credenciais que provam a associação de uma entidade a uma identidade [13]. Em contraste a sistemas tradicionais, nos quais tal verificação é necessária para cada serviço acessado, mecanismos de SSO fornecem um identificador único aos usuários para que eles possam se autenticar em todos os serviços que o suportam; desta forma, a autenticação é centralizada em um IdP [14, 15].

Basicamente, um mecanismo SSO é composto por três elementos:

- *Identity Provider* **(IdP)**: O Provedor de Identidades é o servidor de autenticação responsável pela emissão de identidades digitais de uma maneira segura, sempre que necessário;
- **SP**: O Provedor de Serviço é a aplicação que requer a autenticação das entidades que desejam acessá-lo (sujeitos), antes de prestar o serviço; e
- **Subject**: O Sujeito é a entidade (*e.g.*, usuário final ou aplicação) que deseja acessar o serviço, podendo fazê-lo depois de ser autenticado com a ajuda do IdP.

A Figura 5.1 ilustra a arquitetura básica das soluções SSO e as interações entre essas entidades. O processo de autenticação SSO é iniciado com o sujeito acessando o SP e informando qual IdP deve ser usado para a autenticação (*e.g.*, selecionando-o a partir de uma lista de IdPs suportados) (1). Em seguida, o agente de autenticação incorporado na aplicação do sujeito obtém as credenciais do IdP (2), no qual as credenciais do sujeito são armazenadas. Este IdP então executa o processo de autenticação junto ao SP, concedendo acesso ao sujeito para os serviços que está autorizado a utilizar (3). Do ponto de vista do SP, há uma pequena diferença entre este processo e a autenticação tradicional: embora a aplicação ainda receba informações referentes às credenciais do sujeito, no caso da autenticação via SSO a entidade responsável por fornecer essas credenciais é o IdP, não o próprio sujeito [29].

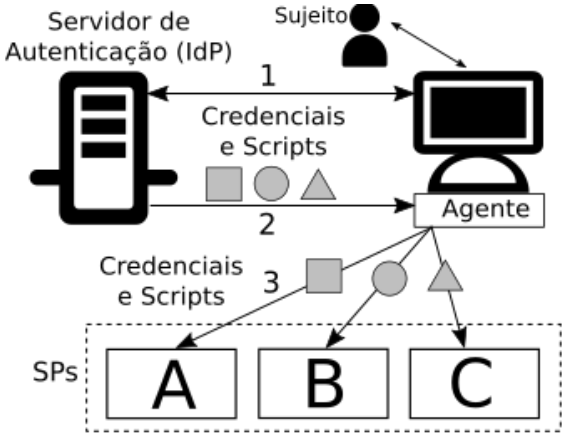


Figura 5.1: Sistema SSO. Adaptado de [29].

A autenticação executada desta maneira é dita "centralizada em um IdP", pois uma única execução do processo de autenticação junto a esta entidade concede acesso a vários SPs, assim, há variados recursos de computação e aplicações em um ambiente distribuído. Em outras palavras, depois que o IdP verifica que a reivindicação do usuário sobre uma identidade é válida, o IdP pode atuar em favor do usuário sempre que algum processo de autenticação for necessário. Adicionalmente, sempre que for necessário atualizar credenciais e dados pessoais (*e.g.*, um endereço residencial), essa operação pode ser realizada diretamente no IdP, que se torna responsável por sincronizar essa alteração junto a todos os serviços. Em sistemas que adotam senhas como parte do seu mecanismo de autenticação, por exemplo, o usuário não precisa lembrar de diversas senhas para diferentes serviços, nem mesmo ignorar as melhores práticas de segurança que desaconselham a reutilização de senhas: uma única senha, cadastrada junto ao IdP, é suficiente. O IdP pode então proteger as senhas de forma adequada: por exemplo, ao usar um esquema de *hashing* de senhas [3], dificultam-se ataques de dicionário mesmo no caso de seu banco de dados ser invadido/roubado; já com o uso de *captchas* e/ou mecanismos de bloqueio (temporário ou permanente), pode-se frustrar tentativas de teste exaustivo de senhas *online* [30]. O IdP pode ainda considerar a senha do usuário como uma "senha mestre", utilizando-a para derivar uma chave criptográfica para cifrar dados diversos do usuário, fornecendo proteção adicional contra vazamento de informações pessoais. De modo similar, em sistemas baseados em biometria, o IdP é a única entidade que coleta e armazena

a informação biométrica do usuário, devendo protegê-la adequadamente para reduzir o risco de exposição e eventuais ataques, como a replicação dessas informações [31]. Tais medidas são especialmente importantes ao se considerar que, como o IdP é responsável por armazenar as credenciais dos usuários, ele torna-se um ponto central de falha do sistema: se o IdP for comprometido, todo o sistema pode ficar ameaçado.

Combinados, esses elementos criam um *Metassistema de Identidades*, *i.e.,* uma arquitetura interoperável de identidades digitais. Esse metassistema permite aos usuários gerenciar várias identidades digitais usando múltiplas tecnologias subjacentes, possivelmente de implementações e provedores distintos [32]. De fato, soluções SSO têm evoluído constantemente e atualmente contam com diversos mecanismos para sincronização e armazenamento de credenciais. Exemplos de soluções e protocolos utilizam certificados (*e.g.*, X.509 [33]), *tokens* (*e.g.*, Kerberos [34]), e mecanismos de compartilhamento de atributos (*e.g.*, OpenID [35]). Embora essa pluralidade leve à possibilidade de se atender as mais diversas necessidades das aplicações, a escolha da solução de SSO mais adequada para um determinado cenário pode tornar-se bastante complexa sem o amparo de uma taxonomia que deixe claras as características a serem avaliadas. Para auxiliar nesta tarefa, a Seção 5.3 apresenta uma taxonomia consolidada com esse propósito.

5.3. Taxonomias para Mecanismos de Autenticação Única

Historicamente, a existência de um grande número de soluções de SSO resultou no aparecimento de diversas taxonomias na literatura [36–38], cada uma voltada a classificar esses protocolos em contextos específicos. Embora as taxonomias existentes não abordem as soluções visando especificamente nuvens computacionais, elas fornecem um panorama bastante completo dos métodos SSO que podem ser adotados também nesse cenário. Por esse motivo, algumas das principais taxonomias são analisadas nesta seção.

5.3.1. Taxonomia de Pashalidis e Mitchell

A taxonomia de Pashalidis [36] é a mais aceita para a classificação de sistemas SSO [37, 39–41]. Ela permite classificar sistemas SSO em duas categorias principais:

- **Pseudo-SSO**: utiliza um componente SSO para gerenciar as credenciais de autenticação para cada SP. O usuário autentica-se no componente, que é responsável por transmitir a operação aos diferentes SPs. O relacionamento entre identidade e SP é $n:1$ – o usuário pode ter múltiplas identidades em um mesmo SP e cada identidade está relacionada a um único SP.

- **True SSO**: Sistemas *True* SSO utilizam um *Authentication Service Provider* (ASP), um papel que é geralmente designado ao IdP, para estabelecer a relação com cada SP. Essa relação requer algum nível de confiança entre as entidades, que é assegurado por um acordo contratual [36]. O processo de autenticação ocorre exclusivamente entre o usuário e o ASP, e os SPs são notificados sobre o resultado por meio de asserções de autenticação. Nessas asserções estão contidos a identidade do usuário e o estado de autenticação no SP. O relacionamento entre identidade e SP pode ser $n:m$, já que o usuário pode utilizar várias identidades em um SP, e também pode utilizar uma identidade em vários SPs.

Sistemas Pseudo-SSO, assim como os sistemas *True* SSO, podem ser do tipo local ou baseado em *proxy*, o que resulta nas quatro categorias apresentadas na Tabela 5.1. Essa taxonomia apresenta uma classificação sucinta, porém bem definida dos sistemas SSO existentes, abstraindo quatro modelos de gerenciamento de identidades em sistemas SSO. Dessa forma, essa taxonomia tornou-se, de certo modo, a base para a classificação de sistemas SSO, abordando os possíveis modelos existentes.

Tabela 5.1: Taxonomia SSO de Pashalidis e Mitchell [36]

	Local	Proxy
Pseudo-SSO	**Pseudo-SSO Local**: O componente Pseudo-SSO reside na máquina do usuário, atuando como um repositório de credenciais e *scripts* para autenticar o usuário. O usuário inicia o processo de autenticação no componente e então o processo é redirecionado aos SPs.	**Pseudo-SSO baseado em Proxy**: O componente Pseudo-SSO reside em um servidor *proxy* externo, o qual deve ser de confiança do usuário para armazenar as credenciais e scripts de autenticação. A autenticação é iniciada pelo usuário e é redirecionada ao *proxy*, que executa o processo de autenticação.
True SSO	**True SSO Local**: O usuário autentica-se no ASP local, que envia as asserções de autenticação aos SPs. Deve existir um nível adequado de confiança entre o ASP local e os SPs, bem como uma infraestrutura de segurança apropriada.	**True SSO baseado em Proxy**: O ASP reside em um servidor externo que opera como um intermediário entre os usuários e SPs, emitindo asserções a respeito da autenticação. O ASP deve ser confiável, uma vez que ele pode facilmente personificar o usuário enviando uma asserção a um SP.

5.3.2. Arquitetura de Clercq

A taxonomia apresentada por Clercq [37], ilustrada na Figura 5.2, classifica as arquiteturas SSO com base na sua complexidade, distinguindo-as em duas categorias: simples ou complexas.

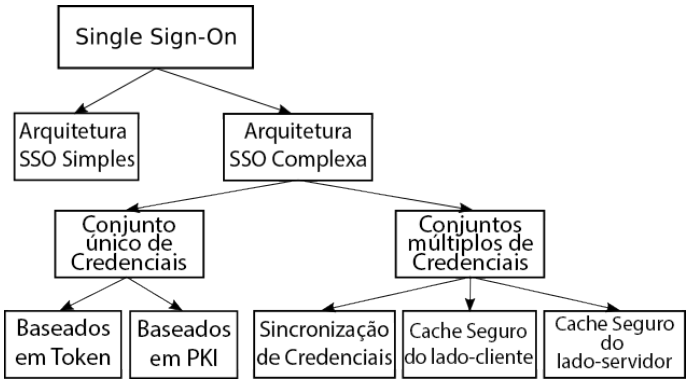


Figura 5.2: Arquitetura de Clercq [37].

Arquiteturas SSO simples usam uma única autoridade de autenticação com um mesmo conjunto de credenciais para o usuário. Alguns exemplos incluem o IBM Tivoli [42] e Microsoft RADIUS [43]. Ter uma autoridade única não implica necessariamente que existe somente um servidor de autenticação e um banco de credenciais disponível. De fato, para melhor desempenho e escalabilidade, uma única autoridade de

autenticação pode ser composta por múltiplos servidores de autenticação e vários bancos de credenciais [37].

Já arquiteturas SSO complexas envolvem múltiplas autoridades de autenticação com um ou vários conjuntos de credenciais para cada usuário, permitindo cobrir domínios administrativos distintos, implementados em diferentes plataformas (potencialmente gerenciados por várias organizações). As arquiteturas complexas são divididas em duas subclasses: sistemas que utilizam um único conjunto de credenciais e sistemas que empregam múltiplos conjuntos. Os primeiros geralmente contam com dois tipos de credenciais:

- **Baseados em *Token***: Os usuários são autenticados na autoridade de autenticação e recebem um *token*, que pode ser utilizado em autenticações subsequentes em serviços de mesmo domínio ou para assegurar a identidade do usuário em uma segunda autoridade (*e.g.*, Kerberos [34]); e
- **Baseados em PKI**: Usam uma *Public Key Infrastructure* (PKI), o que significa que o usuário inicialmente gera um par de chaves criptográficas assimétricas e obtém um certificado sobre sua chave pública. O processo de registro normalmente envolve uma *Registration Authority* (RA), que verifica o conjunto de credenciais do usuário, e uma *Certification Authority* (CA), que é responsável por emitir o certificado digitalmente assinado para a chave pública apresentada pelo usuário. A posse do certificado e da chave privada correspondente permite ao usuário gerar *tokens* assinados para autenticação (*e.g.*, em um protocolo do tipo desafio-resposta).

Arquiteturas com múltiplos conjuntos de credenciais, por sua vez, podem ser subdivididas em três tipos:

- **Sincronização de credenciais**: As credenciais dos usuários são sincronizadas entre várias aplicações. Dessa forma, todas elas têm acesso ao mesmo conjunto de credenciais (*e.g.*, mesma senha). Para facilitar o processo de sincronização, em sistemas desse tipo as credenciais são geralmente armazenadas em um único banco de dados, acessível por todas as aplicações.
- **Cache seguro no lado-cliente**: Os usuários podem armazenar as credenciais localmente, as quais são protegidas por uma senha mestre; e
- **Cache seguro no lado-servidor**: Os usuários armazenam suas credenciais em um servidor remoto, que se torna o banco de credenciais primário. Esse banco mapeia credenciais primárias do usuário a credenciais ditas secundárias, que podem ser armazenadas pelos serviços correspondentes.

É interessante notar que, mesmo que a taxonomia de Clercq [37] inclua mais classes que a taxonomia de Pashalidis [36], elas utilizam conceitos similares em suas classificações. Por exemplo, as classes de cache seguro no lado-cliente e no lado-servidor, da arquitetura de Clercq [37], são respectivamente análogas às classes Pseudo-SSO e *True* SSO baseados em *Proxy* da taxonomia de Pashalidis [36].

5.3.3. Taxonomia de Bhargav-Spantzel

A taxonomia de Bhargav-Spantzel [38] aborda a solução SSO do ponto de vista do usuário. Por esse motivo, ela é bem aceita para classificar sistemas SSO no contexto de soluções de gerenciamento de identidades centrados no usuário [44–47]. Basicamente, a taxonomia propõe a classificação baseada em dois paradigmas principais: focado no relacionamento e focado na credencial.

5.3.3.1. Focado no Relacionamento

Um sistema SSO focado no relacionamento gerencia somente a relação entre IdP, SP e o usuário. O usuário consulta um IdP (com o qual tenha uma relação de confiança) para cada transação e obtém as informações dinamicamente. Sistemas focados no relacionamento utilizam *tokens* de curta duração (*short-term tokens*), que são válidos somente durante um número limitado de transações. Os potenciais benefícios e problemas de tais sistemas são:

- **Benefícios:**
  - *Tokens* de curta duração limitam o risco e danos potenciais, caso sejam roubados ou interceptados;
  - Enquanto o IdP estiver *online*, as informações de autenticação estarão atualizadas;
  - Em geral, são sistemas leves e não necessitam de uma aplicação robusta no lado-cliente; e
  - Somente necessitam de uma solução efetiva de criptografia para autenticar tokens.
- **Problemas:**
  - A necessidade de ter um IdP sempre *online* o transforma em um ponto central de falhas, ou *Single Point of Failure* (SPF);
  - Como o IdP sempre está envolvido nas transações, todas as atividades do usuário podem ser monitoradas por ele, logo é necessário confiar no IdP; e
  - Se existir transitividade de *tokens*, o risco de personificação aumenta. Um *token* é considerado transitivo se o componente que o recebe pode, de algum modo, usá-lo se passando pelo seu proprietário. Muitos sistemas SSO incluem mecanismos para prevenir esse problema; *e.g.,* o SAML [48] e o Liberty [49] definem explicitamente qual é a entidade para qual o *token* é destinado. Assim, qualquer tentativa de encaminhar o *token* a uma terceira parte levaria a sua rejeição.

5.3.3.2. Focado na Credencial

Um sistema focado na credencial é caracterizado por gerenciar diretamente as credenciais. Isso inclui, por exemplo, uma aplicação cliente que armazena *tokens* de longa duração

(*long term token*) em um banco de dados local: neste caso, o usuário pode reutilizar as credenciais emitidas pelo IdP para múltiplas transações, mesmo sem contatar o IdP novamente. Em termos práticos, tais sistemas utilizam credenciais com validade longa (*e.g.*, certificados X.509 [33]), pois caso contrário as credenciais armazenadas se tornariam inúteis rapidamente. Os benefícios e problemas de soluções focadas na credencial são:

- **Benefícios:**
  - Por definição, o IdP não está envolvido nas transações. Assim, não é possível monitorar as atividades dos usuários e, se o IdP ficar *offline*, a disponibilidade do sistema de autenticação não é afetada de forma crítica (exceto que novos *tokens* não poderão ser emitidos); e
  - Os *tokens* gerados para credenciais de longa duração são necessariamente não-transitivos, pois caso contrário os usuários estariam vulneráveis a ataques de personificação. A não-transitividade é assegurada por mecanismos como desafio-resposta empregando *nonces*, um número ou *string* arbitrária que só pode ser utilizada uma única vez. Dessa forma, um *token* gerado para um certo contexto não pode ser reutilizado em outro.
- **Problemas:**
  - O roubo ou o compartilhamento indesejado de credenciais podem acarretar em riscos e danos graves ao sistema;
  - A perda de credenciais requer um mecanismo para revogá-las. Se uma credencial é revogada, o usuário não pode acessar qualquer serviço associado a ela enquanto uma nova credencial não for emitida; e
  - Sistemas focados na credencial geralmente levam a uma carga alta de trabalho na aplicação cliente, exigindo aplicações mais robustas e recursos computacionais suficientes no lado-cliente.

5.3.3.3. Limitações

Mesmo que a taxonomia de Bhargav-Spantzel leve a uma abstração sucinta e ampla para mecanismos SSO, o fato de se voltar a sistemas centrados no usuário faz com que ela seja menos útil ao analisar arquiteturas centradas na identidade. Mais precisamente, soluções centradas no usuário permitem que o usuário utilize diversos provedores de identidades; dessa forma, o usuário (entidade) pode utilizar várias identidades para se autenticar nos mais diversos serviços, como ocorre comumente na Internet (*e.g.*, Google com OpenID Connect e Facebook Connect). Em contraste, soluções centradas na identidade utilizam um único provedor de identidades, logicamente centralizado, que gerencia as identidades de usuários (*e.g.*, Kerberos). Portanto, tal taxonomia não cobre de maneira adequada cenários em que é importante garantir que cada entidade seja inequivocamente identificada para propósitos administrativos (*e.g.*, um ambiente corporativo ou federado).

5.3.4. Comparação e Avaliação das Taxonomias Existentes

Um primeiro aspecto que pode ser considerado em uma comparação entre as três taxonomias apresentadas refere-se a cobertura de paradigmas centrados no usuário e/ou centrado na identidade. Desse ponto de vista, enquanto a taxonomia de Bhargav-Spantzel [38] aborda exclusivamente soluções centradas no usuário, as demais taxonomias abordam completamente o espectro de soluções SSO.

Outro aspecto interessante refere-se ao método ou às categorias empregadas na construção das taxonomias propostas para soluções SSO. A taxonomia de Pashalidis [36] apresenta a classificação dos sistemas principalmente em termos de localidade, distinguindo sistemas Pseudo-SSO e *True* SSO. Já a taxonomia de Bhargav-Spantzel [38] aborda as interações entre os usuários e o IdP, distinguindo sistemas nos quais as interações são fortes e os *tokens* são de curta duração (focado no relacionamento) daqueles com poucas interações e *tokens* de longa duração (focados na credencial).

Mesmo que não seja possível definir qual taxonomia é a mais qualificada para todos os cenários, a revisão de literatura mostra que a taxonomia de Pashalidis [36] é a mais abrangente e bem aceita. Dessa forma, é sensato usá-la como referência para a definição de sub-classes mais detalhadas. De fato, a relação entre a taxonomia de Pashalidis [36] e a arquitetura de Clercq [37] já foi descrita por [41]: como ilustrado na Tabela 5.2, a taxonomia de Pashalidis [36] (nas linhas e colunas) pode ser mapeada aproximadamente para as classes definidas por Clercq [37] (nas células da tabela). A principal limitação deste mapeamento é que não há uma categoria específica na taxonomia proposta para a sincronização de credenciais, abordada explicitamente por Pashalidis [36].

Tabela 5.2: Comparação entre as taxonomias: Pashalidis *vs*. Clercq. Adaptado de [41].

	Sistemas SSO locais	Sistemas SSO baseados em Proxy
Sistemas Pseudo-SSO	Cache Seguro do lado-cliente	Cache Seguro do lado-servidor
Sistemas *True* SSO	Sistemas SSO baseados em PKI	Sistemas SSO baseados em token

Tendo como base a relação entre as taxonomias de Pashalidis [36] e Clercq [37], desenvolvida por [41], pode-se elaborar a relação das mesmas com a taxonomia de Bhargav-Spantzel [38], conforme mostrado na Figura 5.3.

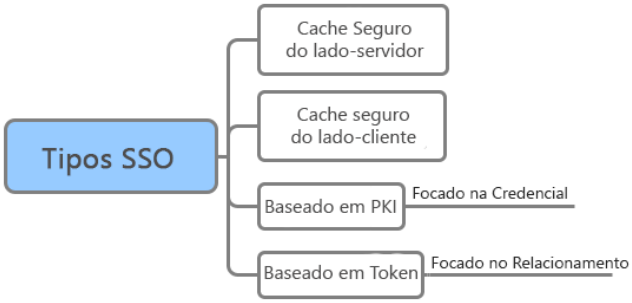


Figura 5.3: Relação entre as taxonomias analisadas neste trabalho.

É possível observar na Figura 5.3 que sistemas focados na credencial e sistemas focados no relacionamento podem ser vistos como subtipos de sistemas baseados em PKI e sistemas baseados em *tokens*, respectivamente. Dessa forma, a taxonomia de Pashalidis [36] (e, consequentemente, a de Clercq) podem ser vistas como mais abrangentes, incluindo categorias para a classificação de sistemas específicos como aqueles abordados em [38].

Embora uma taxonomia genérica permita que vários sistemas sejam classificados sob sua estrutura, a falta de classes com granularidade mais fina, disponíveis em taxonomias mais específicas, podem reduzir a sua utilidade quando uma análise mais detalhada é necessária. Afinal, sistemas com propriedades muito distintas podem se encontrar na mesma categoria (de alto nível), dificultando uma comparação direta dos detalhes não cobertos pela taxonomia. Por outro lado, uma abordagem de taxonomia centrada no usuário (*e.g.*, [38]) não necessariamente abrange sistemas SSO que não são centrados no usuário. No caso específico de nuvens, isso pode dificultar a classificação e comparação de diferentes modelos de autenticação, uma vez que eles devem considerar diversos aspectos, como [32]:

- Várias relações de confiança;
- Várias políticas de controle e acesso, baseadas em papéis e atributos;
- Provisionamento em tempo real;
- Autorização; e
- Auditoria e prestação de contas.

Ainda, uma particularidade do gerenciamento de identidades em nuvens é que nesse tipo de ambiente a identidade dos usuários está fragmentada em silos e geralmente não está inteiramente sob controle do usuário [50]. Tal especificidade motiva a criação de uma taxonomia para sistemas SSO orientado a nuvens que permita a identificação dos sistemas de autenticação que conciliam essas propriedades (*e.g.*, retornando o controle das identidades do sistema aos usuários).

5.3.5. Taxonomia Consolidada

Analisando as taxonomias descritas na Subseção 5.3.4, percebem-se as diferentes dimensões e ainda a ausência de outros aspectos/critérios de acordo com a fundamentação exposta nesta seção. Para superar tal limitação, é aqui elaborada uma proposta de taxonomia para consolidar as taxonomias identificadas e incluir os demais critérios relevantes, mostrada na Figura 5.4. A taxonomia proposta é baseada no modelo definido por Pashalidis [36] mas adota uma abordagem hierárquica em vez de horizontal para explicar as particularidades de sistemas em nuvem. Como resultado, esta abordagem permanece concisa nos níveis iniciais, enquanto os níveis mais profundos permitem a classificação de sistemas SSO específicos, mesmo sem se limitar a um cenário particular (*e.g., centrado no usuário*).

Figura 5.4: Taxonomia proposta para sistemas SSO em nuvens computacionais.

De forma sucinta, as categorias da taxonomia consolidada podem ser descritas como segue:

- **Pseudo-SSO**: Sistemas Pseudo-SSO podem ser classificados em: locais, baseados em *proxy* ou baseados na sincronização de senhas. Como notado na Subseção 5.3.4, essa inclusão explícita dos sistemas baseados na sincronização de senhas evita a necessidade de classificá-los como locais ou baseados em *proxy*, conciliando as taxonomias de [36] e de [37].

  - *Sincronização de Senhas*: Sistemas baseados na sincronização de senhas entre os serviços utilizados. Podem ser tipicamente representados por sistemas que usam um banco mestre para tal sincronização, assim implementando integração transparente entre qualquer outro banco de credenciais;
  - *Cache seguro do lado-cliente*: Sistemas que permitem gerenciamento local de credenciais. As credenciais são armazenadas em um banco local, protegido com uma senha mestre ou outro mecanismo de autenticação (*e.g.*, biometria); e
  - *Cache seguro no lado-servidor*: Sistemas que permitem armazenamento e gerenciamento remoto de credenciais. Essas credenciais são replicadas e protegidas por um servidor remoto de confiança, incluindo credenciais secundárias usadas para autenticar em outros domínios de serviço.

- **True SSO**: Sistemas *True* SSO são classificados em local e baseados em *proxy*. A taxonomia proposta explicitamente divide essas categorias, mas a análise das soluções existentes atualmente mostra que sistemas baseados em PKI são considerados locais e centrados no usuário, enquanto sistemas baseados em *tokens* são baseados em *proxy* e centrados na identidade.

  - *Centrados no usuário*: Sistemas que se concentram no usuário em vez dos SPs. Neste modelo, o usuário pode escolher o IdP que será responsável por processar o pedido de autenticação. Porém, isso não implica que vários IdPs devem ser utilizados. Isso deve-se ao fato de que o usuário tem maior controle sobre quais informações os SPs podem acessar e, dessa forma, a escolha de

IdPs é uma possibilidade. Sistemas centrados no usuário são baseados em asserções sobre o usuário, que são utilizadas para a comunicação entre os IdPs e SPs, ou empregam PKI para este propósito. Sistemas baseados em asserções são tipicamente focados no relacionamento, enquanto aqueles baseados em PKI são geralmente focados na credencial; e

– *Centrados na identidade*: São sistemas que se concentram na identidade do usuário no SP. Esses sistemas também são denominados "centralizados", pois os usuários estão limitados a autenticar em um IdP em particular e possuem pouco controle sobre as informações que o SP tem acesso.

Para avaliar a utilidade da taxonomia proposta, é interessante discutir como a mesma aborda a classificação dos sistemas SSO utilizados em ambientes de nuvens computacionais. Primeiramente, é enfatizado que soluções centradas no usuário, como OpenID [35] e SAML [48] são largamente adotadas e suportadas por serviços de nuvem, tanto em nuvens privadas, como o OpenStack [51], quanto em nuvens públicas, como a *Amazon Web Services* (AWS) [52]. Nesse contexto, como na taxonomia de Bhargav-Spantzel [38], é possível distinguir mecanismos centrados no usuário que são focados no relacionamento daqueles que são focados na credencial.

A taxonomia também permite a caracterização de sistemas SSO em diferentes níveis de especificidade, dependendo do grau de detalhes necessários. Por exemplo, em um nível mais alto, Kerberos e OpenID são ambos sistemas true SSO baseados em *proxy*. Contudo, em uma perspectiva de granularidade fina, o Kerberos pode também ser classificado como centrado na identidade, utilizando *tokens* para a comunicação entre os elementos do sistema. Em contraste, o OpenID é um sistema true SSO baseado em *proxy* e centrado no usuário, utilizando asserções (tipicamente transmitidas por *tokens*) para a comunicação entre as partes. Consequentemente, mesmo que o Kerberos e o OpenID sejam baseados em *tokens* e centralizam o processo de autenticação em uma entidade SSO, os métodos adotados podem ser distinguidos utilizando a taxonomia proposta.

Considerando que a taxonomia proposta seja abrangente, é importante enfatizar que ela pode não permanecer completa com o desenvolvimento de novos sistemas. Todavia, uma vez que novas abordagens ainda possam ser classificadas nas categorias de alto nível abrangidas, é possível expandir a taxonomia para acomodar tais soluções na hierarquia. Ainda, existem soluções SSO que combinam diferentes mecanismos em um mesmo domínio, conduzindo a um sistema mais complexo que não se encaixa diretamente em uma única categoria da taxonomia proposta. Por exemplo, uma nuvem OpenStack com a API de identidade na versão 3 [53] pode utilizar OpenID Connect e Kerberos para autenticar seus usuários. Consequentemente, o sistema de autenticação resultante é baseado em dois modelos distintos, podendo apenas ser classificado em uma categoria de nível mais alto (*e.g.*, sendo classificado genericamente como um sistema true SSO baseado em *proxy*).

Para uma visualização mais completa do ecossistema de autenticação em ambientes de nuvens computacionais, a Tabela 5.3 ilustra a classificação de vários sistemas SSO utilizando a taxonomia proposta. Nesta tabela, sistemas SSO são classificados, inicialmente, utilizando um nível de granularidade grossa. Posteriormente, são classificados em

um nível com granularidade mais fina. É possível observar que vários sistemas SSO utilizados em serviços e soluções de nuvem são abrangidos, mesmo aqueles desenvolvidos para contextos específicos. Portanto, são os níveis de granularidade fina que auxiliam na comparação de soluções distintas, a fim de encontrar a solução mais adequada para um serviço de nuvem e contexto.

Tabela 5.3: Classificação de soluções SSO existentes utilizando a taxonomia proposta

Solução	Categoria Principal	Subcategorias
Azure Active Directory [54]	Pseudo-SSO	Sincronização de credencial
CardSpace [55]	Pseudo-SSO	Cache seguro do lado-cliente
Entrust Cloud PKI [56]	Pseudo-SSO	Cache seguro do lado-servidor
Facebook Connect [57]	True SSO	Centrado no usuário; Asserções; Focado no relacionamento
FIDO [58]	Pseudo-SSO	Cache seguro do lado-cliente
Kerberos-based [34]	True SSO	Identity-centric; Tokens
Liberty Alliance [49]	True SSO	Centrado no usuário; Asserções; Focado no relacionamento
Microsoft Passport [59]	True SSO	Centrado no usuário
OpenID Connect [60]	True SSO	Centrado no usuário; Asserções; Focado no relacionamento
SAML [48]	True SSO	Centrado no usuário; Asserções; Focado no relacionamento
Shibboleth [61]	True SSO	Centrado no usuário; Asserções; Focado no relacionamento
X.509 [33]	True SSO	Centrado no usuário; PKI-based; Focado na credencial

Observando a Tabela 5.3, pode-se constatar que o OpenID Connect e o Facebook Connect são classificados da mesma forma. Vale ressaltar que o Facebook Connect, embora utilize apenas um IdP, é centrado no usuário e não na identidade. Como definido anteriormente, soluções centradas na identidade também são chamadas de centralizadas, uma vez que possuem somente um IdP. Contudo, também foi discutido que soluções centradas no usuário permitem maior controle do usuário sobre suas identidades e não necessariamente fazem uso de múltiplos IdPs. Dessa forma, mesmo que o Facebook Connect utilize somente um IdP, o fato de permitir que os usuários controlem o acesso às suas informações é o que determina sua classificação.

5.4. OpenID e OAuth

Conforme brevemente discutido na Seção 5.1, o OpenID é uma solução de autenticação amplamente utilizada atualmente. Tal aceitação advém de sua grande flexibilidade e facilidade de implementação, que permitem sua integração com outras técnicas para gestão de identidades. De fato, o OpenID Connect é um bom exemplo de integração do OpenID com os mecanismos de delegação de autorização providos pelo OAuth 2. Para uma melhor compreensão das funcionalidades dessas ferramentas, esta seção discute brevemente o OpenID e o OAuth 2 isoladamente, e então a forma como esses mecanismos são combinados no OpenID Connect.

5.4.1. OpenID

O OpenID[1] é um mecanismo de autenticação de código fonte aberto que provê gerenciamento de identidades centrado no usuário, de forma descentralizada, e que permite autenticação SSO [62–64]. Em um sistema usando OpenID, o usuário cadastra suas informações no IdP e utiliza as credenciais obtidas por meio de um agente para negociar a autenticação. Se as credenciais não forem verificadas corretamente, a autenticação falha [65].

Uma abordagem mais detalhada é descrita por [8, 62], no qual o protocolo OpenID oferece um método para provar que o usuário é o dono do identificador OpenID informado. A comprovação é feita através de mensagens trocadas entre o IdP e o SP, autenticadas por meio de um código de autenticação de mensagens. O processo de autenticação utilizando o OpenID envolve um *User-Agent* (UA), o qual é utilizado para a troca de mensagens com o SP e IdP. Na Figura 5.5 é possível observar o fluxo básico de autenticação do OpenID.

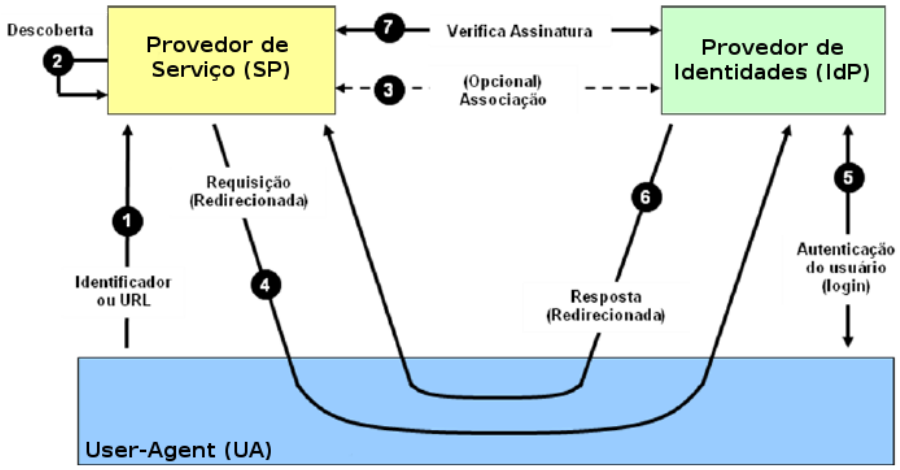


Figura 5.5: Fluxo básico de eventos no OpenID [8]

O processo de autenticação ilustrado na Figura 5.5 é composto pelas seguintes etapas [8]:

1. O usuário fornece ao SP seu OpenID ou a URL do IdP por meio do qual deseja se autenticar;
2. O SP inicia a descoberta da URL do IdP, para o qual é encaminhado o pedido de autenticação;
3. O SP pode estabelecer um segredo compartilhado com o IdP. É com base no segredo compartilhado que a troca de mensagens entre IdP e SP é autenticada;
4. O SP envia o pedido de autenticação ao IdP por meio de um redirecionamento HTTP, com vários parâmetros, realizado no UA;

[1] `http://openid.net/`

5. O IdP pode utilizar qualquer método de autenticação disponível para autenticar o usuário;

6. Em caso de sucesso, o IdP redireciona o UA para o SP, com a URL contendo uma confirmação de que a autenticação foi corretamente concluída; e

7. Por fim, o SP verifica a autenticidade da URL utilizando a chave estabelecida no passo 3 ou por meio de um pedido direto ao IdP.

No OpenID, o IdP precisa ser considerado confiável pelo usuário, pois o IdP é responsável pelo gerenciamento dos identificadores OpenID do usuário [8].

5.4.2. OAuth

O protocolo de autorização OAuth 2[2] provê um arcabouço genérico para permitir que um serviço interaja com outro para obter informações de um usuário, sob a anuência deste último [64, 66]. Por exemplo, um usuário poderia permitir que um serviço de impressão de fotos *A* acesse as suas fotos armazenadas em um serviço de rede social *B* sem que seja necessário fornecer a *A* as credenciais de autenticação de longa duração (e.g., senhas) cadastradas em *B*; assim, o acesso (potencialmente temporário) fica restrito às fotos autorizadas pelo usuário que é dono dos dados. Para que isso seja possível, o serviço *A*, denominado *Cliente* na arquitetura OAuth, deve primeiramente registrar-se junto a um *Servidor de Autorização*, estabelecendo uma chave compartilhada com este último e definindo um identificador único para essa relação de confiança entre eles. Quando o usuário dono dos recursos deseja conceder a *A* acesso aos seus dados armazenados no serviço *B*, denominado *Servidor de Recursos*, os seguintes passos devem ser realizados (veja Figura 5.6):

1. O usuário dono dos recursos, acessando a página do Cliente (no exemplo, `www.print-easy.com`), seleciona o Servidor de Recursos onde os dados estão armazenados (no exemplo, `www.mypics.com`);

2. O Cliente então redireciona o usuário para a página do Servidor de Recursos com uma requisição de autorização, na qual é informada o identificador do Cliente e o escopo (i.e., o tipo de recursos) que ele deseja acessar;

3. O usuário confirma a autorização de acesso aos recursos solicitados (ou uma parte deles), sendo então redirecionado para a página do Cliente com um *código de autorização*;

4. De forma transparente para o usuário, o Cliente apresenta o código de autorização ao Servidor de Autorização OAuth, autenticando-se com a chave previamente estabelecida. Em caso de sucesso, o Cliente recebe um *token de acesso*, o qual prova que o Cliente está autorizado a acessar o conjunto de dados concedido pelo usuário;

5. Finalmente, o Cliente apresenta o *token* de acesso ao Servidor de Recursos, que verifica a sua validade e permite acesso aos recursos protegidos em questão.

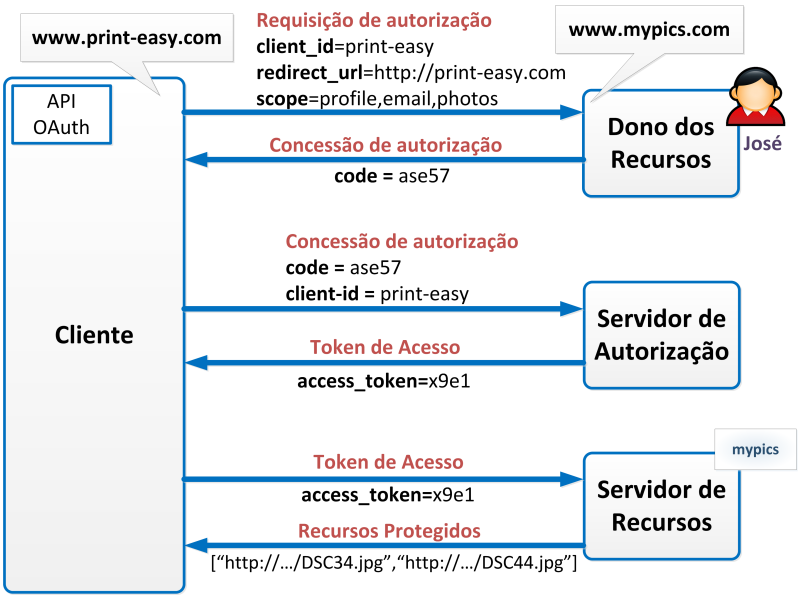


Figura 5.6: Uma visão geral do protocolo OAuth 2.0.

Protocolos como o OAuth possuem um papel relevante em nuvens computacionais (principalmente em nuvens públicas), assim como o OpenID. Esses protocolos fornecem métodos padronizados para permitir que os usuários controlem o número de contas e também quais serviços na nuvem podem acessar seus dados privados. Organizações como Facebook, Google, Microsoft, Yahoo, entre outros, utilizam o protocolo de autorização OAuth para permitir que os usuários tenham acesso aos serviços contratados, de maneira rápida e segura [66, 67]. Embora muitos aspectos de segurança de aplicações em nuvem ainda dependam dos desenvolvedores de serviços, o uso de mecanismos gerenciamento de identidades e acesso é um passo importante tomado por provedores de nuvem para garantir a segurança de seus ambientes [66].

5.4.3. OpenID Connect

O OpenID Connect é a terceira geração da tecnologia de autenticação OpenID. Essa versão consiste basicamente em um protocolo de gerenciamento de identidades usando o OAuth 2, aproveitando-se do fluxo de mensagens diretas REST/JSON e de *Transport Layer Security* (TLS) para garantir uma comunicação segura entre as partes [65]. Em outras palavras, o OpenID Connect permite que um serviço qualquer utilize o protocolo OAuth para ter acesso a um recurso específico do usuário: sua identidade OpenID. Assim, o OpenID Connect permite que usuários se autentiquem em múltiplas aplicações sem ter que gerenciar uma senha em cada uma delas e requer apenas requisições e respostas HTTP/HTTPS, não necessitando de *cookies* ou mecanismo similar para gerenciamento de sessão no SP.

Para explicar como o usuário interage com o OpenID Connect, um cenário fictício é utilizado. O usuário deseja acessar o serviço `exemplo.com`, neste cenário denominado SP, o qual é equivalente ao Cliente no contexto do OAuth 2. Para isso, ele deseja utili-

[2] `http://oauth.net/2/`

zar a sua identidade OpenID armazenada em um IdP capaz de autenticá-lo; desta forma, o IdP no OpenID Connect combina os papéis de Servidor de Autorização e Servidor de Recursos do OAuth 2. Assim, em vez de preencher o formulário de cadastro no SP, o usuário fornece um identificador (*e.g.,* uma URL) que representa sua identidade ou seleciona qual IdP será utilizado para a autenticação, caso o serviço ofereça uma lista com os IdPs compatíveis. A partir deste momento, o SP faz o processo de descoberta para verificar a propriedade do usuário sobre o identificador, através do documento de descoberta fornecido pelo IdP, que armazena informações para o processo de autenticação e autorização (*e.g.,* emissor, *endpoints* de autorização, escopos permitidos, tipos de resposta). Neste cenário, o SP identifica que o usuário pode se autenticar no IdP `openid.idp.com`, com o nome de usuário `user`. O SP então redireciona o usuário para esse IdP, requisitando acesso a sua identidade ao definir como parte do escopo da requisição o recurso "openid". O usuário deve então se autenticar no IdP utilizando as suas credenciais (*e.g.*, par de usuário e senha) e autorizar o acesso do SP às suas informações de identidade. Finalmente, o IdP redireciona o usuário através do UA (*e.g.*, navegador) para o SP, que fornece uma nova conta ao usuário ou o autentica em uma conta já existente e de sua propriedade. O fluxo descrito pode ser observado na Figura 5.7.

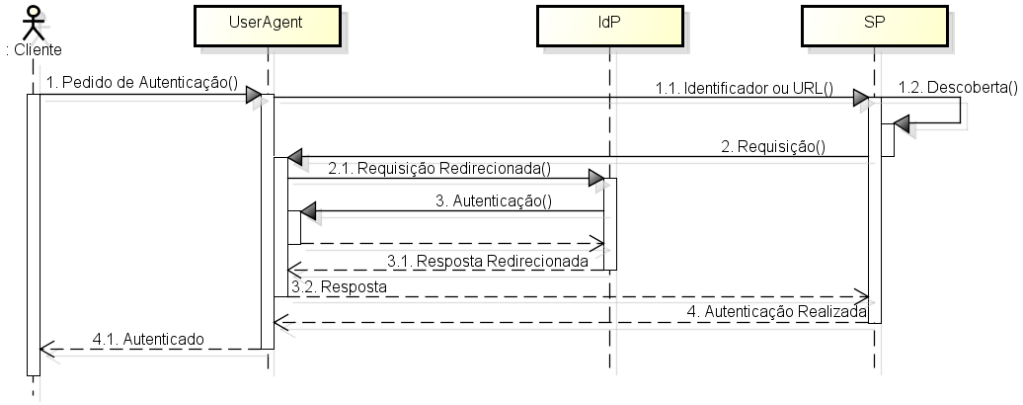


Figura 5.7: Processo de autenticação com OpenID Connect.

Essa integração dos processos de autenticação e autorização provida pelo OpenID Connect fornece maior segurança quando comparado ao uso dos protocolos OpenID e OAuth separadamente [63]. Esse fato, aliado a sua simplicidade e abertura, faz com que o OpenID Connect seja atualmente um dos principais mecanismos de SSO do mercado.

5.5. OpenStack

O OpenStack é um projeto de computação em nuvem criado em julho de 2010, oriundo de uma iniciativa entre RackSpace[3] e NASA[4]. Essa iniciativa tem por objetivo oferecer um serviço de computação em nuvem que possa ser executado em *hardware* de servidores padrão. Sua primeira versão oficial, o Austin, surgiu quatro meses depois. Com o intuito de realizar atualizações e correções em períodos curtos de tempo, a partir da quinta versão

[3]http://www.rackspace.com/
[4]http://www.nasa.gov/

do OpenStack ficou estabelecido o período de seis meses [51] entre o lançamento de cada nova versão.

O projeto OpenStack é uma coleção de componentes de código aberto utilizados por organizações para configurar e gerenciar nuvens computacionais do tipo IaaS. Esse projeto visa construir uma comunidade *open source* com pesquisadores, desenvolvedores e empresas que compartilhem um objetivo comum: criar uma nuvem simples de ser implementada, consideravelmente escalável e com vários recursos avançados [68, 69]. Assim, todo o código do OpenStack é aberto aos desenvolvedores, sendo que qualquer pessoa pode utilizar ou submeter modificações ao projeto, bem como construir aplicativos sobre a plataforma.

5.5.1. Componentes

Até a versão OpenStack Kilo, de abril/2015, o OpenStack possuía a organização dos seus componentes dividida em componentes oficiais e componentes incubados (em teste). Entretanto, novos componentes incubados eram incluídos a cada nova versão e componentes incubados considerados "estáveis" eram incorporados aos demais serviços, fazendo com que o número de componentes aumentasse a cada novo lançamento. Como vários desses novos componentes visavam atender a serviços bastante específicos, muitas vezes eles não eram adotados amplamente pela comunidade usuária. Por esta razão, o comitê gestor do OpenStack decidiu a partir da versão OpenStack Liberty organizar os componentes em dois grupos: *Core Services* e *Optional Services*.

Os componentes denotados *Core Services* são considerados essenciais, devendo estar presentes na instalação básica. São eles:

- Nova: serviços de computação. É o software que controla a infraestrutura de IaaS, alocando ou liberando recursos computacionais, como máquinas virtuais.
- Neutron: serviços de rede. Provisiona serviços de rede para os componentes que são gerenciados pelo Nova. Especificamente, permite que os usuários criem e anexem interfaces/comutadores de rede virtuais (vNICs/vSwitches) a esses componentes; e
- Cinder: serviço de blocos de armazenamento. Permite que volumes de armazenamento sejam criados, destruídos e vinculados a serviços virtualizados, como máquinas virtuais;
- Glance: serviço de armazenamento de imagens. Fornece um catálogo e repositório para imagens de máquinas virtuais, contêineres, *etc*. Essas imagens são manipuladas pelo Nova, responsável pelo gerenciamento das mesmas;
- Swift: serviço de armazenamento de objetos. Permite armazenar e recuperar arquivos que não estejam montados em um diretório. É um sistema de armazenamento a longo prazo para os dados permanentes ou estáticos; e
- Keystone: serviço de identidade. Fornece autenticação e autorização para todos os serviços do OpenStack.

Os componentes incluídos no *Optional Services* não são considerados essenciais, devendo ser instalados conforme as necessidades de cada provedor de nuvem. Nos serviços opcionais também encontram-se inseridos os antigos serviços incubados. Serviços disponíveis em Março/2017:

- Barbican: serviço de gerenciamento de chaves. Fornece uma API REST projetada para armazenamento, fornecimento e gerenciamento de chaves criptográficas, senhas e certificados X.509;
- Ceilometer: serviço de contabilidade. Monitora e mede o OpenStack para efetuar cobranças, medição de desempenho, escalabilidade e geração de estatísticas;
- Congress: serviço de governança. Promove *Policy as a Service* através de qualquer conjunto de serviços da nuvem para fornecer serviços de governança e conformidade para infraestruturas dinâmicas;
- Designate: serviço de *Domain Name Service* (DNS). Fornece serviço de *DNS as a Service*;
- Heat: serviço de orquestração. Orquestra a nuvem, permitindo o gerenciamento dos demais componentes por meio de uma interface única;
- Horizon: painel de controle/*dashboard*. Fornece uma interface web modular para todos os componentes do OpenStack, tornando possível seu gerenciamento direto a partir de um navegador web comum;
- Ironic: serviço de *bare-metal*. Propicia a utilização de imagens de máquina física (*bare-metal*) ao invés de *Virtual Machine*s (VMs);
- Magnum: serviço de contêineres. Disponibiliza um mecanismo de orquestração de contêineres para gerenciadores de contêineres como Docker e Kubernetes;
- Manila: serviço de compartilhamento de sistemas de arquivos. Fornece uma versão evoluída do Cinder com suporte a sistemas de arquivo distribuídos e recursos aprimorados de compartilhamento;
- Murano: serviço de catálogo de aplicações. Habilita que desenvolvedores e administradores da nuvem possam publicar diversas aplicações *cloud-ready* em um catálogo de aplicações navegável;
- Sahara: serviço elástico de *map reduce*. Fornece aos usuários um modo facilitado de provisionar cluster do tipo Hadoop através da especificação de um conjunto de parâmetros pré-definidos;
- Trove: serviço de banco de dados. Fornece serviço de banco de dados relacionais e não relacionais em um modelo *Database as a Service*; e
- Zaqar: Serviço de mensagens. Serviço de mensagens en nuvem *multi-tenant* para desenvolvedores web e *mobile*. Combina mecanismos do Amazon SQS com um suporte de semântica para eventos em *broadcast*.

Um ponto marcante nas diversas versões do OpenStack é a evolução das suas APIs. Cada componente possui a sua própria API em determinada versão, sendo que a evolução de uma versão comumente implica em incompatibilidade com versões anteriores. Deste modo, desenvolvedores necessitam constantemente atualizar os seus softwares caso desejem atualizar a versão do OpenStack.

Para dar uma visão geral da arquitetura do OpenStack, a Figura 5.8 ilustra como alguns dos componentes mais usuais do OpenStack estão relacionados. A Figura 5.8 também apresenta a hierarquia e conexões do OpenStack em alto nível, sendo que no nível de APIs existem muito mais conexões [51]. Ressalta-se aqui que, apesar do suporte a TLS nas APIs de alguns componentes, esse recurso de segurança não vem habilitado na sua configuração padrão.

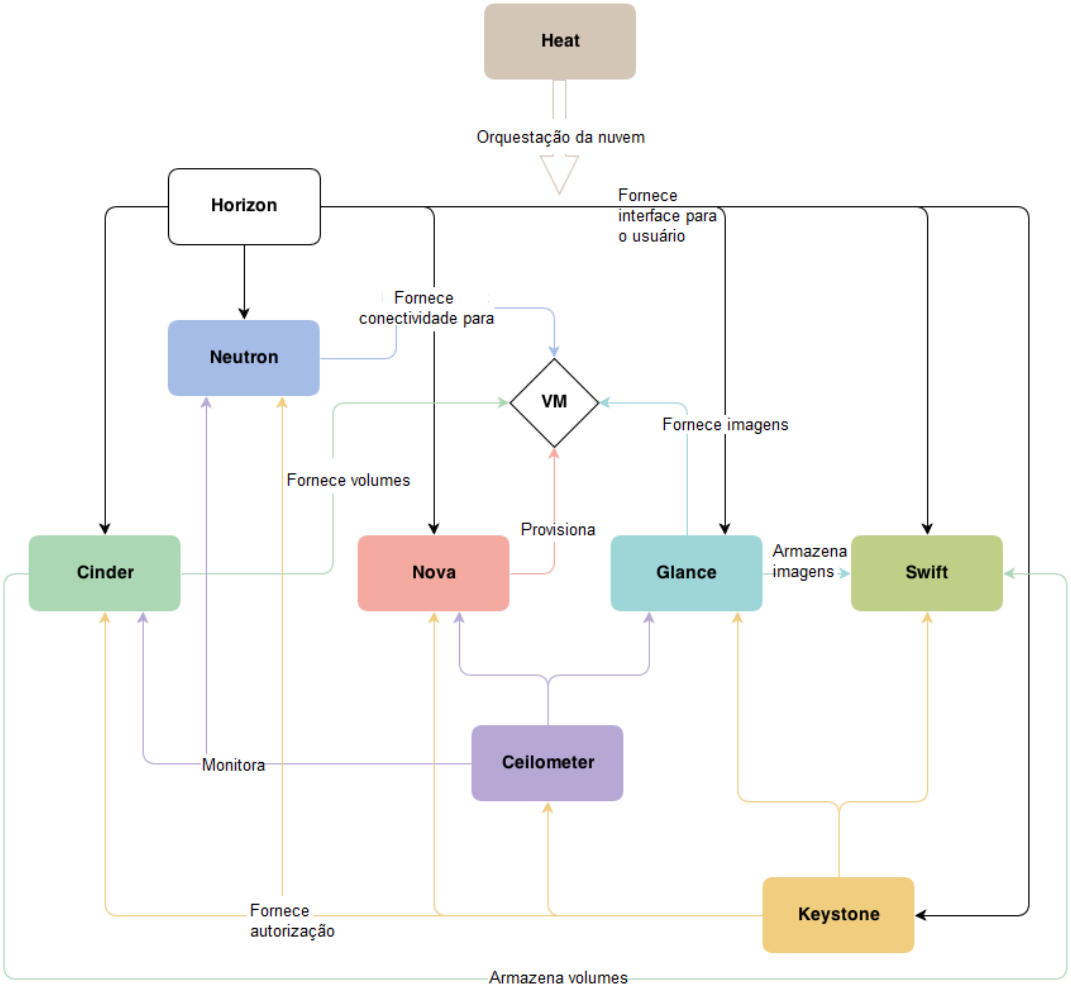


Figura 5.8: Visão geral da Arquitetura do OpenStack.

A Figura 5.8 abrange os componentes mais comuns empregados em instalações mais populares do OpenStack, tendo como ponto central a VM. Contudo, esse ponto central pode ser também um contêiner (usando o Magnum) ou uma imagem *bare-metal* (usando o Ironic). Independentemente da unidade de computação, pode-se perceber como o Keystone interage com todos os componentes, pois ele é o responsável pelos processos de autenticação e autorização em qualquer serviço do OpenStack.

Em um primeiro momento, o Keystone pode ser acessado através de uma interface web disponível no Horizon ou através de sua API específica. A interface web no Horizon é um processo interativo direto com o usuário que faz uso dos mecanismos internos do OpenStack para autenticação e autorização, não sendo objeto deste trabalho. As API já proporcionam o uso tanto do mecanismo interno como interfaces para uso de mecanismos externos de autenticação/autorização. Neste sentido, a API do Keystone possui suporte às seguintes soluções de identidade federada: LDAP, SAML2, ADFS, Keystone-to-Keystone, OpenID Connect e ABFAB. Esse suporte é tipicamente disponibilizado através de plugins e extensões do OpenStack.

Em um segundo momento, quando ocorre o processo de autenticação de usuários, o Keystone gera um *token* de acesso para uso interno do OpenStack após verificar a validade das credenciais do usuário em questão. Esse *token* é empregado para autorizar e autenticar o acesso aos recursos disponíveis dentro da nuvem OpenStack, possuindo validade e expiração (Figura 5.9).

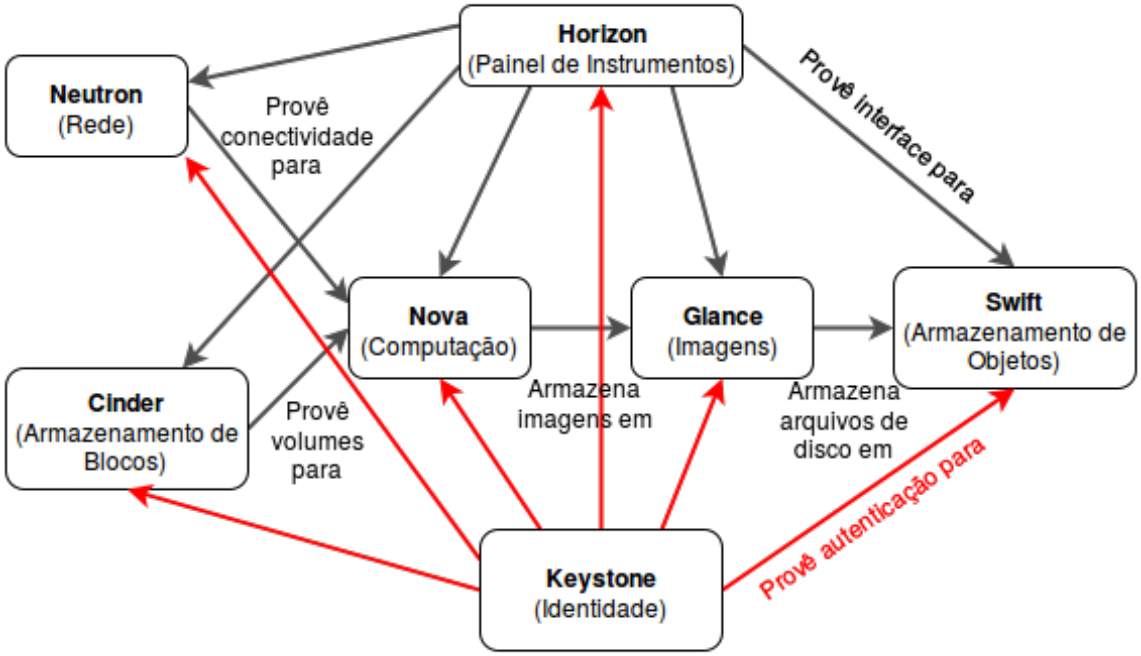


Figura 5.9: Visão geral interação do Keystone com demais componentes do OpenStack.

Internamente ao OpenStack, o acesso aos recursos é proporcionado através de um *token* que criado no momento da autenticação. Sendo assim, a Figura 5.9 permite identificar como o Keystone relaciona-se com outros componentes do OpenStack. Em um nível macro, a Figura 5.10 ilustra um exemplo das trocas de mensagens (em alto nível) no processo de autenticação de um usuário e o uso das credenciais para a criação de uma VM. Na Figura 5.10, a etapa de envio da credenciais do Usuário ao Keystone reflete o momento do processo no qual o OpenID Connect pode ser empregado.

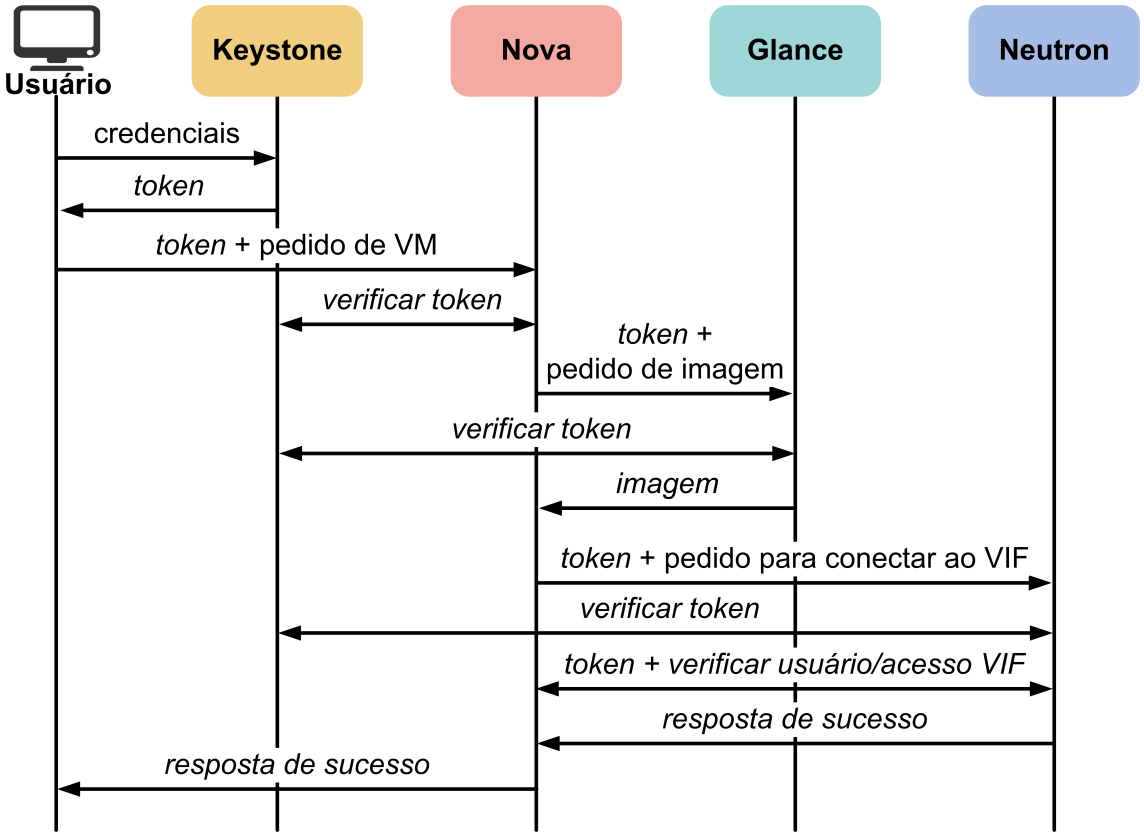


Figura 5.10: Visão macro da solicitação de criação de uma VM.

5.6. Estudo de caso OpenID Connect com o OpenStack

Como citado brevemente na Seção 5.5, o Keystone suporta a adição de extensões e *plugins* para os mais diversos propósitos. Dessa forma, o Keystone permite a utilização da extensão `OS-FEDERATION` para autenticação única. A extensão foi baseada no fato de que o Keystone é executado sobre o Apache HTTP Server. Uma vez que o Apache permite a instalação de *plugins* e módulos para as mais diversas finalidades, a extensão foi desenvolvida para usar esse aspecto. Assim, se um IdP *Security Assertion Markup Language* (SAML) for utilizado, então estão disponíveis os *plugins* `mod_shib` e `mod_auth_melon`; se um IdP OpenID Connect for utilizado, existe o plugin `mod_auth_openidc` [70].

O OpenID Connect utiliza *claims*, que requisitam e fornecem atributos dos usuários registrados. As *claims* do OpenID Connect são fornecidas através de um arquivo JSON assinado e cifrado [15, 70]. Para a conversão dos atributos do usuário obtidos do IdP através dos *claims*, são utilizadas regras de mapeamento.

Um mapeamento especifica quais usuários podem acessar o serviço e em qual grupo e projeto os mesmos devem ser alocados. Regras de mapeamento possuem duas seções: uma local e outra remota. A seção local representa o recurso do Keystone para o qual os usuários serão mapeados (*e.g.,* nome, email) e a seção remota representa os atributos do usuário no cabeçalho HTTP, como pode ser observado na Figura 5.11.

```
[
    {
        "local": [
            {
                "group": {
                    "id": "4e48b0c139c94cf088a665d919ee10f1"
                }
            }
        ],
        "remote": [
            {
                "type": "HTTP_OIDC_ISS",
                "any_one_of": [
                    "https://accounts.google.com"
                ]
            }
        ]
    }
]
```

Figura 5.11: Mapeamento com IdP do Google [70].

A Figura 5.11 ilustra um exemplo básico de mapeamento. Qualquer usuário que empregue o Google para se autenticar, essencialmente utilizando o emissor `https://accounts.google.com`, é mapeado para o grupo indicado na seção local. Além do mapeamento, é necessário criar o provedor de identidades no OpenStack e seu respectivo protocolo SSO, *i.e.,* OpenID Connect. Também é necessário configurar o Keystone para aceitar as requisições de autenticação SSO e habilitar a escolha do tipo de autenticação no Horizon. Essa etapa é realizada nos arquivos de configuração do Keystone, do Horizon e nos arquivos de *hosts* virtuais do Apache. Uma vez que estas configurações são definidas, os usuários podem se autenticar utilizando o IdP configurado.

5.6.1. Autenticação com OpenID Connect no OpenStack

Na Subseção 5.4.3 são descritas as etapas do processo de autenticação do OpenID Connect. No OpenStack essas interações ocorrem da mesma forma. Porém, para melhor entendimento do processo de autenticação do OpenID Connect no OpenStack o mesmo foi dividido neste estudo de caso da seguinte forma:

- Comunicação entre User-Agent e Keystone;
- Comunicação entre User-Agent e IdP; e
- Comunicação entre Keystone e IdP.

Ainda, com o objetivo de facilitar a compreensão do fluxo, cada canal de comunicação é descrito em maiores detalhes, indicando as informações que trafegam entre as partes.

5.6.1.1. Comunicação entre User-Agent e Keystone

O Keystone é responsável por gerenciar as identidades do OpenStack e isso envolve os processos de autenticação e autorização, realizados através de requisições nos seus *endpoints*. O Keystone possui dois *endpoints*: um escutando requisições na porta TCP/5000;

e outro na porta TCP/35357. O *endpoint* da porta TCP/35357 não é empregado nesse processo de autenticação, uma vez que é utilizado para tarefas administrativas apenas. A Figura 5.12 expõe o diagrama de sequência da comunicação entre o UA e o OpenStack durante o processo de autenticação.

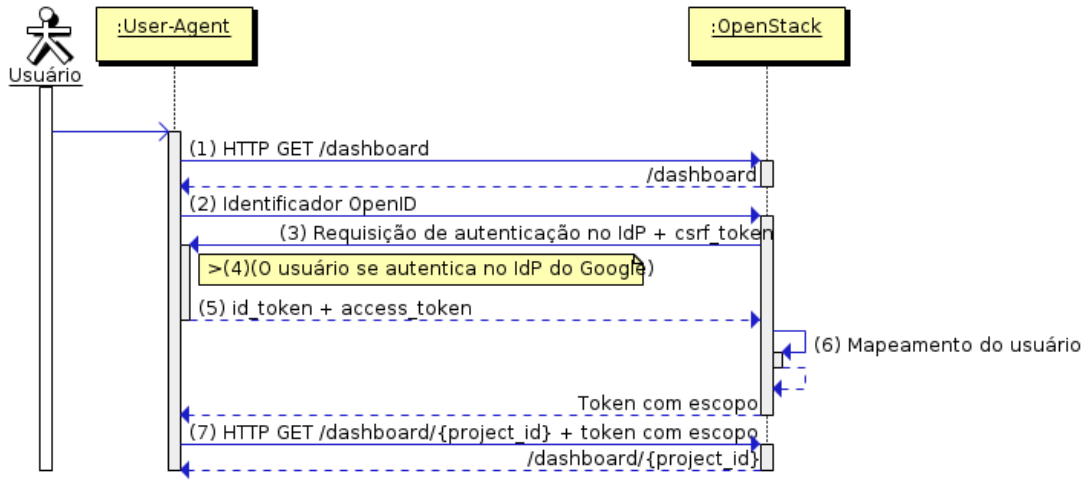


Figura 5.12: Diagrama de sequência da comunicação entre o UA e o SP no processo de autenticação SSO.

Na Figura 5.12 é possível observar o tráfego de dados entre o UA e o OpenStack:

1. O usuário, por meio do UA, requisita ao OpenStack o *dashboard* para se autenticar;
2. O usuário seleciona o OpenID Connect para se autenticar no OpenStack;
3. O OpenStack envia uma requisição ao IdP do Google, redirecionada pelo UA;
4. Nesta etapa, que é descrita em mais detalhes posteriormente, o usuário se autentica no Google, que retornará o *id_token* e o *access_token*;
5. O *id_token* e o *access_token* serão redirecionados ao OpenStack;
6. Através do *id_token*, que é um *JSON Web Token* (JWT), o OpenStack obtém as *claims* necessárias para autenticar o usuário, como o *iss*, *email*, e *name*. O *access_token* será utilizado caso exista a necessidade de obter outras *claims* do usuário através de chamadas de API (e.g., verificar se o usuário pertence a determinado grupo); e
7. Após a autenticação e mapeamento do usuário, o UA redireciona o usuário para o *dashboard* do seu projeto.

5.6.1.2. Comunicação entre User-Agent e IdP

Nessa etapa, o usuário deve se autenticar com sua conta no Google e então autorizar a aplicação a acessar seus dados. Nesse ponto o OAuth 2 é responsável por requisitar a

permissão do usuário, o detentor dos recursos, para autorizar o acesso da aplicação, a terceira parte, aos recursos solicitados. O fluxo de autenticação e autorização é mostrado na Figura 5.13.

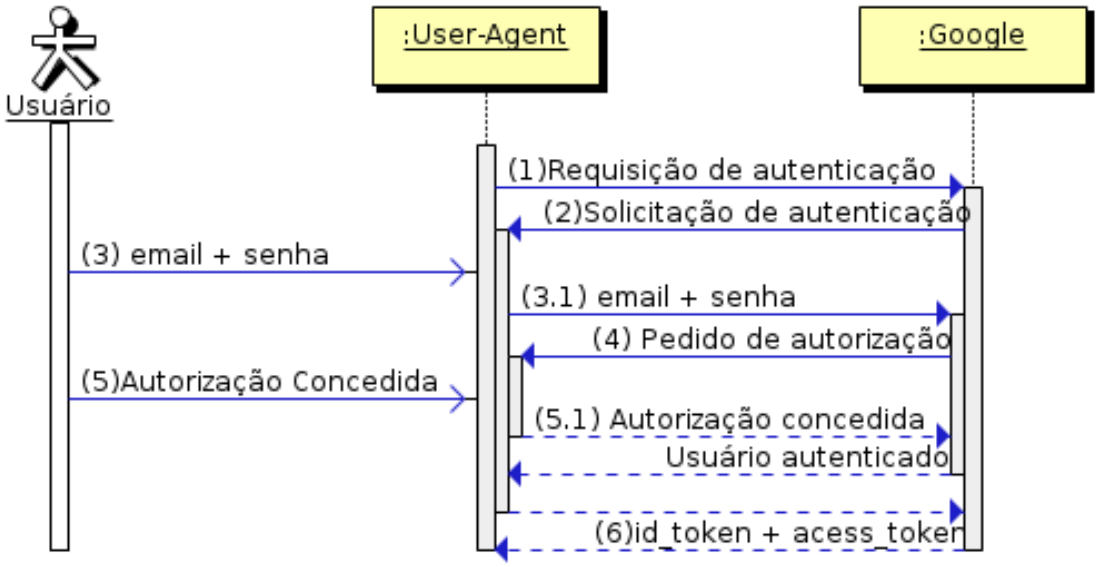


Figura 5.13: Diagrama de sequência: processo de autenticação/autorização Google.

A autenticação no Google (diagrama da Figura 5.13) é iniciada a partir de uma requisição de autenticação, gerada pelo OpenStack e segue como descrito:

1. Uma requisição redirecionada do OpenStack inicia o processo de autenticação;
2. O Google solicita que o usuário se autentique, encaminhando a página de *login*;
3. O usuário informa suas credenciais;
4. O Google então solicita que o usuário autorize a aplicação a acessar seus dados, informando os dados que estarão acessíveis ao OpenStack através da tela de consentimento;
5. O usuário autoriza a aplicação do OpenStack a acessar suas informações; e
6. O Google retorna então o *id_token* e o *access_token* ao UA, que redirecionará posteriormente ao OpenStack.

Após a autenticação com o Google o usuário é redirecionado ao OpenStack, através do UA, portando o `id_token` e o `access_token`. Nesse momento, o processo de autenticação continua na etapa *5* da Figura 5.12.

5.6.1.3. Comunicação entre Keystone e IdP

A comunicação entre o OpenStack e o IdP do Google ocorre através do redirecionamento das requisições através do UA. Ainda que na configuração do OpenID Connect seja necessário informar uma *Uniform Resource Locator* (URL) para o processo de descoberta, o OpenStack faz o acesso somente ao documento para obter as informações necessárias para autenticar o usuário no IdP (e.g., *endpoints*, escopos suportados). Dessa forma, não existe tráfego direto entre o OpenStack e o IdP do Google no processo de autenticação. O diagrama da Figura 5.14 mostra as mensagens trocadas entre o OpenStack e o IdP do Google.

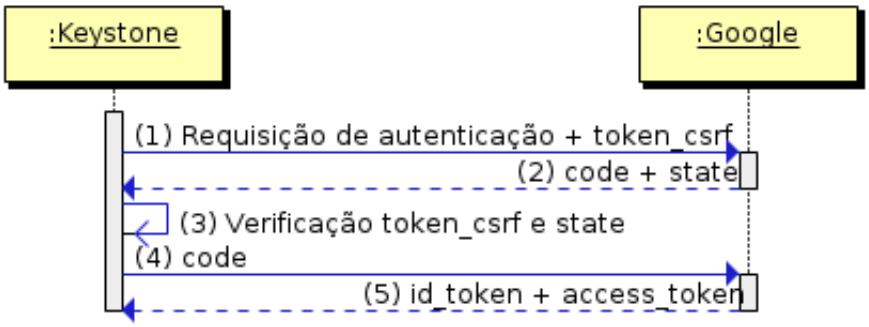


Figura 5.14: Diagrama de sequência da comunicação entre o OpenStack e o IdP do Google.

Para melhor entendimento, o UA foi omitido no diagrama da Figura 5.14. Todas as mensagens trocadas entre o OpenStack e o IdP do Google passam pelo UA. Uma breve descrição do fluxo das mensagens:

1. O OpenStack envia um *token Cross-Site Request Forgery* (CSRF) e o pedido de autenticação ao IdP do Google;

2. O IdP responde o pedido de autenticação, passando o parâmetro *state* e o parâmetro *code*;

3. O *token state* é comparado com o *token* CSRF da sessão e validado;

4. O OpenStack utiliza o *token code* e as informações do cliente criado no Google Developers Console [5] para solicitar o *id_token* e o *access_token*; e

5. O IdP retorna o *id_token* e o *access_token* para o OpenStack.

É possível solicitar outras informações do usuário, através da APIs, e utilizá-las para o mapeamento. Nesse caso, após o OpenStack obter o *access_token*, é feita uma requisição ao IdP solicitando o acesso à API e então os dados do usuário são obtidos.

[5]https://console.developers.google.com/

5.6.2. Configuração do OpenID Connect no OpenStack

Para habilitar a autenticação única com o OpenID Connect no OpenStack são necessários alguns passos que envolvem a configuração e a habilitação do método de autenticação. Primeiramente, é necessária a criação das credenciais no IdP, nesse caso o Google[6]. Em seguida, é realizada a configuração do Apache e dos arquivos de *hosts* virtuais do Keystone e do Horizon. Os arquivos de configuração do Keystone e do Horizon também devem ser alterados, para habilitar a opção de autenticação com o OpenID Connect. Por fim, é necessário registrar o IdP e as regras de mapeamento dos usuários no OpenStack. Ao final desse processo, a autenticação única estará habilitada para os usuários.

5.6.2.1. Criação das Credenciais no IdP

Inicialmente, é preciso a criação de um projeto e das suas credenciais no painel de desenvolvedores do Google. A criação do projeto e das credenciais é necessária para que o OpenStack tenha acesso aos dados dos usuários pois, do contrário, não seria possível solicitar as informações ao Google. São necessárias as credenciais do tipo "ID do cliente OAuth"para aplicações *web*, uma vez que é necessário definir a origem das requisições e a URL de redirecionamento. Para este exemplo, a origem das requisições será o próprio *Fully Qualified Domain Name* (FQDN) com o prefixo `http://`, ou seja, `http://nuvem.com`. Já a URL de redirecionamento será `http://nuvem.com:5000/v3/auth/OS-FEDERATION/websso/oidc/redirect`, através da qual o OpenStack manipulará a resposta recebida do Google. A Figura 5.15 ilustra a criação das credenciais no painel do Google.

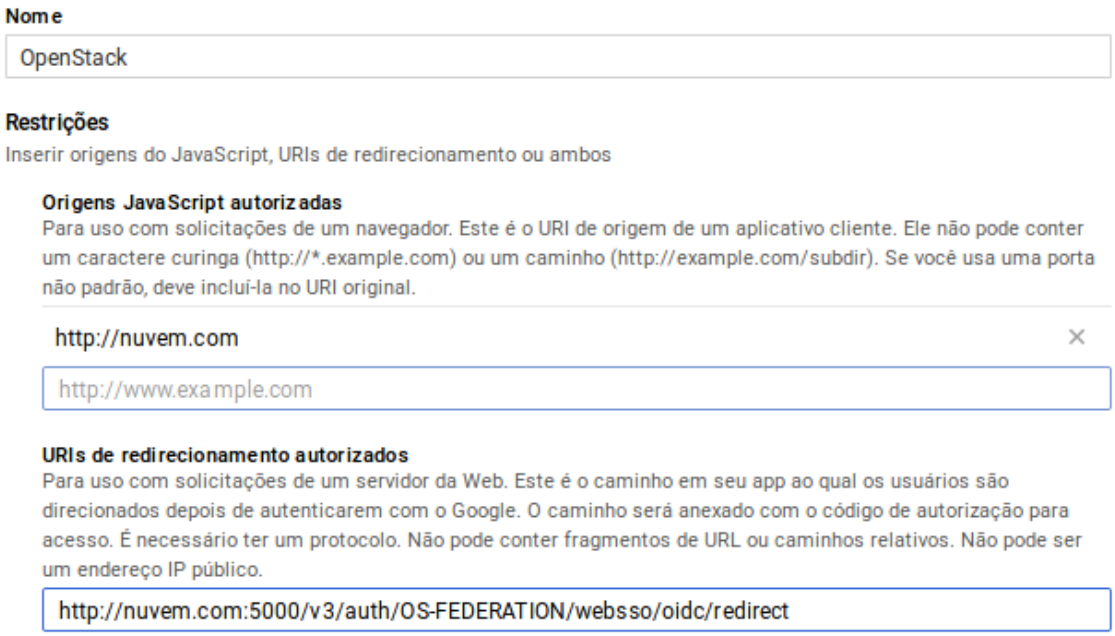


Figura 5.15: Criação das credenciais no painel de desenvolvedores do Google.

[6] `https://console.developers.google.com/apis`

Ao final desse processo, o Google retornará o identificador e a chave secreta que será utilizada na configuração do *plugin* `mod_auth_openidc`. Para habilitar o uso do *plugin* OpenID Connect no Keystone, é necessário instalar o *plugin* `mod_auth_openidc` e configurar o arquivo de *host* virtual do Keystone para aceitar a autenticação SSO a partir do IdP configurado.

5.6.2.2. Configuração do Plugin do OpenID Connect no Apache

Para a instalação do *plugin* `mod_auth_openidc` é necessário instalar o `hiredis` primeiramente, uma biblioteca cliente minimalista para o banco de dados Redis. Após ser instalado, o *plugin* `mod_auth_openidc` deve ser propriamente carregado e configurado no arquivo de *host* virtual do Keystone. No RDO OpenStack[7], este arquivo é localizado em `/etc/httpd/conf.d/10-keystone_wsgi_main.conf`. Para o funcionamento correto, o *plugin* do OpenID Connect deve ser carregado no início do arquivo.

```
LoadModule auth_openidc_module /usr/lib64/httpd/
   modules/mod_auth_openidc.so
```

Outras informações devem ser inseridas dentro do bloco `<VirtualHost *:5000>`, abaixo da linha `DocumentRoot`.

```
OIDCClaimPrefix "OIDC-"
OIDCResponseType "id_token"
OIDCScope "openid email profile"
OIDCProviderMetadataURL https://accounts.google.com/.
   well-known/openid-configuration
OIDCClientID <google_client_id>
OIDCClientSecret <google_client_secret>
OIDCCryptoPassphrase openstack
OIDCRedirectURI http://nuvem.com:5000/v3/auth/OS-
   FEDERATION/websso/oidc/redirect

<LocationMatch "/v3/auth/OS-FEDERATION/websso/oidc">
    AuthType openid-connect
    Require valid-user
    LogLevel debug
</LocationMatch>
```

Essas informações definem atributos utilizados no processo de autenticação com o IdP, nesse caso o do Google. São definidos o tipo de resposta, o escopo desejado, a URL do documento de descoberta, o identificador do cliente e a chave secreta, além do *Uniform Resource Identifier* (URI) de redirecionamento após a autenticação. Dessa forma, o Keystone utilizará essa configuração para os usuários que optarem por se autenticar com o Google.

[7] `https://www.rdoproject.org/`

Ainda com relação ao Apache, é necessário fazer com que o OpenStack responda pelo FQDN quando solicitado, caso contrário responderá somente pelo endereço *Internet Protocol* (IP). Contudo, o Google impede que um endereço IP seja usado para redirecionar os usuários após a autenticação. Dessa forma é obrigatório o uso de um FQDN. Para que o OpenStack redirecione corretamente os usuários, uma entrada deve ser criada no arquivo `/etc/hosts` apontando o endereço IP da interface externa para o FQDN. Para esse exemplo, o IP utilizado será `10.0.0.1` e o FQDN será `nuvem.com`. O arquivo de *hosts* ficará com o seguinte aspecto:

```
127.0.0.1       localhost
(...)
10.0.0.1        nuvem.com
```

Além disso, é necessário adicionar um novo `ServerAlias` no arquivo `/etc/httpd/conf.d/15-horizon_vhost.conf` definindo um novo nome, nesse caso o FQDN, para o servidor. A nova linha deverá ser inserida abaixo dos *aliases* existentes. O arquivo terá esse aspecto:

```
(...)
ServerAlias 10.0.0.1
ServerAlias nuvem.com
(...)
```

5.6.2.3. Configuração do Keystone e Horizon

Após configurar o Apache, é necessário ajustar as configurações do Keystone no arquivo `/etc/keystone/keystone.conf`. Na seção `[auth]` é necessário adicionar o método de autenticação através do OpenID Connect e também definir o *plugin* que irá gerenciar a autenticação. Na seção `[federation]` é necessário o atributo para obter o identificador do IdP e os *dashboards* confiáveis. O arquivo terá o seguinte conteúdo:

```
(...)
[auth]

# Allowed authentication methods. (list value)
methods = external,password,token,oauth1,oidc

oidc = keystone.auth.plugins.mapped.Mapped
(...)
[federation]

remote_id_attribute = HTTP_OIDC_ISS

trusted_dashboard = http://nuvem.com/dashboard/auth/
   websso/
(...)
```

Uma vez que o Apache e o Keystone estejam configurados, é necessário configurar o Horizon para habilitar a opção de autenticação com o OpenID Connect. Para isso, é necessário configurar o Horizon para utilizar a *Identity API v3*, uma vez que a mesma não é utilizada por padrão para autenticar os usuários, além de definir a URL do Keystone atualizada com o FQDN. Também é necessário configurar as opções de autenticação para que os usuários possam escolher o método mais adequado. Essas informações são inseridas no arquivo `/etc/openstack_dashboard/local_settings`:

```
(...)
OPENSTACK_KEYSTONE_URL = http://nuvem.com:5000/v3
(...)
OPENSTACK_API_VERSIONS =
        "identity": 3
}

WEBSSO_ENABLED = True

WEBSSO_CHOICES = (
        ("credentials", _("Keystone␣Credentials")),
        ("oidc", _("OpenID␣Connect"))
)

WEBSSO_INITIAL_CHOICE = "oidc"
```

Ao final desse processo, é necessário reiniciar o servidor *web* para que as alterações tenham efeito.

5.6.2.4. Registro do IdP no OpenStack

O próximo passo consiste em criar um registro para o IdP no OpenStack e as regras de mapeamento que serão utilizadas para redirecionar os usuários aos seus respectivos projetos. As regras de mapeamento devem ser armazenadas em um arquivo, para facilitar o registro. Por padrão, esse arquivo é denominado `mapping.json`. Para este exemplo, é criada uma regra que mapeia qualquer usuário que possua uma conta no Google para um grupo em específico. Assim é criado um grupo e um projeto é atribuído a esse grupo:

```
# openstack group create grupo_google -f value -c id
# openstack role add _member_ --group grupo_google --
   project usuarios_google
```

O primeiro comando retornará o identificador do grupo criado, que será utilizado na regra de mapeamento:

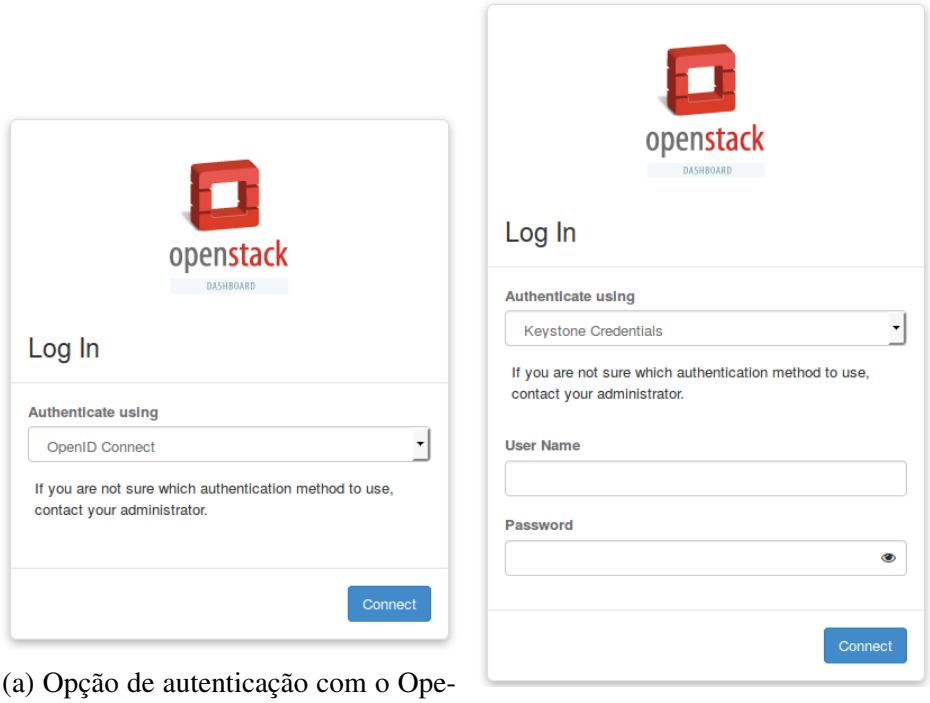
```
{
        "local": [
                {
                        "group": {
                                "id": "<id_do_grupo>"
                        }
                }
        ],
        "remote": [
                {
                        "type": "HTTP_OIDC_ISS",
                        "any_one_of": [
                                "https://accounts.
                                   google.com"
                        ]
                }
        ]
}
```

Essa regra permite que qualquer usuário que tenha um identificador válido, emitido por `https://accounts.google.com`, será mapeado para o grupo recém criado. Outros atributos podem ser utilizados para mapear os usuários (*e.g.,* email, organização, nome e sobrenome).

Por fim, basta registrar o IdP e a regra criada no OpenStack.

```
# openstack identity provider create google --remote-
   id https://accounts.google.com
# openstack mapping create google_mapping --rules
   mapping.json
# openstack federation protocol create oidc --
   identity-provider google --mapping google_mapping
```

Uma vez que a configuração esteja finalizada e correta, os usuários estarão aptos a selecionar o Google como provedor de identidades do OpenID Connect. Assim, OpenStack permite que os usuários escolham o método de autenticação mais adequado, como ilustra a Figura 5.16.



(a) Opção de autenticação com o OpenID Connect

(b) Opção de autenticação padrão

Figura 5.16: Opções de autenticação no OpenStack

A Figura 5.16a ilustra a opção de autenticação com o OpenID Connect. Uma vez que o usuário clique no botão "Connect", ele será redirecionado para se autenticar em sua conta do Google. Uma vez autenticado, será solicitado ao usuário a concessão de acesso à aplicação, ou seja, o OAuth solicitará ao usuário permissão para que a aplicação do OpenStack tenha acesso aos seus dados. As permissões solicitadas são mostradas na Figura 5.17.

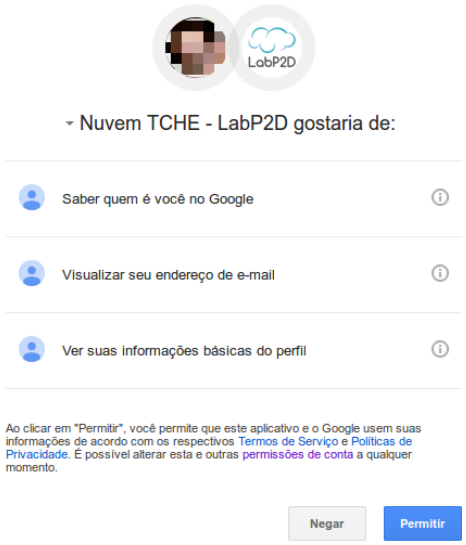


Figura 5.17: Tela de consentimento OAuth

Se o usuário permitir que o OpenStack acesse seus dados, o processo de autenticação terá continuidade e o usuário será mapeado de acordo com as regras. Do contrário, o processo de autenticação é encerrado. É importante ressaltar que as permissões aqui solicitadas foram definidas na configuração do arquivo de *host* virtual do Keystone no Apache. O parâmetro `OIDCScope "openid email profile"` define:

- Identificador OpenID: Responsável por solicitar o identificador do usuário no Google, ou seja, saber quem é o usuário no Google;
- Email: Responsável por solicitar o email do usuário; e
- Perfil (*Profile*): Responsável por solicitar acesso ao perfil público do usuário, coletando informações como o nome e sobrenome.

É possível especificar outros escopos e solicitar dados de outros serviços do IdP. No caso específico do Google, é possível solicitar acesso às APIs de qualquer serviço. Como exemplo, é possível autenticar somente usuários que fazem parte de um grupo dentro do domínio da organização no Google Apps, requisitando acesso às APIs de grupos de domínio e mapeando os usuários que estão no grupo especificado. Assim, é possível mapear os usuários de acordo com as funções desempenhadas dentro da organização.

5.7. Considerações

A taxonomia proposta consolida as diversas taxonomias já existentes, consistindo em uma evolução para a classificação de soluções SSO com um viés orientado para computação em nuvem. A taxonomia consolidada se mostrou mais abrangente e permite a classificação das soluções em um nível de granularidade mais fina, quando comparada com as demais taxonomias. Adicionalmente, a classificação de várias soluções SSO permitiu atestar a coerência da taxonomia proposta ao mesmo tempo que fornece um referencial teórico para outros pesquisadores.

O usos de soluções SSO baseadas no OpenID, como OpenID Connect, tem possibilitado um meio facilitado de autenticação e autorização para diversos serviços web. Contudo, constata-se que o uso de serviços de SSO ainda é pouco significativo no caso de nuvens computacionais. Algumas nuvens experimentais, como TryStack[8], usam mecanismos de SSO a fim de facilitar o acesso de usuários a sua plataforma OpenStack de testes e experimentação. Entretanto, o TryStack emprega o Facebook Connect que consiste em uma versão modificada do OpenID e OAuth desenvolvida pelo Facebook. Sendo assim, o Facebook Connect não é nativamente compatível com o plugin do OpenID Connect disponível na plataforma OpenStack e exige o desenvolvimento de software adicional (*e.g., proxy*) para permitir o seu uso.

A solução de IdP da Google, por outro lado, segue os padrões do OpenID / OAuth tornando possível o seu uso no OpenStack através do plugin do OpenID Connect. Porém, a integração não é trivial como mostrado no estudo de caso deste trabalho, exigindo cautela em diversos pontos para evitar problemas de segurança.

[8] `http://trystack.org/`

Nos últimos anos, soluções como o OpenID Connect, proporcionaram um novo paradigma em relação ao gerenciamento de identidades na Internet. As nuvens computacionais tiveram um crescimento considerável nos últimos anos e algumas delas já utilizam mecanismos de autenticação única para prover de seus recursos aos usuários. O uso de SSO com IdP externos proporciona uma alternativa para tratar com volumes consideráveis de usuários que se cadastram ou deixam de usar um serviço, facilitando as tarefas operacionais. Contudo não se deve ignorar os aspectos de segurança que devem ser observados. Uma outra pesquisa [71], de alguns dos autores deste trabalho, faz uma análise destes aspectos de segurança e deve ser levada em consideração quando realizar um estudo de viabilidade de uso de IdP externos para autenticar usuários em nuvens computacionais.

Agradecimentos

O presente trabalho foi em parte financiado pelo CNPq, através da Bolsa de Produtividade em Pesquisa 305350/2013-7.

O presente trabalho foi em parte financiado pelo CNPq, através da Bolsa de Produtividade Desenvolvimento Tecnológico e Extensão Inovadora Nível 2 311667/2014-7.

Os autores agradecem o apoio do Laboratório de Processamento Paralelo e Distribuído (LabP2D) no Centro de Ciências tecnológicas (CCT) da Universidade do Estado de Santa Catarina (UDESC).

Os autores agradecem o apoio do Laboratório de Arquitetura e Redes de Computadores (LARC) do Departamento de Engenharia de Produção e Sistemas Digitais (PCS) da Escola Politécnica da Universidade de São Paulo (USP).

Esse trabalho foi financiado com recursos do Edital PIPES 001/2015 da UDESC.

Referências

[1] M. E. Haloui and A. Kriouile, "A decision-support model enabling a proactive vision of cloud computing adoption," in *2016 2nd International Conference on Cloud Computing Technologies and Applications (CloudTech)*, May 2016, pp. 192–198.

[2] P. Mell and T. Grance, "The nist definition of cloud computing," NIST, Tech. Rep., 2011.

[3] E. Andrade, M. Simplicio, P. Barreto, and P. Santos, "Lyra2: efficient password hashing with high security against time-memory trade-offs," *IEEE Transactions on Computers*, vol. 65, no. 10, pp. 3096–3108, 2016, see also: http://eprint.iacr.org/2015/136.

[4] S. Barhate and M. Dhore, "User Authentication Issues In Cloud Computing," *IOSR Journal of Computer Engineering (IOSR-JCE)*, pp. 30–35, 2016. [Online]. Available: http://www.iosrjournals.org/iosr-jce/pages/Conf.15013(4).html

[5] H.-K. Oh and S.-H. Jin, "The Security Limitations of SSO in OpenID," in *10th International Conference on Advanced Communication Technology, 2008. ICACT 2008*, vol. 3, feb 2008, pp. 1608–1611.

[6] J. B. Abdo, J. Demerjian, H. Chaouchi, K. Barbar, and G. Pujolle, "Single-Sign-On in operator centric mobile cloud architecture," in *MELECON 2014 - 2014 17th IEEE Mediterranean Electrotechnical Conference*, apr 2014, pp. 151–155.

[7] D. S. Brands, *A Technical Overview of Digital Credentials*. Credentica.com, 2002.

[8] R. R. M. Sakuragui, "Gerenciamento de identidades com privacidade do usuário em ambiente web." Ph.D. dissertation, Universidade de São Paulo, 2011.

[9] F. Corradini, E. Paganelli, and A. Polzonetti, "The e-Government digital credentials," *International Journal of Electronic Governance*, vol. 1, no. 1, p. 17, 2007. [Online]. Available: http://www.inderscience.com/link.php?id=14341

[10] A. Herzberg and Y. Mass, "Relying Party Credentials Framework," in *Topics in Cryptology — CT-RSA 2001*, G. Goos, J. Hartmanis, J. van Leeuwen, and D. Naccache, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, vol. 2020, pp. 328–343, dOI: 10.1007/3-540-45353-9_25. [Online]. Available: http://link.springer.com/10.1007/3-540-45353-9_25

[11] R. M. Needham and M. D. Schroeder, "Using Encryption for Authentication in Large Networks of Computers," *Commun. ACM*, vol. 21, no. 12, pp. 993–999, Dec. 1978. [Online]. Available: http://doi.acm.org/10.1145/359657.359659

[12] N. M. Gonzalez, M. A. T. Rojas, M. V. M. d. Silva, F. Redígolo, T. C. M. d. B. Carvalho, C. C. Miers, M. Näslund, and A. S. Ahmed, "A Framework for Authentication and Authorization Credentials in Cloud Computing," in *2013 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications*, Jul. 2013, pp. 509–516.

[13] M. Bishop, *Introduction to Computer Security*, 1st ed. Addison-Wesley Professional, nov 2004.

[14] M. Urueña, A. Muñoz, and D. Larrabeiti, "Analysis of privacy vulnerabilities in single sign-on mechanisms for multimedia websites," *Multimedia Tools and Applications*, vol. 68, no. 1, pp. 159–176, jan 2014.

[15] I. Sette and C. Ferraz, "Integrating cloud platforms to identity federations," in *2014 Brazilian Symposium on Computer Networks and Distributed Systems (SBRC)*, 2014, pp. 310–318.

[16] B. C. Neuman, "Proxy-based authorization and accounting for distributed systems," in *[1993] Proceedings. The 13th International Conference on Distributed Computing Systems*, May 1993, pp. 283–291.

[17] C. Li and Z. Liao, "An Extended ACL for Solving Authorization Conflicts," in *2009 Second International Symposium on Electronic Commerce and Security*, vol. 1, May 2009, pp. 30–34.

[18] R. S. Sandhu and P. Samarati, "Access control: principle and practice," *IEEE Communications Magazine*, vol. 32, no. 9, pp. 40–48, Sep. 1994.

[19] G.-J. Ahn and R. Sandhu, "Role-based Authorization Constraints Specification," *ACM Trans. Inf. Syst. Secur.*, vol. 3, no. 4, pp. 207–226, Nov. 2000. [Online]. Available: http://doi.acm.org/10.1145/382912.382913

[20] R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman, "Role-based access control models," *Computer*, vol. 29, no. 2, pp. 38–47, Feb. 1996.

[21] R. Sandhu, E. Bertino, J. Jaeger, R. Kuhn, and C. Landwehr, "Panel: The next generation of acess control models (panel session): Do we need them and what should they be?" in *Proceedings of the Sixth ACM Symposium on Access Control Models and Technologies*, ser. SACMAT '01. New York, NY, USA: ACM, 2001, pp. 53–. [Online]. Available: http://doi.acm.org/10.1145/373256.373262

[22] J. Longstaff and J. Noble, "Attribute Based Access Control for Big Data Applications by Query Modification," in *2016 IEEE Second International Conference on Big Data Computing Service and Applications (BigDataService)*, Mar. 2016, pp. 58–65.

[23] V. C. Hu, D. Ferraiolo, R. Kuhn, A. R. Friedman, A. J. Lang, M. M. Cogdell, A. Schnitzer, K. Sandlin, R. Miller, and K. Scarfone, "Guide to attribute based access control (ABAC) definition and considerations (draft)," *NIST special publication*, vol. 800, no. 162, 2013.

[24] W. W. Armour, J. Kickenson, J. Koilpillai, and M. A. Salim, "NIST SP 500-299: Cloud Computing Security Reference Architecture," may 2013. [Online]. Available: http://bigdatawg.nist.gov/_uploadfiles/M0007_v1_3376532289.pdfâĂŐ

[25] M. Sugumaran, B. Murugan, and D. Kamalraj, "An Architecture for Data Security in Cloud Computing," in *2014 World Congress on Computing and Communication Technologies (WCCCT)*, feb 2014, pp. 252–255.

[26] X.-e. You and Y. Zhu, "Research and design of Web Single Sign-On scheme," in *2012 IEEE Symposium on Robotics and Applications (ISRA)*, Jun. 2012, pp. 383–386.

[27] D. W. Chadwick, K. Siu, C. Lee, Y. Fouillat, and D. Germonville, "Adding federated identity management to OpenStack," *Journal of Grid Computing*, vol. 12, no. 1, pp. 3–27, 2013.

[28] M. Ahmadi, M. Chizari, M. Eslami, M. J. Golkar, and M. Vali, "Access control and user authentication concerns in cloud computing environments," in *2015 1st International Conference on Telematics and Future Generation Networks (TAFGEN)*, May 2015, pp. 39–43.

[29] A. Volchkov, "Revisiting single sign-on: a pragmatic approach in a new context," *IT Professional*, vol. 3, no. 1, pp. 39–45, jan 2001.

[30] B. B. Zhu, J. Yan, G. Bao, M. Yang, and N. Xu, "Captcha as graphical passwords: A new security primitive based on hard AI problems," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 6, pp. 891–904, 2014.

[31] J. Galbally, A. Ross, M. Gomez-Barrero, J. Fierrez, and J. Ortega-Garcia, "From the iriscode to the iris: A new vulnerability of iris recognition systems," *Black Hat Briefings USA*, 2012.

[32] K. Hamlen, P. Liu, M. Kantarcioglu, B. Thuraisingham, and T. Yu, "Identity management for cloud computing: Developments and directions," in *Proceedings of the Seventh Annual Workshop on Cyber Security and Information Intelligence Research*, ser. CSIIRW '11. New York, NY, USA: ACM, 2011, pp. 32:1–32:1. [Online]. Available: http://doi.acm.org/10.1145/2179298.2179333

[33] D. Cooper, "Internet X.509 public key infrastructure certificate and certificate revocation list (CRL) profile," 2008. [Online]. Available: https://tools.ietf.org/html/rfc5280

[34] MIT, "Kerberos: The network authentication protocol," 2015. [Online]. Available: http://web.mit.edu/kerberos/

[35] OpenID, "Openid," 2016. [Online]. Available: http://openid.net/

[36] A. Pashalidis and C. J. Mitchell, "A taxonomy of single sign-on systems," in *Australasian Conference on Information Security and Privacy*. Springer, 2003, pp. 249–264.

[37] J. D. Clercq, "Single sign-on architectures," in *Proceedings of the International Conference on Infrastructure Security (InfraSec'02)*. London, UK, UK: Springer-Verlag, 2002, pp. 40–58.

[38] A. Bhargav-Spantzel, J. Camenisch, T. Gross, and D. Sommer, "User centricity: A taxonomy and open issues," in *Proceedings of the 2nd ACM Workshop on Digital Identity Management*, ser. DIM '06. New York, NY, USA: ACM, 2006, pp. 1–10. [Online]. Available: http://doi.acm.org/10.1145/1179529.1179531

[39] P. B. Tiwari and S. R. Joshi, "Single sign-on with one time password," in *2009 First Asian Himalayas International Conference on Internet*, Nov 2009, pp. 1–4.

[40] K. Botzum, "Single sign on—a contrarian view," *Open Group Website*, pp. 1–8, 2001.

[41] M. Linden and I. Vilpola, "An empirical study on the usability of logout in a single sign-on system," in *Information Security Practice and Experience*, ser. Lecture Notes in Computer Science, R. H. Deng, F. Bao, H. Pang, and J. Zhou, Eds. Springer Berlin Heidelberg, apr 2005, no. 3439, pp. 243–254, dOI: 10.1007/978-3-540-31979-5_21.

[42] IBM, "Cloud & smarter infrastructure manage your business infrastructure in real time with tivoli," jun 2016. [Online]. Available: https://www.ibm.com/software/tivoli

[43] Microsoft, "RADIUS Authentication, Authorization, and Accounting (Windows)," 2015. [Online]. Available: https://msdn.microsoft.com/en-us/library/bb892012(v=vs.85).aspx

[44] J. Han, Y. Mu, W. Susilo, and J. Yan, "A generic construction of dynamic single sign-on with strong security," in *Security and Privacy in Communication Networks*. Springer, 2010, pp. 181–198.

[45] S. Suriadi, E. Foo, and A. Jøsang, "A user-centric federated single sign-on system," *Journal of Network and Computer Applications*, vol. 32, no. 2, pp. 388–401, 2009.

[46] Y. Zhang and J. L. Chen, "Universal identity management model based on anonymous credentials," in *2010 IEEE International Conference on Services Computing*, July 2010, pp. 305–312.

[47] ——, "A Delegation Solution for Universal Identity Management in SOA," *IEEE Transactions on Services Computing*, vol. 4, no. 1, pp. 70–81, Jan. 2011.

[48] OASIS, "Saml xml," 2016. [Online]. Available: http://saml.xml.org/

[49] L. Alliance, "Liberty alliance project," 2016. [Online]. Available: http://www.projectliberty.org/

[50] M. Ates, S. Ravet, A. Ahmat, and J. Fayolle, "An identity-centric internet: Identity in the cloud, identity as a service and other delights," in *2011 Sixth International Conference on Availability, Reliability and Security (ARES)*, aug 2011, pp. 555–560.

[51] Openstack, "Openstack open source cloud computing software," 2016. [Online]. Available: http://www.openstack.org/

[52] Amazon, "Amazon web services," 2016. [Online]. Available: http://aws.amazon.com/pt/

[53] Openstack, "Openstack identity api v3," 2016. [Online]. Available: http://developer.openstack.org/api-ref/identity/v3/index.html

[54] Microsoft, "Azure active directory," 2016. [Online]. Available: https://azure.microsoft.com/pt-br/services/active-directory/

[55] ——, "Introducing Windows CardSpace," 2006. [Online]. Available: https://msdn.microsoft.com/en-us/library/aa480189.aspx

[56] Entrust, "Entrust," 2016. [Online]. Available: http://www.entrust.com/

[57] Facebook, "Facebook login," 2016. [Online]. Available: https://www.facebook.com/about/login/

[58] FIDO, "FIDO Alliance," 2016. [Online]. Available: https://fidoalliance.org/

[59] Microsoft, "Microsoft passport," 2016. [Online]. Available: https://login.live.com/

[60] OpenID, "OpenID Connect," 2016. [Online]. Available: http://openid.net/connect/

[61] Shibboleth, "Shibboleth," 2016. [Online]. Available: https://shibboleth.net/

[62] R. Khan, J. Ylitalo, and A. Ahmed, "OpenID authentication as a service in OpenStack," in *2011 7th International Conference on Information Assurance and Security (IAS)*, Dec. 2011, pp. 372–377.

[63] M. Cordeiro Domenech, E. Comunello, and M. Silva Wangham, "Identity management in e-Health: A case study of web of things application using OpenID connect," in *2014 IEEE 16th International Conference on e-Health Networking, Applications and Services (Healthcom)*, Oct. 2014, pp. 219–224.

[64] Z. Obrenović and B. d. Haak, "Integrating User Customization and Authentication: The Identity Crisis," *IEEE Security Privacy*, vol. 10, no. 5, pp. 82–85, Sep. 2012.

[65] L. Lynch, "Inside the Identity Management Game," *IEEE Internet Computing*, vol. 15, no. 5, pp. 78–82, Sep. 2011.

[66] J. Sendor, Y. Lehmann, G. Serme, and A. Santana de Oliveira, "Platform-level Support for Authorization in Cloud Services with OAuth 2," in *2014 IEEE International Conference on Cloud Engineering (IC2E)*, Mar. 2014, pp. 458–465.

[67] M. N. Ko, G. P. Cheek, M. Shehab, and R. Sandhu, "Social-networks connect services," *Computer*, no. 8, pp. 37–43, 2010.

[68] X. Wen, G. Gu, Q. Li, Y. Gao, and X. Zhang, "Comparison of open-source cloud management platforms: OpenStack and OpenNebula," in *2012 9th International Conference on Fuzzy Systems and Knowledge Discovery*, may 2012, pp. 2457–2461.

[69] K. Jackson, *Openstack Cloud Computing Cookbook*. Packt Publishing Ltd, 2012.

[70] S. Martinelli, H. Nash, and B. Topol, *Identity, Authentication, and Access Management in OpenStack: Implementing and Deploying Keystone*, 1st ed. O'Reilly Media, Dec. 2015.

[71] G. C. Batista and C. C. Miers, "Security analysis of the openid connect protocol integration with an openstack cloud using an external idp," in *2016 XLII Latin American Computing Conference (CLEI)*, Oct 2016, pp. 1–12.

Capítulo 6

Comunicação por Luz Visível: Conceitos, Aplicações e Desafios

Luiz Eduardo Mendes Matheus (UFJF), Alex Borges Vieira (UFJF), Jean H. F. Freire (UFMG), Luiz F. M. Vieira (UFMG), Marcos A. M. Vieira (UFMG), Omprakash Gnawali (University of Houston)

Abstract

Mobile devices have become very popular during the past few years. Moreover, the exponentially increase on processing and storage capacities of these devices, created a huge demand for wireless communication and Internet access. In this context, Visible Light Communication (VLC) is an alternative to complement the current wireless infrastructure of the Internet. VLC applications can offer a number of improvements, such as data rate increase, ease of implementation and low-cost devices. Therefore, the purpose of this short course is to present to the participants the state-of-art, as well as the main concepts and challenges related to this ascending area.

Resumo

Dispositivos móveis tornaram-se muito populares nos últimos anos. Além disso, a capacidade de armazenamento e processamento de tais dispositivos cresce exponencialmente, gerando uma grande demanda por recursos wireless na Internet. Neste sentido, a Comunicação por Luz Visível (VLC) é uma alternativa para complementar a atual infraestrutura wireless da Internet. Aplicações VLC podem oferecer uma série de melhorias, como aumento na taxa de dados, facilidade de implementação e baixo custo de dispositivos. Sendo assim, a proposta deste minicurso é apresentar aos participantes o estado da arte, assim como os principais conceitos e desafios relacionados a esta área em constante crescimento.

6.1. Introdução

A quantidade de dispositivos móveis ao redor do mundo aumentou consideravelmente nos últimos anos. *Smartphones*, *tablets* e sensores estão cada vez mais comuns no dia a dia

das pessoas. Estes dispositivos tornam-se cada vez mais potentes –no sentido amplo da palavra– frente à demanda por mais recursos, tanto para acesso a conteúdo na Internet, quanto para sensoriamento contínuo e intermitente. Além de uma notória popularização e massificação de dispositivos móveis, há a aproximação iminente da próxima grande onda da computação: a Internet das Coisas (IoT), onde todo e qualquer dispositivo terá conectividade e processamento. Dispositivos comuns (televisão, microondas, geladeira, veículos) passam a representar nós na rede, exigindo ainda mais recursos, seja dos próprios dispositivos, seja da infraestrutura de rede de suporte.

Considerando este cenário, cada vez mais comum nos dias de hoje, um problema que surge e chama a atenção da comunidade acadêmica é a crescente lotação da faixa do espectro eletromagnético destinada ao WiFi[1] [De Vries et al., 2014]. Esta lotação de faixa, conhecida como "*WiFi Spectrum Crunch*", ocorre em ambientes onde há grande demanda de recursos *wireless*. Em muitos casos, nesses ambientes, a infraestrutura existente não consegue fornecer os recursos apropriados para comunicação sem fio.

Esta expressão foi repetida diversas vezes por grandes veículos de comunicação durante os últimos anos e, alarmou tanto comunidade acadêmica quanto a indústria. Por consequência, novas tecnologias –e.g., cooperação com o WiFi [Baylis et al., 2014, Haas, 2013]– vêm sendo estudadas com o objetivo de evitar tal situação.

Entre as novas tecnologias destinadas a resolver o problema de lotação de frequência, a Comunicação por Luz Visível (VLC) [Pohlmann, 2010, Haas, 2015] tem se mostrado promissora. De fato, o interesse em formas de comunicação ótica sem fio cresceu, principalmente, dada a possibilidade de cooperação com sistemas de frequência de rádio. Outro fator que atrai pesquisadores é a possibilidade de trabalhar com frequências muito maiores que as utilizadas em dispositivos WiFi, possibilitando assim comunicações sem fio em altíssimas velocidades (teoricamente na ordem de terabytes/s). Além disso, novas tecnologias de emissão de luz, como lâmpadas LEDs, têm se tornado mais populares e acessíveis, possibilitando novas perspectivas para cenários de comunicação ótica sem fio [Haruyama, 2013, Wang et al., 2013b]. Esta onda de estudos e exploração do espectro da luz visível trouxe uma série de inovações que já estão presentes no mercado, como a tecnologia LiFi [Tsonev et al., 2013], apresentada em 2011 e já comercializada por empresas especializadas em VLC.

Em suma, o objetivo deste minicurso é apresentar os principais conceitos envolvendo VLC, apresentar aplicações e desafios encontrados na área, e por fim, demonstrar a utilização da tecnologia em um ambiente real. Isso inclui as recentes aplicações de Comunicação por Luz Visível em áreas emergentes e discussão das questões em aberto que podem fomentar novos trabalhos de pesquisa e desenvolvimento na área.

O restante do minicurso está organizado da seguinte forma: inicialmente, é oferecida uma introdução fundamental para a área, junto de um breve histórico e motivação, ambos nesta Seção. Em seguida, na Seção 6.2, é dada uma visão geral da Comunicação por Luz Visível. Na Seção 6.3, serão estudados componentes de um sistema VLC. Na Seção 6.4, detalhes técnicos de codificação são abordados. A Seção 6.5 introduz uma

[1]Will we ever face a wireless "spectrum crunch"? - `http://www.bbc.com/future/story/20131014-are-we-headed-for-wireless-chaos`, 2013.

série de aplicações encontradas na literatura em que se utiliza VLC, e a Seção 6.6 apresenta os principais desafios encontrados na área atualmente. Por fim, na Seção 6.8, são apresentados as discussões finais do tema, assim como a perspectiva da área.

6.1.1. O que é VLC?

Comunicação por Luz Visível é o nome dado aos sistemas em que dados são enviados através da modulação das ondas de luz no espectro visível, ou seja, utiliza-se apenas a faixa do espectro eletromagnético que varia entre 390 nm a 700 nm. De maneira geral, qualquer sistema em que a informação possa ser transmitida utilizando algum tipo de luz visível aos olhos humanos pode ser chamado de Comunicação por Luz Visível. Entretanto, a ideia deste tipo de comunicação é a transferência de dados de maneira imperceptível aos olhos do ser humano, de forma que, o que se enxerga é apenas a iluminação comum, sem nenhuma alteração perceptível. Há diversas outras nomenclaturas criadas ao longo dos anos para tecnologias similares, como OWC (*Optical Wireless Communication*) e Li-Fi (*Light Fidelity*), por exemplo, que serão abordados com detalhes na seção 6.2.

As ondas de rádio vêm sendo estudadas por diversos trabalhos desde o início do século XX [Garratt, 1994]. Isso levou a uma série de descobertas acerca das propriedades desse tipo de onda, trazendo diversas novas tecnologias para o dia-a-dia das pessoas ao redor do mundo, seja na área militar, ou na área médica. A eficiência de comunicações por rádio melhorou muito devido aos avanços nas pesquisas. A luz vista como forma de comunicação, por sua vez, é um elemento que conquistou a atenção dos acadêmicos há pouco mais de 10 anos [Dimitrov e Haas, 2015], e ainda é pouco explorado, quando comparado à faixa de radiofrequência do espectro eletromagnético.

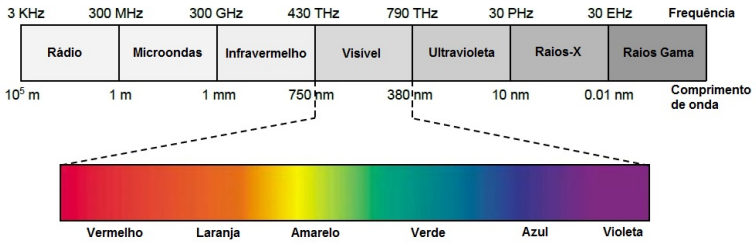


Figura 6.1: Espectro eletromagnético [Pathak et al., 2015]

A figura 6.1 apresenta o intervalo do espectro eletromagnético, desde as baixas frequências, onde estão localizadas as ondas de rádio, até as frequências mais altas, onde se situa a radiação gama. Dentre este intervalo, está o espectro visível da luz, com ondas que variam entre 390 nm e 700 nm. Qualquer informação que é transferida através da modulação das ondas de luz neste intervalo pode ser considerada um tipo de Comunicação por Luz Visível. É importante observar que a faixa destinada às ondas de rádio, do qual o WiFi faz parte, abrange frequências que vão de 3 KHz a 300 GHz. Em termos de frequência, o intervalo da luz visível abrange frequências que variam de 430 THz a 770 THz, o que é muito maior, se comparado ao de ondas de rádio.

A Comunicação por Luz Visível é uma tecnologia muito promissora, pois a implementação de sistemas VLC está intimamente ligada à rápida e crescente adoção de lâmpadas LEDs ao redor do mundo, assim como a aproximação iminente do paradigma

de iluminação inteligente [Sevincer et al., 2013]. A indústria tem reagido à nova onda de tecnologias de iluminação. Por exemplo, a *Philips* HUE[2], já comercializa uma lâmpada LED que pode ser controlada pelo *smartphone*. Dessa forma, no futuro, as lâmpadas LEDs irão desempenhar dois papeis diferentes: iluminação e comunicação.

6.1.2. História da Comunicação por Luz Visível

Os estudos envolvendo VLC não são um privilégio do século XXI. A luz sempre esteve presente entre os elementos utilizados pelo ser humano para se comunicar. Desde os tempos remotos, a utilização da luz como meio de comunicação já era presente em diversas culturas ao redor do mundo, seja na utilização de sinais de fumaça ou no uso de tochas

No final do século XVIII, na França Napoleônica, o engenheiro Charles Chappe inventou o telégrafo [Dilhac, 2001]. Este mecanismo consistia em duas barras laterais, chamadas indicadoras, anexadas em uma barra longa, denominada reguladora, como pode ser visto na figura 6.2. Através da rotação das hastes laterais, era possível criar uma série de símbolos diferentes. Estes equipamentos eram colocados em torres, em uma distância média de 10 a 15 quilômetros. Com essa estrutura e uma codificação eficiente, era possível criar até 98 combinações diferentes, que podiam ser enxergadas a quilômetros de distância, com o auxílio de um par de telescópios equipados nas torres. Em poucas décadas, a França já estava equipada com centenas de telégrafos, formando uma grande rede de comunicação, que serviu aos interesses franceses por mais de 50 anos, sendo substituído posteriormente pelo telégrafo elétrico.

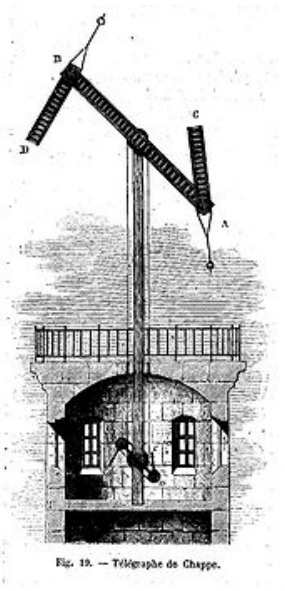


Figura 6.2: Telégrafo [Dilhac, 2001]

Mais tarde, no final do século XIX, outra invenção se destacou no estudo de Comunicação por Luz Visível. Em 1880, Alexander Graham Bell e seu assistente, Charles Tainter, se comunicaram em uma distância de 213 metros, utilizando o Fotofone [Bell, 1880]. Este dispositivo, criado pelo próprio Graham Bell, era formado por um emissor e um receptor, como pode ser visto na figura 6.3. De maneira sucinta, o funcionamento deste sistema se dava da seguinte forma: a luz solar era refletida em um espelho, atingindo uma superfície fina de vidro, que vibrava de acordo com a voz da pessoa. Após isso, a luz era transportada através de uma segunda lente para o receptor, onde um espelho parabólico refletia a luz em uma célula de selênio, cuja resistência variava de acordo com a intensidade da luz recebida. Apesar da popularidade do telefone, outro dispositivo patenteado por Graham Bell, o cientista sempre considerou o fotofone sua maior invenção.

Os estudos envolvendo meios óticos só ganharam força a partir da década de 1970, quando estudos demonstraram o potencial de uma comunicação ótica sem fio (no caso, infravermelho(IR)) em um ambiente interno, onde havia a possibilidade de se explorar bandas do espectro eletromagnético na escala de THz [Gfeller e Bapst, 1979], e o sistema

[2]Your personal Wireless lighting system - `www2.meethue.com/en-us/about-hue/`, 2017.

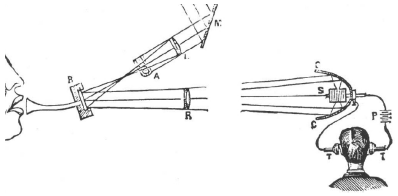


Figura 6.3: Fotofone criado por Alexander Graham Bell [Bell, 1880]

construindo chegava a alcançar taxas de 1 Mbps. Outro trabalho que explorou o potencial do infravermelho em ambientes internos foi realizado no final da década de 1990, em que o sistema implementado atingiu taxas de 50 Mbps [Marsh e Kahn, 1996].

No início dos anos 2000, lâmpadas LEDs passaram a ser consideradas para experimentos envolvendo VLC. Tanaka *et al.* foram os primeiros a utilizar uma lâmpada LED branca para iluminação e comunicação em um ambiente interno, em 2003, atingindo taxas de 400 Mbps[Tanaka et al., 2003]. Este foi o primeiro passo para uma vasta gama de trabalhos envolvendo VLC realizados no século XXI. Após este, muitos outros vieram com grandes inovações, como novas modulações, novas tecnologias de lâmpadas LEDs.

Um marco importante na história da Comunicação por Luz Visível foi realizado no ano de 2011, quando Harald Haas fez a primeira demonstração do Li-Fi (*Light Fidelity*), durante um TED Talk[3]. Essa apresentação tornou-se muito popular, alcançando milhões de visualizações em alguns meses. Evidentemente, a comunidade acadêmica reagiu à novidade, e, como consequência, a quantidade de pesquisas na área cresceu consideravelmente, como pode ser visto na figura 6.4, que apresenta a quantidade de trabalhos com a palavra-chave *"Visible Light Communication"* na plataforma *IEEE Xplore* ao longo dos anos.

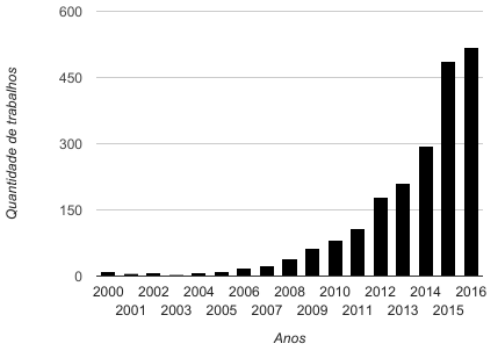


Figura 6.4: Quantidade de trabalhos envolvendo VLC na última década.

Atualmente, VLC é amplamente estudada. Grandes figuras da indústria e da academia, como a Nasa [Boroson et al., 2012, Luzhanskiy et al., 2016], Disney [Schmid et al., 2013] e Philips[4] possuem produtos e pesquisas na área.

6.1.3. Importância do estudo de VLC

O interesse em torno de tecnologias envolvendo VLC aumentou consideravelmente nos últimos anos. A importância dos estudos da luz como meio de comunicação tem se tornado cada dia mais evidente. Movimentos de órgãos de padronização internacionais e

[3]Wireless data from every light bulb - `https://www.ted.com/talks/harald_haas_wireless_data_from_every_light_bulb`, 2017.

[4]Perfect light,precise location - `http://www.lighting.philips.com/main/systems/themes/led-based-indoor-positioning.html`, 2016.

corporações importantes demonstram o interesse cada vez maior nesta tecnologia.

O aumento exponencial de demanda por dispositivos móveis ao redor do mundo tornou-se um grande incentivo para o estudo de VLC como complemento das tecnologias WiFi [Arnon, 2015], principalmente em ambientes internos, onde a luz pode ser controlada e há menor interferência de fatores externos. Como a maioria das pessoas tem a tendência a ficar a maior parte do tempo em ambientes fechados, a implementação de tecnologias VLC seria mais simples, utilizando-se de uma infraestrutura já existente.

Por mais que tenha havido um esforço no estudo de VLC nas últimas décadas, um dos fatores que limitava consideravelmente as pesquisas era a tecnologia utilizada nas lâmpadas comerciais, em sua maioria incandescentes ou CFL. A popularização de lâmpadas LEDs na última década ofereceu uma nova oportunidade para as pesquisas. Atualmente, diversos dispositivos utilizam LEDs, além das lâmpadas convencionais. Monitores, *smartphones* e automóveis estão entre os dispositivos equipados com LEDs. Com o avanço da tecnologia e popularização de preços de lâmpadas LEDs, grande parte das pesquisas passou a focar na utilização desses tipos de lâmpadas como emissores, e até mesmo como receptores [Schmid et al., 2013, Wang e Chi, 2014, Tsonev et al., 2013].

A possibilidade de altíssimas taxas de dados proporcionada pelas frequências, na ordem de THz, é outra grande motivação para estudos na área. Atualmente, a tecnologia WiFi alcança, no máximo, taxas de até 300 Mbps. Sistemas que utilizam a Comunicação Ótica sem Fio já alcançaram taxas de até 42.8 Gbps[5].

O espectro da luz visível traz uma série de oportunidades para estudos envolvendo VLC. Ao contrário da radiofrequência, o espectro da luz visível não é licenciado, ou seja, dispositivos podem transmitir em qualquer frequência sem a necessidade de uma licença [Jovicic et al., 2013]. Enquanto no espectro de ondas de rádio a frequência varia de KHz a GHz, no espectro visível a frequência é na grandeza de THz, ou seja, 1.000 vezes maior. Além disso, ao contrário do infravermelho e ultravioleta, que podem apresentar perigo para a saúde humana, o espectro da luz visível não apresenta nenhum perigo. Por fim, as ondas de rádio podem atravessar obstáculos como paredes, de modo que uma rede sem fio convencional pode ser interceptada por invasores a qualquer momento. Em sistemas VLC, o que se vê é o que se transmite. Em outras palavras, devido às propriedades da luz, um sistema VLC em locais fechados se torna muito mais seguro [Classen et al., 2016].

De maneira geral, o cenário VLC atual é muito rico e amplo, com oportunidades de estudo nas mais diversas áreas, desde aplicações como *LightId*[6] até métodos de modulação variados considerando propriedades particulares da luz [Monteiro e Hranilovic, 2014].

6.2. Visão Geral

Esta seção apresenta uma visão geral sobre VLC. Inicialmente, será feito um estudo acerca das principais tecnologias que se relacionam com VLC de alguma forma, entre

[5] `https://www.tue.nl/en/university/news-and-press/news/17-03-2017-wi-fi-on-rays-of-light-100-times-faster-and-never-overloaded/#top`, 2017.

[6] Light ID Technology - `http://www.panasonic.com/global/corporate/technology-design/technology/lightid.html`, 2016.

elas *Light Fidelity* (Li-Fi), *Optical Wireless Communication*, *Free Space Optical Communication*. É importante destacar as semelhanças e diferenças entre estes tipos de comunicação. Além disso, serão abordados os principais trabalhos responsáveis por alavancar os estudos na área [Heydariaan et al., 2016, Yin e Gnawali, 2016, Pathak et al., 2015, Schmid et al., 2013] e impulsionar os interesses para a comercialização desta tecnologia em setores diferentes [Tsonev et al., 2013]. Também será discutido a relação do crescente interesse na área de VLC e a popularização dos LEDs (*Light Emitting Diodes*). Neste sentido, serão apresentados os principais tipos de LED, e sua importância para a área. Após isto, pretende-se apontar uma série de vantagens e desvantagens que sistemas VLC possuem, quando comparados aos sistemas de radiofrequência. Por fim, serão abordadas as principais tendências acadêmicas e comerciais.

6.2.1. VLC e outras nomenclaturas

Assim como ocorre com a frequência de rádio, há uma série de nomenclaturas dadas às diferentes tecnologias que envolvem o estudo da luz como forma de comunicação. Algumas das principais nomenclaturas serão detalhadas abaixo:

Optical Wireless Communications (OWC): A comunicação ótica sem fio envolve qualquer tipo de transferência de dados em que o meio utilizado é o meio ótico. Em outras palavras, todo o espectro da luz pode ser utilizado como forma de comunicação, seja ele infravermelho, visível ou ultravioleta [Uysal e Nouri, 2014].

Free-Space Optical Communication: Apesar de ter um conceito similar à OWC, esta nomenclatura tem sido muito utilizada para transmissões de grande escala, como comunicações entre satélites e torres na Terra [Chan, 2006]. Uma comunicação no espaço livre envolve o envio de dados em meios sem barreiras, como o ar, atmosfera e o espaço. Aplicações que utilizam tal nomenclatura tendem a ser muito complexas, lidando com turbulências atmosféricas [Zhu e Kahn, 2002, Hou et al., 2015] e equipamentos de alto custo [Khalighi e Uysal, 2014].

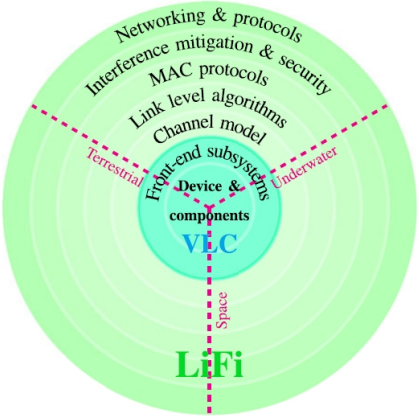


Figura 6.5: Diferenças entre Li-Fi e VLC [Haas et al., 2016]

Visible Light Communication (VLC): A utilização de comunicação ótica sem fio se tornou muito popular nos últimos anos. Em especial, estudos envolvendo o espectro da luz visível são cada vez mais comuns, visto que esta área tem um grande potencial acadêmico e comercial [Arnon, 2015]. A Comunicação por Luz Visível engloba todas as frequências do espectro visível da luz, ou seja, ondas que vão de 430 THz a 790 THz [Pathak et al., 2015].

Light Fidelity (LiFi): O termo Li-Fi foi cunhado em 2011, durante um TED *Talk*, como dito anteriormente, onde o professor Harald Haas fez uma demonstração prática dos potenciais da tecnologia. Pode-se dizer que o Li-Fi é um tipo de VLC [Tsonev et al., 2013]. Entretanto, o próprio criador do termo publicou, em 2015, um trabalho onde destaca as

principais diferenças entre VLC e Li-Fi [Haas et al., 2016]. Entre as diferenças entre as duas tecnologias, pode-se destacar comunicação multi-usuário bidirecional e a alta velocidade, aspectos presentes no conceito de Li-Fi. A figura 6.5 apresenta as principais técnicas necessárias para se alcançar o Li-Fi.

6.2.2. LEDs: uma grande oportunidade para VLC

Vários fatores contribuíram para o crescente interesse em VLC. Dentre todos, o que mais se destaca é a utilização de lâmpadas LEDs para manipulação das ondas de luz. Devido à suas características, como preço, por exemplo, as lâmpadas LEDs tornaram-se o principal meio utilizado para comunicação por luz visível. Além disso, as lâmpadas LEDs tornam-se cada vez mais populares, integrando diversos ambientes em que seria vantajoso utilizar a luz como forma de comunicação. Por isso, torna-se cada vez mais comum a escolha deste tipo de lâmpada em sistemas VLC [Karunatilaka et al., 2015].

As vantagens de lâmpadas LEDs em comparação com outros tipos são inúmeras. LEDs são muito eficientes em termos de consumo de energia, reduzindo em até 75% o consumo de energia elétrica, quando comparado à lampadas incandescentes, por exemplo, além de durarem até 25 vezes mais tempo[7]. Lâmpadas LEDs são muito mais seguras, pois esquentam menos, além de não conterem substâncias nocivas à saúde, como mercúrio. Por fim, a eficiência de luminosidade das lâmpadas LEDs é muito maior, quando comparada a outros tipos de lâmpadas. Todas essas vantagens contribuem para um ambiente em que as lâmpadas LEDs passam a oferecer mais do que apenas iluminação básica.

Existem diversos tipos de LEDs, cada um com suas particularidades, o que os torna apropriados para aplicações VLC. Dependendo do composto utilizado na fabricação do chip, a luz é emitida em uma região específica do espectro visível, ou seja, o fóton será emitido com um comprimento de onda característico, resultando em uma cor. Arseneto de gálio (GaAs), Fosfeto de Galio (GaP), são exemplos de compostos utilizados em LEDs. Os principais tipos de LEDs são listados abaixo, junto com seus respectivos detalhes.

Phosphor Converted LEDs (pc-LEDs): Os LEDs do tipo pc-LED são amplamente utilizados, de baixa complexidade e baixo custo. Consistem de um chip de LED azul revestido com uma camada de fósforo, cuja função é converter parte da luz azul em verde, amarelo e vermelho, enquanto uma fração da luz azul é emitida, resultando na luz branca. Esse tipo de LED tem uma banda limitada, devido à resposta lenta do fósforo.

Multi-chip LEDs: São LEDs cuja estrutura consiste em três ou mais *chips* e que emitem luzes de cores diferentes. Normalmente, os diferentes *chips* emitem as cores RGB, a fim de produzir a luz branca. A grande vantagem deste tipo de LED é a capacidade de controlar as cores que são emitidas, através da intensidade de cada *chip*. É importante ressaltar que uma modulação foi criada especialmente para este tipo de LED, chamada *Color-Shift Keying*. A seção 6.4 abordará mais detalhes deste tipo de modulação.

LEDs Orgânicos (OLEDs): Este tipo de LED consiste de uma série de filmes orgânicos finos entre dois condutores. Quando a corrente elétrica é aplicada, a luz é emitida. São muito utilizados em *displays* de *smartphones*. A grande vantagem deste tipo de tecnologia é a possibilidade de construir dispositivos transparentes e flexíveis.

[7]LED Lighting - `https://www.energy.gov/energysaver/led-lighting`, 2016.

Entretanto, tanto em termos de frequência quanto em termos de duração, este tipo de LED ainda é ineficiente quando comparado aos outros tipos [Karunatilaka et al., 2015].

μ-LEDs: Os μ-LEDs são normalmente acoplados em *displays*, possibilitando uma comunicação paralela de alta densidade, atingindo velocidades muito altas.

6.2.3. VLC *vs* RF: vantagens e desvantagens

Nas últimas décadas, o mundo viveu um grande avanço nas tecnologias de comunicação. Em termos de comunicação sem fio, o WiFi se consolidou como o principal meio de acesso à Internet. Entretanto, fatores, como crise no espectro WiFi, necessidade de comunicações em alta velocidade, impulsionam novas tecnologias e pesquisas. Nesse cenário, estudos em VLC focam, em sua maioria, em sua utilização como complemento do WiFi, de forma a suprir as necessidades da atual demanda de conteúdo em redes sem fio [Ayyash et al., 2016]. Além desses motivadores, alguns autores propõem o uso da tecnologia oferecida pelo uso da luz como meio de comunicação com o objetivo de substituir o Wi-Fi em alguns cenários [Dimitrov e Haas, 2015]. Em ambos os casos, deve-se considerar as vantagens e desvantagens que o VLC oferece, quando comparado ao WiFi. A tabela 6.1 apresenta uma comparação geral entre frequência de rádio e VLC.

Uma grande vantagem do VLC é a utilização de uma infraestrutura já existente para implementar a forma de comunicação. As lâmpadas LEDs, amplamente comercializadas nos dias de hoje, já desempenham o papel de iluminação. Com o VLC, tais lâmpadas passam a transmitir dados utilizando-se dessa iluminação já existente. Ou seja, a energia utilizada para a comunicação não aumentaria os custos [Burchardt et al., 2014]. Além disso, muitas das pesquisas dos últimos anos tem seu foco na utilização de dispositivos de baixo custo na implementação de sistemas VLC, como é o caso de Wang *et al.* , que utilizaram micro-computadores (*Beaglebone*) e LEDs de baixo custo para o desenvolvimento de uma plataforma *open-source* para estudos direcionados na área [Wang et al., 2015b]. Outro exemplo importante na literatura é o trabalho de pesquisadores na *Disney Research Center*, responsáveis pelo desenvolvimento de um sistema VLC que faz uso de lâmpadas LEDs presentes no mercado [Schmid et al., 2014].

Uma vantagem da luz visível é o tamanho do espectro disponível, se comparado à frequência de rádio. A divisão das frequências na faixa do espectro eletromagnético destinada às ondas de rádio é extremamente restrita, sendo regulada por cada país, e coordenadas por órgãos internacionais de telecomunicação. Dessa forma, cada país possui seu próprio regulamento quanto às frequências alocadas para cada tipo de uso, que vai desde uso militar, até transmissões de conteúdo em rádios AM e FM. Por ser uma tecnologia relativamente nova, os dispositivos WiFi transmitem o sinal em duas faixas: 2,4 GHz e 5 GHz, ambas localizadas em regiões do espectro destinadas à dispositivos não licenciados. Entretanto, esta situação não ocorre com a luz. O espectro da luz visível é totalmente livre, gerando diversas possibilidades comerciais e acadêmicas [Burchardt et al., 2014].

Devido às suas propriedades de propagação, a luz passa a oferecer vantagens em termos de segurança, quando comparada às ondas de rádio. Quando um ponto de acesso WiFi é configurado, as ondas de rádio podem se propagar de acordo com a capacidade de emissão da antena, podendo chegar a centenas de metros. Neste processo, as ondas ultrapassam paredes e outras superfícies sólidas, podendo representar um risco para a segu-

rança, visto que tentativas de *eavesdropping* e *sniffing* podem ocorrer [Burchardt et al., 2014]. A luz, por sua vez, não segue este comportamento. Suas ondas não ultrapassam paredes e outras superfícies, oferecendo um ambiente muito mais seguro, onde basicamente o que está sendo transmitido é o que se vê [Rohner et al., 2015]. Essa possibilidade de manipulação das ondas de luz constitui outra grande vantagem desta forma de comunicação.

Por último, uma das maiores vantagens da luz como forma de comunicação é a frequência alta das ondas (na grandeza de THz), fato que possibilita velocidades altíssimas. Atualmente, em termos de WiFi, a maior velocidade alcançada é próxima de 1 Gbps, no padrão WiGig [Hansen, 2011]. Graças à alta frequência das ondas de luz, pesquisas com VLC já obtiveram resultados extremamente expressivos, atingindo velocidades superiores a 100 Gbps [Azhar et al., 2013, Gomez et al., 2015].

	Radiofrequência	VLC
Espectro	$\sim$ 300 GHz	$\sim$ 400 THz
Infraestrutura	Ponto de acesso	Iluminação
Interferência e Ruídos	Baixa	Alta
Segurança	Limitada	Alta
Cobertura	Alta	Limitada
Complexidade do sistema	Alta	Baixa
Interferência Eletromagnética	Sim	Não

Tabela 6.1: Comparação entre radiofrequência e VLC, adaptado de [Karunatilaka et al., 2015].

6.2.4. Tendências na área

Com o crescimento do interesse em sistemas VLC, novas oportunidades e tecnologias surgiram nos últimos anos. É importante ressaltar que paralelo a essa nova tendência, há muitos estudos que buscam aumentar o desempenho do WiFi em redes *wireless* presentes nos dias de hoje, como WLANs e 4g. Dessa forma, as tendências e o futuro da área são promissores, contemplando trabalhos que tratam de VLC de forma individual, e aqueles que integram esta tecnologia à outras existentes, como WiFi.

Muitos trabalhos recentes focam no desenvolvimento de sistemas híbridos, em que a arquitetura de rede integra ambos WiFi e VLC. Neste âmbito, trabalhos como PLiFi tem surgido e chamado atenção da comunidade acadêmica [Hu et al., 2016]. O PLiFi é um sistema híbrido, cuja arquitetura une as tecnologias WiFi e VLC, através do uso de PLC (*Power Line Communication*). A conexão entre as lâmpadas LEDs e a Internet é feita em alta velocidade, utilizando ferramentas acessíveis. Este trabalho oferece uma solução nova, cuja proposta visa mitigar alguns desafios conhecidos na área, como *Uplink* e mobilidade, cujos detalhes podem ser vistos na seção 6.6.

Além disso, há muito se discute acerca da implementação de VLC em futuras tecnologias sem fio, principalmente no que diz respeito ao 5G. *Smartphones, tablets* e o conceito cada vez mais sólido de IoT levanta uma série de questões em relação às futuras demandas de dispositivos sem fio, exigindo que a tecnologia atual evolua e se adapte para atender as atuais e futuras aplicações. Para isso, pesquisas nos últimos anos

apontam para o uso de VLC integradas ao 5g, próxima tecnologia celular [Wu et al., 2014, Ayyash et al., 2016]. Estes trabalhos encontram, em sua premissa, o fato de que a maioria do tráfego atual na Internet é feito em ambientes fechados, local em que há possibilidade de se utilizar de uma infraestrutura preexistente para a implementação de sistemas VLC.

Dispositivos como *Smartphones* têm se tornado cada vez mais indispensáveis na vida de grande parte da população mundial. Dessa forma, há um esforço por parte de algumas pesquisas em integrar a comunicação por luz visível aos *Smartphones*, utilizando componentes como a câmera e os LEDs externos [Duque et al., 2016]. Como os *smartphones* já são equipados com a câmera, o grande desafio é adaptar os dispositivos sem nenhuma modificação no *hardware*, a fim de integrá-los em sistemas de comunicação por luz visível [Schmid et al., 2016a].

VLC também se destaca em grandes projetos, como o *Lunar Laser Communication Demonstration*. A Nasa lançou, em 2013, uma espaçonave na missão denominada LADEE (*Lunar Atmosphere and Dust Environment Explorer*), com o objetivo de coletar dados da lua. Acoplado à espaçonave, estava o LLST (*Lunar Lasercom Space Terminal*), um terminal responsável por enviar os dados coletados para a Terra, através de Comunicação Ótica sem Fio [Boroson et al., 2012]. Taxas de transferência de 622 Mbps foram alcançadas, 6 vezes maior quando comparada às tecnologias de rádio utilizadas até então[8].

Por fim, com o avanço nas pesquisas e a inevitável comercialização de produtos e aplicações que utilizam VLC, a área acadêmica continua tendo interesse constante nessa forma de comunicação. Pesquisadores já desenvolverem plataformas de pesquisa, com o objetivo de padronizar pesquisas e levar essa tecnologia a mais universidades ao redor do mundo [Hewage et al., 2016]. Um exemplo dessa iniciativa é a plataforma *open-source* OpenVLC, criada em 2014, e atualmente está presente em mais de 15 grupos de pesquisa ao redor do mundo [Wang e Giustiniano, 2014a]. Os aspectos desta plataforma serão explicados com mais detalhes na seção 6.7.

6.3. Arquitetura de Comunicação

Os componentes principais que integram sistemas VLC –transmissor e receptor– geralmente são constituídos de 3 camadas comuns a eles [Khan, 2016]. As camadas física, enlace e de aplicação compõem um modelo de referência para a arquitetura destes sistemas [Schmid et al., 2013]. Assim, esta seção tem 2 objetivos principais. Primeiro, dar uma visão geral das camadas que compõem sistemas VLC. Cada uma destas camadas será melhor estudada em futuras seções. Segundo, discutir as topologias de redes suportadas e os desafios encontrados, como mobilidade [Khan, 2016]. Mais ainda, discutiremos os diferentes dispositivos apontados pela literatura, e como eles se adequam a propósitos específicos, como por exemplo, uma topologia de redes criada em um ambiente interno, onde busca-se uma infraestrutura próxima ao WiFi [Schmid et al., 2016b].

6.3.1. Visão geral da arquitetura VLC

Como dito nas seções anteriores, a Comunicação por Luz Visível faz uso da luz para transmitir informações. Além disso, a proposta de sistemas VLC é fornecer ilumina-

[8]NASA | LLCD: Proving Laser Communication Possible - `https://www.youtube.com/watch?v=wJMPd2FJp5g`, 2017.

ção e comunicação ao mesmo tempo. Dessa forma, sistemas VLC sempre terão componentes para transmitir e receber luz. Na maioria dos trabalhos disponíveis na literatura, utilizam-se lâmpadas LEDs como transmissores. Nessas lâmpadas são feitas a modulação da intensidade (IM) para o envio de dados. Do outro lado, fotosensores são responsáveis por captar essa luz de maneira direta (*Direct Detection*), convertendo-a em corrente [Medina et al., 2015]. Em VLC, é importante que a iluminação não seja afetada pela manipulação da luz no envio de informações, daí a influência de cada tipo de lâmpadas LED no desempenho de um sistema VLC.

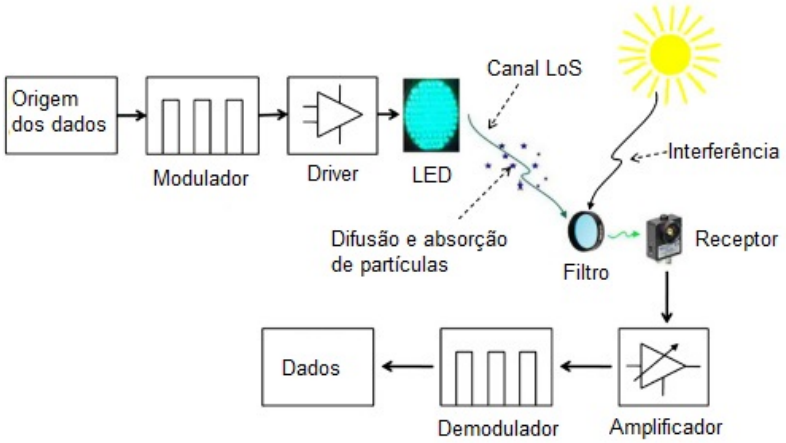


Figura 6.6: Arquitetura de um sistema VLC, adaptado de [Cui et al., 2012]

A figura 6.6 apresenta uma visão geral da arquitetura de um sistema VLC. As lâmpadas LEDs transmitem os dados através da modulação da intensidade (*Intensity Modulation*). O receptor deve estar na linha de visão do LED, para que receba os raios de luz contendo a informação. De fato, durante a transmissão da luz, haverá perda devido à difusão das partículas e à interferência inerente da luz do ambiente. Para diminuir a interferência, podem ser utilizados filtros. O receptor é atingido pelos raios de luz, alterando diretamente a corrente. Amplificadores são utilizados para que os sinais sejam menos propensos a interferências e ruídos [Schmid et al., 2014]. Por fim, para se obter a informação original, ocorre a demodulação. Abaixo seguem mais detalhes sobre os principais componentes de um sistema VLC.

Transmissores: Em geral, são utilizadas lâmpadas LEDs para sistemas VLC. Uma lâmpada LED pode conter vários LEDs, como é o caso de diversas lâmpadas comercializadas atualmente. Estas lâmpadas contêm um *driver* responsável por controlar a corrente que passa pelos LEDs, influenciando diretamente na intensidade da iluminação. Em outras palavras, a corrente que chega ao LED é controlada através de *transistores*, que manipulam os sinais de luz que o LED emite em alta frequência, e assim, torna a comunicação imperceptível aos olhos humanos [Pathak et al., 2015].

Receptores: Os receptores são responsáveis por captar a luz e convertê-la em corrente elétrica. Normalmente, são utilizados fotodiodos como receptores em sistemas de Comunicação por Luz Visível [Schmid et al., 2014]. Entretanto, fotodiodos são extremamente sensíveis, e captam ondas além do espectro da luz visível, como ultravioleta e infravermelha [Wang et al., 2015b]. Dessa forma, em um ambiente externo e exposto à

luz solar, por exemplo, o fotodiodo falharia em receber os dados devido à alta interferência do ambiente externo. Para isso, outros componentes podem ser utilizados para captar a luz. Um deles é a própria câmera dos *Smartphones*, que permite que qualquer celular passe a receber dados em sistemas VLC, como foi discutido na seção 6.2. Além desses dispositivos, os próprios LEDs podem ser utilizados como receptores pois apresentam características de fotosensores[9] [Wang et al., 2014]. A figura 6.7 mostra o diagrama com os componentes do OpenVLC, onde o mesmo LED pode ser utilizado como transmissor (TX) e receptor (RX).

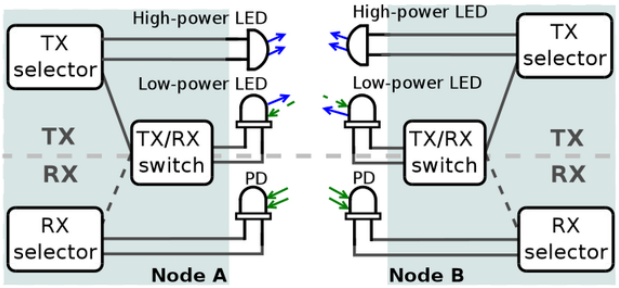


Figura 6.7: Diagrama do OpenVLC, onde se utiliza LEDs como transmissores e receptores [Wang et al., 2015b].

Ao contrário dos fotodiodos, os LEDs têm propriedades que os tornam eficientes em certas situações. Um LED detecta uma faixa de frequência reduzida, quando comparado a fotodiodos, reduzindo a presença de ruídos e interferências. Além disso, a sensibilidade de LEDs é estável, com o passar do tempo. A principal vantagem é o fato de LEDs se comportarem tanto como trasmissores quanto como receptores, o que possibilita a criação de um sistema com apenas um LED em cada ponto[10], além de serem componentes muito acessíveis e populares, facilitando ainda mais seu uso em aplicações VLC.

6.3.2. Modelos de referência

Devido ao crescente interesse em sistemas de Comunicação por Luz Visível por parte de universidades e indústria, houve a necessidade de se padronizar certos aspectos desse tipo de comunicação. Para isso, o comitê do IEEE (*Institute of Electrical and Electronics Engineers*) aprovou, em 2011, a criação do padrão **IEEE 802.15.7**, onde foram definidas as camadas Física e MAC para Comunicação ótica sem fio de curta distância utilizando luz visível [IEE, 2011]. O padrão cobre aspectos necessários para garantir a entrega de dados com taxas suficientes para suportar serviços como multimídia e áudio, além de garantir a compatibilidade com a infraestrutura de luz visível. Além disso, o padrão ainda cobre efeitos na saúde e no meio ambiente. De maneira geral, o padrão aborda questões como topologias de rede, dispositivos considerados para VLC, arquitetura de comunicação, características da camada física e MAC com suporte a *dimming* e *flickering*, assim como especificações de segurança. Abaixo, são apresentados detalhes desses aspectos.

Inicialmente, o documento aborda os tipos de dispositivos em sistemas VLC, entre

[9]LED Sensing - `www.thebox.myzen.co.uk/Workshop/LED_Sensing.html`, 2017.

[10]How to Use LEDs to Detect Light - `http://makezine.com/projects/make-36-boards/how-to-use-leds-to-detect-light/`, 2013.

eles infraestrutura, móveis e veículos, cada um com suas características específicas, como pode ser visto na tabela 6.2. Em seguida, topologias e especificações das modulações são propostas nesse padrão.

	Infraestrutura	Móvel	Veículo
Coordenador fixo	Sim	Não	Não
Fonte de energia	Ampla	Limitada	Moderada
Fonte de luz	Intensa	Fraca	Intensa
Mobilidade Física	Não	Sim	Sim
Alcance	Curto/Longo	Curto	Longo
Taxa de dados	Alta/baixa	Alta	Baixa

Tabela 6.2: Estruturas propostas no IEEE 802.15.7, adaptado de [IEE, 2011]

Grande parte do IEEE 802.15.7 é focado em características das camadas física e MAC, e serão vistas com mais detalhes na seção 6.4. De maneira geral, o padrão divide a camada física em três modos de operação: PHY I, PHY II e PHY III. Qualquer sistema compatível com o IEEE 802.15.7 deve implementar pelo menos os modos PHY I ou PHY II. O sistema que implementar o modo PHY III, também deve implementar o PHY II.

O modo de operação PHY I foi desenvolvido para aplicações externas, com *frames* curtos. Já os modos PHY II e PHY III suportam apenas um tipo de codificação.As taxas de dados do modo PHY I variam de 11 kbps à 266 Kbps, enquanto as do modo PHY II vão de 1,25 Mbps à 96 Mbps. O modo de operação PHY III contempla taxas de 12 Mbps até 96 Mbps. É importante ressaltar que o modo de operação PHY III possui uma modulação particular, desenvolvida para lâmpadas LED *multi-chips*, que será vista com mais detalhes na seção 6.4. As tabelas 6.3,6.4 e 6.5 apresentam detalhes de cada modo de operação, assim como modulações e codificações suportadas por cada uma.

Modulation	RLL code	Optical clock rate	FEC		Data rate
			Outer code (RS)	Inner code (CC)	
OOK	Manchester	200 kHz	(15,7)	1/4	11.67 kb/s
			(15,11)	1/3	24.44 kb/s
			(15,11)	2/3	48.89 kb/s
			(15,11)	none	73.3 kb/s
			none	none	100 kb/s
VPPM	4B6B	400 kHz	(15,2)	none	35.56 kb/s
			(15,4)	none	71.11 kb/s
			(15,7)	none	124.4 kb/s
			none	none	266.6 kb/s

Tabela 6.3: Modo de operação PHY I [IEE, 2011]

Além disso, conceitos como *dimming* e *flickering* são abordados com detalhes, pois um sistema VLC deve permitir que haja manipulação da intensidade da luz de maneira que não influencie na comunicação em si.

Modulation	RLL code	Optical clock rate	FEC	Data rate
VPPM	4B6B	3.75 MHz	RS(64,32)	1.25 Mb/s
			RS(160,128)	2 Mb/s
		7.5 MHz	RS(64,32)	2.5 Mb/s
			RS(160,128)	4 Mb/s
			none	5 Mb/s
OOK	8B10B	15 MHz	RS(64,32)	6 Mb/s
			RS(160,128)	9.6 Mb/s
		30 MHz	RS(64,32)	12 Mb/s
			RS(160,128)	19.2 Mb/s
		60 MHz	RS(64,32)	24 Mb/s
			RS(160,128)	38.4 Mb/s
		120 MHz	RS(64,32)	48 Mb/s
			RS(160,128)	76.8 Mb/s
			none	96 Mb/s

Tabela 6.4: Modo de operação PHY II [IEE, 2011]

Modulation	Optical clock rate	FEC	Data rate
4-CSK	12 MHz	RS(64,32)	12 Mb/s
8-CSK		RS(64,32)	18 Mb/s
4-CSK	24 MHz	RS(64,32)	24 Mb/s
8-CSK		RS(64,32)	36 Mb/s
16-CSK		RS(64,32)	48 Mb/s
8-CSK		none	72 Mb/s
16-CSK		none	96 Mb/s

Tabela 6.5: Modo de operação PHY III [IEE, 2011]

O documento também aborda as questões de segurança em VLC. Neste sentido, a luz possui propriedades diferentes das ondas de rádio, possibilitando diretrizes novas ao tratar a segurança de sistemas VLC. Como a onda de luz é direcionada e visível, caso um receptor não autorizado intercepte o sinal, sua presença pode ser detectada. Mesmo assim, o mecanismo de criptografia proposto é baseado em chaves simétricas, geradas pelas camadas superiores. Entre os serviços de segurança oferecidos pelo mecanismo de criptografia, estão confidencialidade, autenticidade e *replay protection*.

6.3.3. Topologias de rede

Em uma rede de computadores, a topologia de rede define a forma como os dispositivos são organizados e distribuídos e, como a informação será compartilhada entre eles [Tanenbaum et al., 2003]. Isso é feito tanto do ponto de vista lógico quanto físico. Topologias como Ponto-a-Ponto, barramento e anel são bem conhecidas no meio acadêmico e muito utilizadas em sistemas de redes de computadores.

A ideia é análoga em sistemas VLC. Ainda de acordo com o documento de padronização oficial de VLC [IEE, 2011], há três classes de dispositivos considerados para sistemas VLC: infraestruturas, móveis e veículos (tabela 6.2). Dessa forma, o IEEE 802.15.7 define as aplicações em três topologias: *peer-to-peer*, estrela e *broadcast*, como pode ser visto na figura 6.8. Em VPANs (*Visible Light Personal Area Networks*), todos os dispositivos possuem um endereço único de 64 bits.

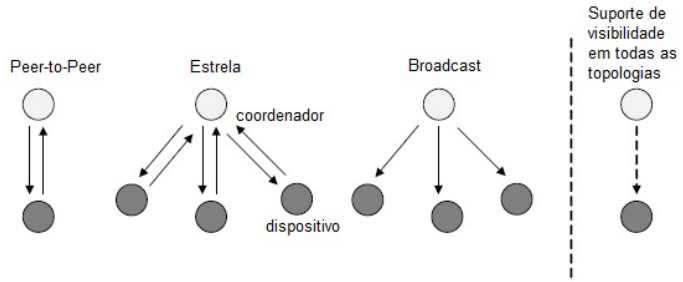


Figura 6.8: Topologias de rede de sistemas VLC de curto alcance [IEE, 2011].

Peer-to-Peer: Em uma topologia peer-to-peer, um dispositivo pode se comunicar com qualquer outro em sua área de cobertura. Além disso, um dos dispositivos conectados deve assumir o papel de coordenador, e isso pode ser feito quando ele assume esse papel por ser o primeiro dispositivo a se comunicar no canal, por exemplo.

Estrela: Em uma topologia de estrela, há um controlador central, chamado coordenador, e a comunicação é estabelecida entre ele e cada dispositivo na rede. É importante ressaltar que toda rede na topologia estrela funciona independente de outras redes em operação, porque existe um identificador associado a uma única rede estrela.

Broadcast: Nesta topologia, um dispositivo pode enviar a informação para outros sem que uma rede seja formada. A comunicação, neste caso, é unidirecional.

Baseado nas topologias descritas anteriormente, o IEEE 802.15.7 propõe uma série de técnicas de modulação na camada física, assim como protocolos específicos na camada de enlace, que serão vistos com detalhes na seção 6.4.

6.4. Mecanismos de codificação

Diversos trabalhos referem apenas à troca de dados utilizando a luz como meio quando tratam VLC. Entretanto, muitas pesquisas atuais envolvem Comunicação por Luz Visível integrada à Internet. Mais ainda, já há definições e modelos de referência acerca da arquitetura VLC [Khan, 2016, Schmid et al., 2013]. Assim, esta seção detalha as camadas física e enlace em sistemas VLC, introduzidas na seção anterior. Inicialmente, como mo-

tivação para esta seção, será feito um levantamento teórico sobre trabalhos disponíveis na literatura que exploram o VLC integrado à Internet. A comunicação por luz visível é análoga ao WiFi, pois deve-se implementar mecanismos e protocolos nas camadas física e de enlace para que haja transferência de dados entre emissor e receptor de forma efetiva. Discutiremos os principais problemas a serem tratados na camada de enlace, como suporte à mobilidade, segurança e visibilidade. A seguir, serão discutidas as principais técnicas utilizadas para codificação e modulação da luz, como *Manchester Encoding*, (*Color Shift Keying*), OOK (*On Off Keying*) e OFDM (*Orthogonal Frequency Division Multiplexing*), implementadas na camada física, assim como os protocolos de múltiplo acesso ao meio implementados na camada de Enlace (CSMA/CA, CSMA/CD, CSMA/CD-HA).

6.4.1. Internet e VLC

Nos últimos anos, a maioria dos trabalhos realizados tem focado nos aspectos da comunicação dos LEDs com os respectivos sensores. Como dito na seção anterior, há um padrão para as camadas Física e Enlace em sistemas VLC, o IEEE 802.15.7 [IEE, 2011]. Algumas pesquisas já abordam uma arquitetura em que essas camadas são integradas com o restante da pilha de protocolos, de forma a habilitar acesso à Internet por meio de VLC.

Comercialmente, há sistemas propostos que envolvem acesso à Internet. E.g., o dispositivo LiFi-X, desenvolvido pela *pureLiFi*[11], evolução do antigo Li-Flame[12], primeiro sistema a utilizar a tecnologia LiFi, cunhada em 2011 pelo cientista Harald Haas. O sistema LiFi-X permite que se instale uma infraestrutura de acesso à Internet completamente baseada em LiFi. O produto oferece aspectos essenciais para comunicações sem fio, como mobilidade, múltiplos usuários e segurança, além de velocidades de 40 Mbps. O sistema é constituído de Pontos de Acesso e Estações. Por ser a primeira solução a utilizar a tecnologia LiFi, a empresa despertou o interesse de centenas de investidores.

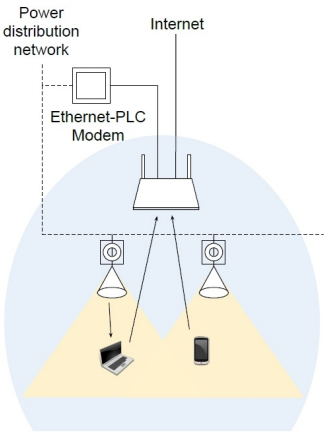


Figura 6.9: Arquitetura PLiFi [Hu et al., 2016]

Outras pesquisas na área tentam trazer soluções mais acessíveis, enquanto buscam maneiras de unir as camadas inferiores dos sistemas criados às superiores. A plataforma híbrida PLifi [Hu et al., 2016] foi criada com o objetivo de unir o WiFi a VLC, em um ambiente interno. Entre os desafios destacados pelo autor, encontra-se a conectividade dos LEDs com a Internet. Como pode ser visto na figura 6.9, na arquitetura do PLifi, o ponto de acesso WiFi utiliza a tecnologia *Power Line Communication* (PLC) utilizando um modem Ethernet-PLC. Por sua vez, a rede PLC se conecta às lâmpadas LEDs através de um modem PLC-VLC. Dessa forma, os pacotes são encaminhados para o LED através do ponto de acesso WiFi.

Com a popularização do conceito de Internet das Coisas, muitas pesquisas passa-

[11]LiFi-X - The fastest, smallest and most secure LiFi system - `http://purelifi.com/lifi-products/lifi-x/`, 2017.

[12]Li-Flame - `http://purelifi.com/lifi-products/li-flame/`, 2017.

ram a abordar VLC integrada a pilha de protocolos da Internet de forma direta, ou seja, sem necessidade de WiFi. Schmid *et al.* apresentaram, em 2015, um trabalho em que o sistema VLC criado utilizava lâmpadas LED comerciais [Schmid et al., 2015]. Na arquitetura proposta, uma lâmpada LED comercial era adaptada, tornando-se um transmissor com Linux e VLC integrados. Para isso, a lâmpada é modificada para conter um *System-on-a-Chip* (SoC) que executa Linux e o *driver* VLC, responsável pela modulação da luz. O *firmware* VLC implementa as camadas Física e MAC, a fim de disponibilizar a criação de rede entre múltiplos dispositivos. A figura 6.10 apresenta a arquitetura do sistema VLC. Neste caso, os elementos da lâmpada são: Módulo SoC com WiFi habilitado, interface de comunicação entre o módulo SoC e o microcontrolador (neste caso, a conexão é feita através da interface *Universal Asynchronous Receiver Transmitter*), microcontrolador contendo o *firmware* VLC, amplificadores, fotodiodos e por fim, o LED.

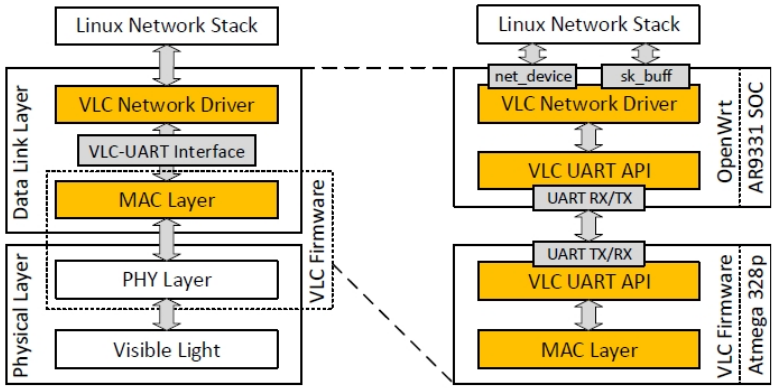


Figura 6.10: Arquitetura do sistema [Schmid et al., 2015].

Outro trabalho que se destaca pela proposta de um sistema integrado à Internet foi desenvolvido por Wang *et al.*. Os autores apresentam a plataforma OpenVLC [Wang et al., 2014] que, diferente do trabalho de Schmid, não utiliza lâmpadas de LED comerciais e SoC para implementação da integração VLC com Internet. O OpenVLC é uma plataforma desenvolvida para ser um periférico do conhecido *BeagleBone Board*, por isso ela recebe o nome de "capa"(*shield*). A capa pode ser inserida no *Beaglobone*, e toda a implementação das camadas inferiores é feita no próprio *BeagleBone*, cujo sistema operacional é o Debian.

Como pode ser observado, há um grande interesse tanto por parte da comunidade acadêmica quanto pela indústria no desenvolvimento de tecnologias VLC. Neste sentido, destacam-se os esforços em integrar VLC e WiFi, assim como aqueles trabalhos que buscam uma alternativa focada somente em VLC, integrado à Internet.

6.4.2. Camada Física

Em um modelo de referência padrão, como o OSI, a camada física é a responsável pela transmissão de dados (bits) através de um canal de comunicação, seja ele um fio de cobre trançado ou ondas de rádio. As questões que devem ser abordadas na camada física de um sistema envolvem a representação dos sinais 0 e 1, como estes sinais serão enviados, como a comunicação é estabelecida, e como é encerrada, envolvendo elementos de interface

eletrônica e sincronização [Tanenbaum et al., 2003]. Na Comunicação por Luz Visível, não é diferente. As camadas superiores enviam normalmente os dados para as camadas inferiores, e a camada física envolve todos os processos citados anteriormente, sendo a luz o meio de transmissão que está abaixo da camada física. Dessa forma, questões como MIMO (*Multiple-Input-Multiple-Output*), modulações e codificações serão vistas nesta seção, mas antes é necessário entender alguns fatores que influenciam as decisões acerca da implementação da camada física em sistemas VLC.

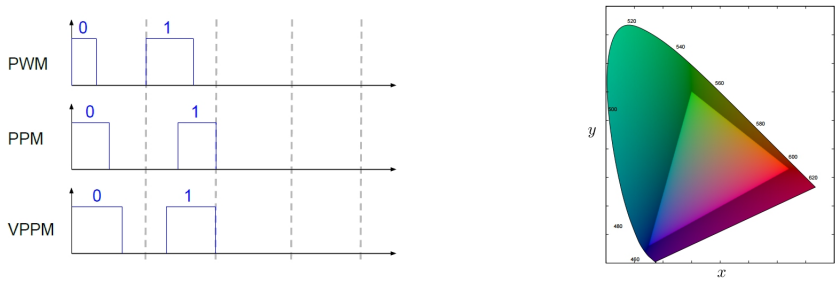
Fluxo luminoso e perda de caminho (*Path Loss*): Em termos de camada física em VLC, a primeira coisa que deve ser levada em consideração é o fato de lâmpadas LED terem duas principais funções: **iluminação e comunicação**. Portanto, é necessário entender os requisitos em termos de luminosidade para que a comunicação ocorra de maneira satisfatória. Os parâmetros fotométricos, como são chamados, determinam uma série de características da luz, como brilho, cor, entre outros, numa perspectiva de visão humana. Por outro lado, os parâmetros radiométricos medem a energia da radiação eletromagnética da luz. Através desses parâmetros, é possível calcular o fluxo luminoso, que representa a energia emitida por um LED. Baseado no fluxo luminoso, é possível calcular um valor importante para a camada física: a perda de caminho (*Path Loss*) [Cui et al., 2010].

Propagação: Outra propriedade que é de grande importância para a camada física é a propagação da onda de luz. Visto que em ambientes internos a tendência é que se tenha múltiplos transmissores (lâmpadas LEDs, por exemplo) e as superfícies podem refletir a luz emitida, é importante compreender o impacto da luz refletida em sistemas VLC.

Ruídos: Em um sistema VLC, ruídos são fatores de extrema importância para a eficiência da comunicação. Durante o dia, em ambientes externos, por exemplo, a luz solar pode gerar interferências, prejudicando a comunicação e, podendo até inibi-lá. Neste caso, pode-se utilizar filtros, com o objetivo de eliminar as ondas destrutivas. Outra opção levada em consideração em alguns trabalhos é a utilização de LEDs como receptores, visto que eles são "fotodiodos seletivos", como discutido na seção 6.3.

Tendo em vista os fatores detalhados acima, outro ponto essencial da camada física deve ser abordado: a modulação. Considerando que em uma Comunicação por Luz Visível há a transmissão de sinais analógicos baseados na intensidade da luz, tais sinais devem ser convertidos para sinais digitais, a fim de representar os *bits*. Para isso, é feita a modulação destes sinais [Tanenbaum et al., 2003]. Diferentemente de outros tipos de comunicação, a modulação da luz deve buscar uma alta taxa de dados ao mesmo tempo em que não interfere com a luz percebida pelo ser humano [Arnon, 2015]. Um destes requisitos é o escurecimento (*dimming*). As lâmpadas de diversos locais residenciais e corporativos são equipadas com circuitos de escurecimento, para que a intensidade da luz possa ser controlada. Neste sentido, de acordo com o padrão IEEE 802.15.7, uma Comunicação por Luz Visível deve suportar este escurecimento, sem que haja problemas na comunicação. O segundo requisito diz respeito à oscilação (*flickering*). A técnica de modulação aplicada não pode causar nenhum tipo de oscilação perceptível pela visão humana [Roberts et al., 2011]. Dito isso, abaixo são apresentados detalhes dos principais métodos de modulação existentes na literatura para sistemas VLC.

***On-Off Keying* (OOK):** A modulação OOK é a mais simples considerando o funcionamento de lâmpadas LEDs. Neste tipo de modulação, os bits 0 e 1 são transmitidos



(a) (b)

Figura 6.11: (a) Comparação entre métodos de modulação por pulso, adaptado de [Pathak et al., 2015] (b) Diagrama de cromaticidade CIE 1931 [Monteiro e Hranilovic, 2014]

através da luz apagada e acesa, respectivamente. Neste caso, o bit 0 pode ser representado pela diminuição da intensidade da luz, ao invés de apagá-la completamente. Este método é prático e de fácil implementação, sendo estas suas principais vantagens. Muitos trabalhos atuais utilizam este tipo de modulação nos seus sistemas [Wang et al., 2014, Schmid et al., 2015]. Como dito anteriormente, a modulação utilizada deve levar em consideração a percepção humana da luz. Sendo assim, a modulação OOK apresenta um contratempo: caso o valor 100001 seja enviado, na teoria, a lâmpada ficaria desligada por muito tempo, podendo causar oscilações perceptíveis aos olhos humanos. Para isso, há algumas medidas propostas no IEEE 802.15.7. A primeira técnica é a de redefinição dos níveis ON e OFF, ou seja, o bit 0 passa a ser representado por outra intensidade de luz. Outra possibilidade é utilizar variações do OOK, como o ***Variable* OOK**, onde é possível obter escurecimento, o que é feito através da inserção de períodos de compensação na onda modulada, dependendo do nível de escurecimento desejado. Por padrão, os sinais na modulação OOK sempre são enviados com um símbolo Manchester simétrico.

***Variable Pulse Position Modulation* (VPPM):** Outra técnica muito utilizada em trabalhos envolvendo VLC é a VPPM. Este método faz uso de dois tipos de modulação diferentes: Modulação por posição de pulso (PPM), para impedir a oscilação de intensidade na luz, e Modulação por Largura de Pulso (PWM), para permitir escurecimento. A modulação por posição do pulso funciona da seguinte forma: A duração do símbolo é dividida em um número de *slots* de mesma duração, e um pulso é transmitido em um desses *slots*, sendo que a sua posição determina o seu valor [Elgala et al., 2011]. Uma das vantagens da PPM é a facilidade de implementação. No entanto, apenas um pulso é emitido para cada símbolo, o que faz com que a taxa de dados seja limitada. Por sua vez, o método PWM ajusta o comprimento dos pulsos, de acordo com o nível de escurecimento desejado, sendo que os pulsos carregam o sinal modulado em forma de uma onda quadrada [Pathak et al., 2015]. A figura 6.11a apresenta o funcionamento de cada uma das modulações por pulsos abordadas aqui.

***Color Shift Keying* (CSK):** No CSK, o sinal é modulado através da intensidade das três cores que compõem um tipo de lâmpada LED denominado *multi-chip*, ou *TriLED*. Esta lâmpada é composta de três ou mais LEDs, normalmente vermelhos, verdes e azuis.

Estas três cores, juntas, são utilizadas para gerar a luz branca. As modulações OOK e VPPM possuem taxas de envio de dados baixas, por isso o padrão IEEE 802.15.7 propôs esta modulação como uma solução para isso, criada especificamente para VLC. A modulação CSK se baseia no diagrama de cromaticidade CIE 1931 [Schanda, 2007]. Há sete bandas de comprimento de ondas disponíveis, de onde a origem RGB pode ser escolhida. Esta origem determina os vértices de um triângulo no qual os pontos da constelação dos símbolos CSK estão. O ponto de cor de cada símbolo é produzido modulando a intensidade dos chips RGB, por isso a limitação a lâmpadas triLEDs. Singh *et al.* fizeram um estudo detalhado onde apresentam uma primeira avaliação da modulação CSK proposta em [IEE, 2011] para diferentes combinações de bandas de cores (CBC), onde levou em consideração parâmetros como eficiência energética e *bit error rate* (BER) [Singh et al., 2013].

***Orthogonal Frequency Division Multiplexing* (OFDM):** Nesta modulação, o canal é dividido em múltiplos sub-portadores ortogonais, e os dados são enviados em *substreams* modulados em cima dos sub-portadores. Uma das grandes vantagens deste método de modulação é a redução de interferência inter-símbolos.

A utilização de lâmpadas LEDs em VLC possibilita uma comunicação **MIMO (*Multiple-Input-Multiple-Output*)**, porque muitas lâmpadas são constituídas de múltiplos LEDs. Cada LED pode ser considerado um transmissor, havendo assim múltiplos transmissores por lâmpada. Técnicas MIMO são amplamente utilizadas em comunicações por radiofrequência, pois uma de suas principais características é o aumento da taxa de dados. Entre os algoritmos MIMO utilizados em VLC, destacam-se *Repetition Coding*, *Spatial Multiplexing* e *Spatial Modulation* [Dimitrov e Haas, 2015]. Muitos trabalhos presentes na literatura implementam esta técnica com o objetivo de aumentar a velocidade da comunicação, atingindo taxas de até 1.1 Gbps [Azhar et al., 2013].

De forma geral, o grande atrativo da Comunicação por Luz Visível está na camada física. Levando em consideração as particularidades da luz visível, que a difere significativamente da radiofrequência, novas abordagens têm sido propostas, e há um esforço por parte da comunidade acadêmica e de membros de comunidades de padronização para que as principais questões relacionadas à camada física do VLC sejam resolvidas, principalmente no que diz respeito à aspectos como modulações e codificações, assim como sua influência em fatores como oscilação e escurecimento da luz, garantindo um futuro promissor para este tipo de comunicação.

6.4.3. Camada de enlace

Muitas aplicações VLC visam múltiplos usuários. Em ambientes internos, como prédios corporativos e residências, pode haver mais de uma pessoa conectada a um ponto VLC (como uma lâmpada LED). Com muitos dispositivos conectados ao mesmo tempo, é necessário criar mecanismos para controlar o acesso ao meio, realizar associação de dispositivos e possibilitar mobilidade [Pathak et al., 2015]. Esta seção tem como objetivo apresentar os três tipos de métodos para múltiplo acesso ao meio (MAC) definidos no padrão IEEE [IEE, 2011]: *Carrier Sense Multiple Access* (CSMA), *Orthogonal Frequency Division Multiple Access* (OFDMA) e *Code Division Multiple Access* (CDMA).

***Carrier Sense Multiple Access* (CSMA):** No IEEE 802.15.7, são propostos dois tipos de protocolos CSMA. No primeiro, os sinais emitidos pelo coordenador são de-

sabilitados. Dessa forma, é utilizado um canal aleatório de acesso não alocado para o CSMA. Portanto, se um dispositivo quiser transmitir, primeiro ele deve esperar por um tempo aleatório chamado *back-off period*, e depois checa se o canal está livre ou não. Caso o canal esteja ocupado, o dispositivo aguarda novamente por um período aleatório antes de tentar acessar o canal novamente. Já no segundo tipo de CSMA proposto no padrão, os sinais dos coordenadores são habilitados, e o tempo é dividido em intervalos de sinais. Um frame dentro de um intervalo de sinal contem informações como *Contention Access Period* (CAP) e *Contention Free Periods* (CFP). Caso um dispositivo queira transmitir no canal, primeiro ele deve localizar o início do próximo *back-off slot*, e esperar por um número aleatório antes de executar o *Clear Channel Assessment* (CCA). Caso o canal esteja ocioso, o dispositivo inicia a transmissão. Caso contrário, aguarda por mais *back-off slots* antes de executar o CCA novamente. Este protocolo já foi implementado em algumas pesquisas presentes na literatura, como o trabalho de Wang *et al.*, em que o protocolo CSMA/CA é expandido a fim de garantir uma comunicação bidirecional entre LEDs [Wang e Giustiniano, 2014a].

***Orthogonal Frequency Division Multiple Access* (OFDMA):** No OFDMA, múltiplos usuários recebem blocos de recursos diferentes para comunicação, chamados de *subcarriers*. Assim como a modulação OFDM é utilizada na camada física, naturalmente usa-se OFDMA para múltiplo acesso. Os principais desafios na implementação deste protocolo em sistemas VLC estão na eficiência energética e complexidade de decodificação [Dang e Zhang, 2012]. Sung *et al.* apresentaram, em seu trabalho, um sistema VLC baseado em OFDMA. Em seus experimentos, taxas de dados de até 13.6 Mbps foram alcançados [Sung et al., 2015]. Recentemente, Lin *et al.* propuseram um sistema VLC bidirecional onde o protocolo NOMA-OFDMA é utilizado. Nos experimentos realizados, o sistema apresentou alta taxa de transferência, assim como banda flexível e capacidade de usuários maior [Lin et al., 2017].

***Code Division Multiple Access* (CDMA):** O CSMA para Comunicação por Luz Visível, também chamado de *Optical* CDMA (OCDMA), consiste em códigos óticos ortogonais (OOC) para se ter acesso ao mesmo canal por usuários diferentes, técnica já utilizada em redes de fibra ótica [Pathak et al., 2015]. Em OCDMA-VLC, um código é designado a cada dispositivo, de forma que os dados possam ser codificados no domínio do tempo através dos estados de lâmpada acesa e apagada. Os códigos OOC tendem a ser longos, para garantir o aspecto ótico, o que pode reduzir a taxa de dados da comunicação.

Wang *et al.* propuseram a utilização do protocolo *Carrier Sensing Multiple Access / Collision Detection and Hidden Avoidance* (CSMA/CD-HA) em seu trabalho, para garantir o uso justo do canal entre todos os nós VLC presentes (Fig. 6.12). Além disso, este protocolo reduz o impacto de colisões e nós ocultos [Wang e Giustiniano, 2016].

Entre os aspectos do protocolo, o autor destaca:

- Possibilidade de mitigar o problema de nós ocultos. Isso ocorre devido ao fato dos *frames* do protocolo possuírem uso duplo: por um lado, eles enviam informações adicionais em banda, e por outro, funcionam como um reconhecedor ativo de recepção de dados, protegendo o transmissor primário de nós ocultos.
- Utilidade da detecção de colisões. A detecção ocorre quando o sensor recebe os

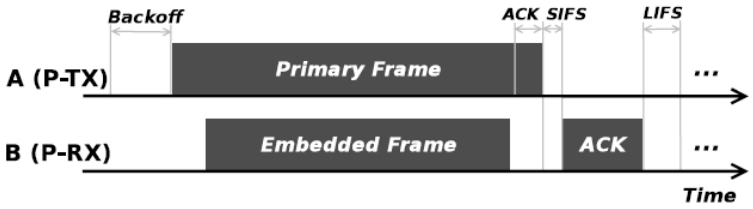


Figura 6.12: Protocolo CSMA/CD-HA [Wang e Giustiniano, 2016].

sinais *HIGH - HIGH*,que é uma sequência inválida, por um número de vezes pré-definida. A partir deste momento, a transmissão do *frame* principal é terminada, aumentando a utilização do canal.

6.5. Aplicações

VLC apresenta uma farta gama de aplicações possíveis, variando desde acesso à Internet com alta velocidade através de lâmpadas de LED, até a comunicação com estações espaciais. Claramente, as aplicações da VLC são as mais diversas e trazem uma nova perspectiva para o que é considerado computação ubíqua e pervasiva. Nesta seção, serão abordadas as potencialidades da comunicação por luz visível, assim como aplicações desta tecnologia que vêm sendo estudadas.

6.5.1. Sistemas internos

Atualmente, grande parte das casas e edifícios corporativos estão equipados com lâmpadas LEDs, como discutido na seção 6.2. Em ambientes internos, em que a infraestrutura conta com diversos transmissores, a principal função das lâmpadas é iluminação, diferente de outras aplicações, em que a luz é utilizada da mesma forma que as ondas de rádio, apenas para transmissão de dados. Dessa forma, há uma série de limitações e regulações para que sistemas VLC atendam aos requisitos de iluminação e comunicação [O'Brien et al., 2008]. Sendo assim, os canais de comunicação e o comportamento das ondas de luz em ambientes fechados devem ser observados cuidadosamente para que tais regulações sejam seguidas adequadamente. Neste sentido, o trabalho de Lee *et al.* analisam as características de dispersão em ambientes internos, considerando a refletância espectral no espectro visível em LEDs [Lee et al., 2011].

Em 2003, Toshiko Komine estudou as possibilidades da utilização de lâmpadas LEDs em sistemas VLC, sendo este um dos primeiros estudos envolvendo VLC no século XXI [Komine e Nakagawa, 2004]. No estudo, Komine *et al.* destacam todas as vantagens de lâmpadas LEDs, que até então eram novidade, além de propor um sistema VLC que utiliza essas lâmpadas para transmitir dados em um ambiente fechado. Na experiência, quatro dispositivos equipados com 3600 LEDs (60x60) foram instalados em uma sala de dimensões 5,0m x 5,0m x 3,0m. A modulação adotada é a *On-Off Keying* (OOK). Ao fim do trabalho, foram discutidos todos os requisitos para que o sistema proposto pudesse servir tanto para iluminação quanto para comunicação, além da influência da luz que é refletida nas paredes e da interferência inter-símbolos. As contribuições deste trabalho são notáveis, ainda mais considerando a época em que foi lançado. Foi provado que era possível estabelecer comunicações sem fio utilizando a luz visível com taxas muito altas,

chegando a 200 Mb/s considerando um campo de visão entre 40 e 50 graus. Além disso, o autor ainda mostra que através do uso de técnicas de rastreamento, é possível diminuir drasticamente o campo de visão, chegando a valores próximos de 5 graus, possibilitando transmissões com taxas de dados de até 10 Gbps. Mais tarde, em 2007, Jelena Grunor utilizou o trabalho de Komine *et al.* como base, mas realizou os experimentos considerando lâmpadas LEDs comerciais (LED azul com fósforo). A análise teórica feita levantou os potenciais para transmissões em alta velocidade e ambientes internos utilizando dois tipos de modulação diferentes: banda base e DMT, com a qual alcançou uma velocidade em torno de 200 Mbps [Grubor et al., 2007].

A infraestrutura de iluminação presente em salas residenciais e corporativas, em que lâmpadas são equipadas com múltiplos LEDs, oferece a possibilidade de aplicar técnicas como MIMO, discutido na seção 6.4. Neste sentido, muitos trabalhos focam em sistemas internos com MIMO. Dambul *et al.* realizaram uma pesquisa em que utiliza receptores de imagem para extrair as informações de luz do transmissor, no caso uma matriz de LEDs 2x2 [Dambul et al., 2011]. Neste trabalho, a modulação utilizada é NRZ OOK, e a velocidade do canal proposto chega a 2 Mbps. No mesmo ano, Azhar *et al.* propuseram um sistema em que o sensor de imagem utilizado é o mesmo, apesar da modulação escolhida ser diferente. O sistema 2x9 MIMO-OFDM atingiu uma velocidade de 220 Mbps, com o receptor a uma distância de 100cm dos transmissores [Azhar et al., 2010]. Dois anos mais tarde, o mesmo autor trouxe um sistema similar (4x9 MIMO-OFDM), mas o método de modulação foi otimizado, alcançando resultados na escala de gigabits/s. Mais precisamente, a uma distância de 1 m, o sistema alcança uma velocidade de transmissão de 1.1 Gbps [Azhar et al., 2013].

Uma questão essencial que deve ser tratada em ambientes internos é a necessidade de continuidade da transmissão de dados com escurecimento, e até com a luz apagada. Existe uma série de cenários em que é necessário uma iluminação reduzida, mas com transmissão de dados, o que ocorre durante dias ensolarados, por exemplo. Outro exemplo de cenário é durante a noite, antes de dormir, momento em que a luz não é necessária, mas muitas vezes há necessidade de conexão com a Internet. Dependendo do ponto de vista, tais situações podem invalidar trabalhos que consideram a Comunicação por Luz Visível apenas com as lâmpadas acesas. Com o avanço das pesquisas e a possibilidade de se utilizar VLC em dispositivos móveis, surge outro problema: a utilização de LEDs para sistemas VLC em *smartphones* pode ser custosa, em termos de energia [Li et al., 2015]. Pensando nesses problemas, Tian *et al.* propuseram uma primitiva para sistemas VLC onde a troca de dados através da luz ocorre mesmo que a lâmpada permaneça apagada, ou com intensidade baixa [Tian et al., 2016a]. De acordo com o autor, a principal ideia do trabalho consiste na codificação dos dados em pulsos de luz muito curtos, em uma frequência muito alta, de maneira que ao mesmo tempo em que a onda de luz é imperceptível aos olhos humanos, ela pode ser detectada por fotodiodos. Entre as contribuições deste trabalho, destacam-se o novo paradigma em que há Comunicação por Luz Visível mesmo que a lâmpada esteja aparentemente apagada e a diminuição do consumo de energia dos dispositivos, possibilitando novas aplicações para *smartphones*, por exemplo. Considerando que o mesmo autor havia atingido distâncias de até 10cm em um trabalho anterior [Tian et al., 2016b], o *DarkLight* apresenta uma melhoria significante, visto que em distâncias de até 1.3 m, é capaz de enviar dados com uma taxa de dados de 1.6 Kbps.

Considerando a inevitável adoção de lâmpadas LEDs no mundo, e a infraestrutura que este tipo de lâmpada oferece para sistemas VLC, o futuro de VLC em ambientes internos é muito promissor. Os desafios ainda são diversos, indo desde oscilações na luz até ruídos e interferências causadas pela luz natural, por exemplo. Ainda assim, os avanços na área deixam claro que há um grande esforço por parte da comunidade acadêmica para mitigar os problemas recorrentes a fim de se obter um cenário ideal em que se tenha acesso a Internet através de uma lâmpada LED.

6.5.2. Sistemas de trânsito e veiculares

Diversos segmentos da indústria vêm sendo contemplados com a rápida adoção de lâmpadas LEDs. De tais segmentos, pode-se destacar o setor automobilístico, onde lâmpadas LEDs são amplamente utilizadas. O uso de Comunicação por Luz Visível em sistemas veiculares pode ser muito vantajoso, dependendo do ponto de vista. O custo associado a implementação de sistemas VLC onde a infraestrutura já existe é relativamente baixo, e menos complexo, se comparado à sistemas RF. Neste sentido, as estradas oferecem um ambiente rico em fontes de luz, considerando semáforos, postes de luz e faróis de carros. Além disso, sistemas VLC funcionam baseados em LOS (*Line of Sight*), ou seja, o receptor deve estar no campo de visão da luz emitida pelo transmissor, o que pode ser adaptado facilmente para sistemas veiculares [Luo et al., 2014]. Tudo isso colabora para o conjunto de ideias, propostas e soluções envolvendo Sistemas de Transporte Inteligentes (ITS) [Papadimitratos et al., 2009].

Vale ressaltar que, assim como a maioria dos trabalhos envolvendo VLC, uma das primeiras pesquisas a abordar um sistema de Comunicação por Luz Visível no trânsito foi de um grupo japonês, realizado em 2001. Neste trabalho, foi apresentado um sistema de informação de tráfego baseado em lâmpadas LEDs já existentes na época [Akanegawa et al., 2001]. Além disso, ao invés de utilizar o espectro infravermelho para a transmissão de informações, os autores inovaram ao optar por utilizar os raios de luz visível emitidos pelos LEDs, com o objetivo de coletar informações como controle de tráfego, número e localização de semáforos, além de movimentos em direção às lâmpadas LEDs. O desempenho sistema foi analisado, assim como as modulações possíveis e SNR (*Signal-to-Noise Ratio*) necessário.

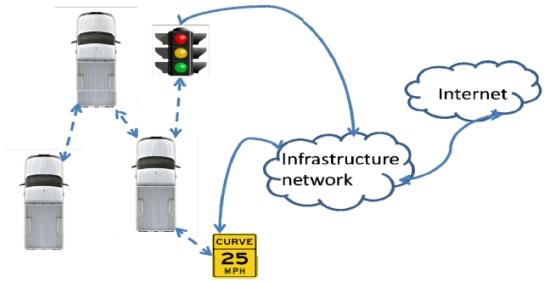


Figura 6.13: Visão geral de uma V2LC [Liu et al., 2011].

O canal de propagação das ondas desempenha um papel muito importante na qualidade da comunicação, principalmente em um ambiente dinâmico como em comunicações veiculares. Cheng *et al.* desenvolveram um trabalho onde compara as ondas de luz e de rádio, assim como suas capacidades e limitações. O primeiro aspecto analisado pelo autor é o padrão de radiação e perda de caminho (*path loss*). De acordo com o autor, entre as vantagens do canal ótico, estão a alta taxa de transferência, a eficiência espacial,

consequência dos dispositivos atuarem em LOS (*line of sight*) e custo baixo, se comparado a dispositivos de rádio [Cheng et al., 2016]. Trabalhos envolvendo a modelagem do canal em Comunicação por Luz Visível são importantes neste momento, visto que os que tratam do canal de rádio já são presentes na literatura.

Uma das possíveis aplicações de VLC em redes veiculares é entre carros, chamada de *Car-to-Car* (C2C), ou *Vehicle-to-Vehicle* (V2V). Dessa forma, o farol de um carro funciona como transmissor, enquanto fotodiodos podem ser equipados e desempenhar o papel de receptor, em comunicações bidirecionais, por exemplo [Liu et al., 2011]. A viabilidade do VLC neste tipo de cenário já foi confirmada em alguns trabalhos presentes na literatura [Kim et al., 2012]. Em 2013, foi proposto um sistema VLC onde é utilizado um sensor de imagem para captar as ondas de luz, atingindo velocidades de até 20 Mbps/pixel [Takai et al., 2013]. Além disso, a performance para diferentes distâncias e condições de luz também foi apresentada, além das limitações envolvendo ruídos vindos de circuitos periféricos. Os mesmos autores apresentaram um sistema mais robusto e resistente à ruídos e interferências um ano mais tarde [Takai et al., 2014]. Neste trabalho, um dos principais aspectos que garantiu a melhoria na performance foi a utilização de uma *flag image*, onde objetos de baixa intensidade de luz são eliminados, enquanto objetos de alta intensidade, como LEDs, são registrados, eliminando grande parte dos ruídos e objetos desnecessários. Em 2014, foi apresentado um modelo matemático para Comunicação por Luz Visível em um cenário C2C, em que a performance do sistema foi medida para diferentes geometrias [Luo et al., 2014].

De maneira mais ampla, uma rede de comunicação sem fio veicular (V2LC) consiste de um ou mais nós móveis, representados por veículos, e estruturas fixas, como semáforos e postes de iluminação. Ambos podem ser equipados com transmissores e receptores, que funcionam simultaneamente, construindo assim uma rede de comunicação dinâmica, capaz de colher e enviar as informações captadas pelos diferentes sensores acoplados nos veículos e nas redondezas [Liu et al., 2011]. *Liu et al.* identificaram e classificaram os tipos de serviços envolvendo V2LC em cinco categorias diferentes, além de desenvolver um protótipo baseado em três princípios: utilização de dispositivos acessíveis para composição do hardware, uso de técnicas analógicas para aumentar a resistência do protótipo à ruídos e por fim a disponibilização de um ambiente de programação flexível para implementação de algoritmos. Os autores experimentaram a plataforma em cada cenário proposto, e mostraram que V2LC é viável em ambientes veiculares. Além disso, foi mostrado através de uma série de simulações que V2LC atende aos requisitos de latência e alcance em cenários de alta densidade de veículos. O único trabalho experimental anterior a este com foco similar foi realizado em 2009, onde os autores propuseram um sistema V2LC unidirecional entre os semáforos e os veículos [Okada et al., 2009].

Uma demonstração prática de um sistema veicular em que se utiliza VLC foi feita em 2015. Nela, Yoo *et al.* propuseram um sistema VLC e usou faróis de carros em seus experimentos, em um ambiente veículo-para-veículo (V2V). O transmissor é composto de um farol LED, e um módulo de *driver*. O farol usado é comercializado, utilizados pela Renault Samsung Motors. Por isso, já está dentro das regulações necessárias de emissão de radiação e distribuição de luz. Para garantir controle de intensidade da luz emitida e transmissão de dados, foi utilizada uma modulação denominada *Inverse M-ary Pulse Position Modulation* (I-M-PPM). O receptor consiste de um fotodiodo, uma lente e um

filtro de cores. Um dos desafios encontrados pelos autores foi à respeito da interferência causada por outras fontes, principalmente durante o dia, com a luz solar. O espectro da luz solar pode ser distribuído em todas as regiões da luz visível, infravermelho e ultravioleta. Por isso, torna-se difícil a filtragem dos sinais. O filtro de cores foi utilizado com o objetivo de diminuir a interferência. Em seus experimentos, foram alcançadas taxas de até 10kbps, a uma distância de mais de 30 metros, durante o dia.

6.5.3. Sistemas de localização

A rápida adoção de lâmpadas LEDs em ambientes internos possibilita uma gama considerável de aplicações VLC, como discutido anteriormente. Entre elas, destacam-se as aplicações envolvendo localização através de VLC. Através de VLC, é possível utilizar os LEDs para determinar a localização de uma pessoa em um ambiente interno, possibilitando a criação de sistemas flexíveis e precisos. Novamente, os sistemas VLC funcionam de forma análoga a outros sistemas já concebidos na sociedade. Neste caso, o *Global Positioning System* (GPS) é um exemplo de uma aplicação criada originalmente para uso militar, que é indispensável no dia-a-dia de muitas pessoas atualmente. Entretanto, este sistema apresenta muitas limitações quando utilizado dentro de edifícios, visto que as ondas emitidas por satélites estão em uma frequência que não ultrapassa qualquer barreira. Além disso, dependendo da precisão necessária para um dado sistema, não é possível contar com o GPS.

Considerando a limitação de um dos sistemas de localização mais populares, surgiram muitas alternativas para posicionamento em ambientes como edifícios e casas. Entre as tecnologias estudadas, figuram RFID, WLAN, UWB e *Bluetooth* [Liu et al., 2007]. Além dessas, a Comunicação por Luz Visível também constitui uma alternativa para localização interna. Entretanto, a Comunicação por Luz Visível apresenta uma série de vantagens quando comparada às outras alternativas. Em primeiro lugar, a luz visível é livre de interferências eletromagnéticas. Além disso, a maioria dos locais já oferece uma infraestrutura (lâmpadas LEDs) para a implementação de sistemas de localização, tornando esta opção acessível e barata.

A partir do momento em que é necessário haver um posicionamento com alta precisão, o receptor deve captar os sinais dos LEDs na sala e calcular a distância em relação a eles. Depois, as medidas são utilizadas por algoritmos e técnicas a fim de estabelecer a posição exata do receptor. Uma dessas técnicas é baseada em **RSS (*received signal strength*)**, muito utilizada em sistemas de radiofrequência. No entanto, quanto maior a distância entre os transmissores e o receptor, mais fraco será o sinal, e obstáculos podem interferir no RSS, pois bloqueiam ou refletem as ondas, limitando assim a precisão deste método. Outro método muito utilizado é através do cálculo do ***Time of Arrival* (TOA)**. Entretanto, esta técnica requer a transmissão de sinais rigidamente sincronizados, o que pode exigir recursos mais caros, tornando o sistema mais caro. Por fim, existe o método em que o cálculo é feito através do ***Angle of Arrival* (AOA)**. Esta técnica não é comum nos sistemas de radiofrequência, visto que depende de LOS, ao contrário de sistemas VLC, que dependem da LOS para funcionamento. Trabalhos que utilizam o método de *fingerprinting* também podem ser encontrados na literatura [Qiu et al., 2016].

O primeiro sistema de localização prático de alta precisão interno foi apresentado

em 2014 [Li et al., 2014]. Levado por fatores como a grande quantidade de lâmpadas LEDs em ambientes internos e a dualidade iluminação/comunicação ofertada por VLC, Liqun Li *et al.* criaram o sistema Epsilon, que fornece localização com alta precisão, custo baixo e de fácil implementação. O algoritmo de localização utilizado pelo sistema é o de trilateração, onde a posição do receptor é calculada a partir da força do sinal recebido (RSS) medida entre múltiplos transmissores. Ao contrário de métodos de localização via Wi-Fi presentes na época, o Epsilon é capaz de calcular a localização em escala de centímetros (aproximadamente 0.4 m), aumentando consideravelmente a precisão.

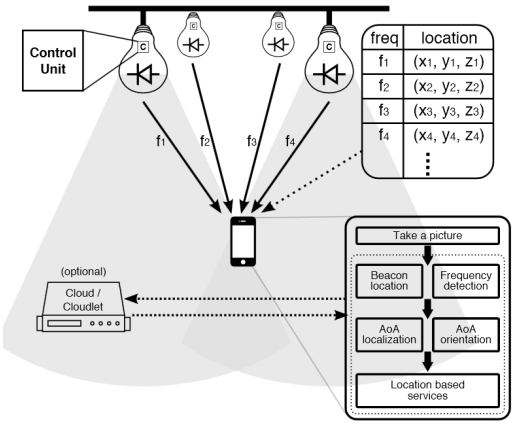


Figura 6.14: Arquitetura do sistema Luxapose. [Kuo et al., 2014].

Outro exemplo de sistema de localização interna é o Luxapose, apresentado em 2014 [Kuo et al., 2014]. Enquanto o sistema Epsilon utiliza fotodiodos como receptores, a inovação do Luxapose consiste na utilização de sensores de imagens como receptores, como a câmera de um *smartphone*, por exemplo. Neste trabalho, os circuitos de lâmpadas LEDs são modificados para que dados sejam transmitidos utilizando a modulação *On-Off Keying*. O sistema consiste de três principais componentes: sinais de luz enviados pelos LEDs, *smartphones* e um servidor na nuvem, como pode ser visto na figura 6.14 . O sinal de luz enviado por cada LED contem sua identidade de coordenadas. Baseado no ângulo de chegada (*angle-of-arrival* (AoA)) do sinal transmitido e orientação da câmera do *smartphone*, é feita a triangulação para que o receptor seja localizado. Os recursos da nuvem são utilizados para auxiliar no processamento da imagem e orientação no sistema de coordenadas. O Luxapose supera o Epsilon pois além de calcular a localização em escalas de aproximadamente 0.1 m, a orientação do dispositivo também é calculada.

Sistemas híbridos também são encontrados na literatura. Prince *et al.* já haviam proposto um algoritmo de localização onde são utilizadas tanto RSS quanto AoA para determinar a posição do receptor [Prince e Little, 2012]. O sistema foi modelado e os resultados mostraram a viabilidade da utilização de VLC para posicionamento, além das vantagens em comparação com sistemas estudados em que se utiliza *Bluetooth* ou Wi-Fi. Um trabalho similar foi realizado em 2014, em que o canal de comunicação de um sistema com um receptor e múltiplos emissores foi modelado, e o algoritmo proposto também utiliza RSS e AoA para medir a posição do receptor, com precisão [Yang et al., 2014].

Localização baseada em VLC é, sem dúvidas, uma aplicação muito importante que estará presente nas tecnologias do futuro, principalmente no que diz respeito à loca-

lização de alta precisão. Além da a academia, serviços como *ByteLight*[13] mostram que a indústria também tem mostrado interesse em aplicações de posicionamento.

6.5.4. Comunicação aquática

Nos últimos anos, redes aquáticas tornaram-se muito populares, principalmente com o avanço das tecnologias relacionadas à redes de sensores sem fio. A água oferece muitos desafios quando comparada aos meios de comunicação tradicionais, como o ar. Veículos operados remotamente (ROV) e comunicação entre mergulhadores também são aplicações que exigem o uso de comunicação aquática [Rust e Asada, 2012].

Atualmente, a maioria dos trabalhos acadêmicos tem seu foco em três ondas distintas para sistemas de comunicação aquáticos: ondas acústicas, eletromagnéticas e óticas. A comunicação acústica é a mais utilizada nos dias de hoje. Sua principal vantagem consiste na baixa redução do som na água. Entretanto, fatores como a velocidade de propagação muito baixa, interferência de ruídos e temperatura próximas da superfície são limitações deste tipo de comunicação. Por sua vez, o uso de ondas eletromagnéticas em radiofrequência é naturalmente limitado pelas propriedades do meio. Frequências utilizadas em aplicações comuns de radiofrequência (2.4 GHz ISM) têm comprimentos de onda e velocidades limitados devido à alta atenuação da água, principalmente em águas salgadas. Por outro lado, as velocidades proporcionadas por ondas de rádio altas, possibilitando uma comunicação mais rápida e eficiente. Por fim, existe a possibilidade de se utilizar ondas de luz para comunicação aquática, assunto abordado nesta seção. A tabela 6.6 apresenta uma comparação sucinta dos tipos de comunicação citados anteriormente.

Parâmetros	Acústica	Radiofrequência	Ótico
Velocidade (m/s)	1500	2.255 x 10^8	2.255 x 10^8
Taxa de dados	$\sim$ Kbps	$\sim$ Mbps	$\sim$ Gbps
Distância	$\sim$ km	$\sim$ 10m	$\sim$ 10m - 100m
Largura de banda	kHz	MHz	MHz
Potência de transmissão	Dezenas de Watts	Dezenas de Watts	Poucos Watts
Par. de performance	Pressão	Condutividade	Absorção, turbidez

Tabela 6.6: Comparação dos tipos de comunicação aquática, adaptado de [Kaushal e Kaddoum, 2016].

O uso de ondas óticas oferece uma grande vantagem no que diz respeito à taxa de dados. Como discutido anteriormente, a frequência das ondas de luz são até 1000 vezes maiores que as de radiofrequência, oferecendo taxas que chegam a Gbps. Entretanto, os sinais óticos são rapidamente absorvidos pela água, além de se espalharem mais facilmente devido à partículas suspensas e planktons [Lanbo et al., 2008]. Ainda assim, diferentes frequências têm comportamento diferentes na água. Um exemplo disso é o fato de a absorção ser maior próximo da região vermelha do espectro, enquanto na parte próxima a violeta, é menor. Portanto, o ambiente e a aplicação tornam-se fatores importantes para a escolha do comprimento de onda do sistema [Karunatilaka et al., 2015].

[13]ByteLight™ Services: Indoor Positioning - `http://hydrel.acuitybrands.com/sitecore/content/acuitybrandscorporate/home/solutions/services/bytelight-services-indoor-positioning`, 2017.

Schill *et al.* apresentaram um sistema aquático em que se utiliza VLC, e realizou experimentos tanto no ar quanto na água para testá-lo [Schill et al., 2004]. No ar, a distância entre receptor e transmissores chegou a 2,02 m, na cor ciano, sem que houvesse perda de dados. Na água, a distância alcançada chegou a 1,7 m. De maneira geral, para que um sistema VLC aquático obtenha sucesso no funcionamento, uma série de fatores devem ser rigorosamente observados.

Primeiramente, **a absorção e espalhamento** são fenômenos que atuam na perda de intensidade ou mudança de direção de um sinal ótico na água. Quando a água é iluminada com um raio de luz, uma fração é absorvida pela própria, enquanto outra é espalhada, deixando o restante do raio de luz inafetado. Os valores de absorção mudam de acordo com o ambiente aquático em que a luz está sendo emitida. Dessa forma, o raio de luz sofrerá muito menos interferência por parte da absorção nas águas claras do oceano em comparação com as águas turvas de um porto, por exemplo. No mar, materiais inorgânicos (sais e moléculas de água) e orgânicos (fitoplânctons) contribuem para o aumento da absorção e espalhamento dos raios de luz.

A **turbulência** da água é outro fator que interfere nas ondas de luz, devido à variação na refração das ondas causadas por alterações na densidade e salinidade da água [Hou et al., 2013].

Como dito anteriormente, sistemas VLC baseiam-se na premissa LOS, ou seja, o receptor deve estar no campo de visão do transmissor. Neste sentido, outro fator importante para aplicações aquáticas é o **alinhamento**, pois o raio ótico pode ser muito estreito, além da interferência de correntes oceânicas, exigindo medidas adicionais para garantir LOS. Uma dessas medidas pode ser vista no trabalho de Simpson *et al.*, onde transmissores e receptores inteligentes são capazes de alinhar e ajustar o campo de visão de acordo com parâmetros como qualidade da água [Simpson et al., 2012].

Assim como em outras aplicações VLC, o tipo de modulação mais utilizado em sistemas aquáticos é baseada na modulação da intensidade (IM), onde os dados são modulados digitalmente através da alteração da intensidade da luz. Este tipo de modulação, unido ao esquema de detecção direta, é simples e barato, portanto é amplamente utilizado. Técnias como *On-Off Keying*, *Pulse Position modulation* são muito comuns em sistemas aquáticos. Oubei *et al.* desenvolveram um sistema capaz de transmitir dados em ambientes aquáticos utilizando a modulação OOK-NRZ. O sistema funciona de maneira satisfatória em distâncias de até 7 m entre o receptor e o transmissor, e atinge velocidade de até 2.3 Gbps [Oubei et al., 2015]. Técnicas de MIMO associadas à modulação OOK também foram apresentadas na literatura com o objetivo de diminuir os efeitos da turbulência [Jamali et al., 2016]. Outros tipos de modulação como OFDM também são utilizados em aplicações aquáticas [Xu et al., 2016].

Em suma, vantagens como alta taxa de dados e baixo consumo de energia fazem da Comunicação por Luz Visível uma grande candidata a figurar entre as tecnologias de comunicação aquática no futuro, devido à sua eficiência e robustez. Além disso, sistemas híbridos que atuam tanto com a comunicação acústica quanto ótica são capazes de enviar dados em velocidades maiores e trocar o modo de funcionamento de acordo com aspectos da água, podendo ser utilizados em robôs marinhos, por exemplo. Assim como em outras áreas, os desafios de sistemas aquáticos VLC são muitos, mas os estudos apontam para

uma diretriz promissora, onde sistemas assim serão cada vez mais eficientes.

6.6. Desafios

O objetivo desta seção é explicar com detalhes alguns desses desafios, como *Flickering*, interferência e *uplink*, além de apresentar algumas soluções propostas para contornar tais desafios.

6.6.1. Flickering

O *Flickering* é um dos principais desafios envolvendo VLC, e pode ser definido como a flutuação no brilho da luz. Este problema pode ser encontrado em trabalhos que implementam sistemas VLC em ambientes internos, como em um escritório, ou supermercado. Dependendo do mecanismo de modulação das ondas de luz implementado, pode haver oscilações perceptíveis aos olhos humanos, podendo causar desconfortos e riscos para a saúde [Wilkins et al., 2010]. Deste modo, existe a necessidade de modular as ondas de forma que a menor de suas frequências seja maior que o limite em que o ser humano percebe, que é menor do que 3 KHz. Existe, portanto, uma dualidade entre frequências altas que resultam em taxas de dados maiores e o *flickering* causado por elas.

Modulações	*Intra-frame flicker*	*Inter-frame flicker*
OOK	Codificação RLL	Padrão ocioso/visibilidade
VPPM	Não ocorre	Padrão ocioso/visibilidade
CSK	Mesma potência para fontes diferentes	Padrão ocioso/visibilidade

Tabela 6.7: Tipos de modulação e suas respectivas maneiras de mitigar o *flicker*.

O padrão IEEE 802.15.7 para VLC de curta distância define algumas medidas para mitigar este problema. O *flicker* pode ser categorizado de duas formas, em VLC: intra-frame e inter-frame [IEE, 2011]. O primeiro determina a oscilação de brilho detectada entre frames, já a segunda forma é definida pela oscilação entre transmissões adjacentes. No caso do *flicker intra-frame*, a solução utiliza modulações ou codificações que corrijam o *flicker*, como a Manchester, 4B6B, e VPPM. Um exemplo prático de utilização de uma destas técnicas pode ser encontrado na plataforma de pesquisas OpenVLC [Wang et al., 2014]. A plataforma utiliza como modulação o método *On-Off Keying* com codificação Manchester. Esta solução resolve o problema de *flickering* pois na codificação Manchester sempre há símbolos 0 e 1 para representar um sinal, e como na modulação OOK os sinais são representados pelo LED aceso ou apagado, o problema se resolve. Já no caso de *flickering* inter-frame, a oscilação ocorre entre o tempo ocioso do LED e o tempo de envio de informação [Oh, 2013]. Para que este tipo de oscilação seja mitigado, são propostos padrões para que, durante o tempo em que o LED esteja ocioso, o brilho seja mantido em uma frequência acima da detectada pelos humanos. A tabela 6.7 apresenta os principais métodos de modulação utilizados em sistemas VLC, assim como alternativas para que não haja oscilações no brilho intra-frame e inter-frame.

6.6.2. Interferência e ruídos

Em uma rede WiFi, tem-se o problema de interferência causada por outros dispositivos que emitem sinais na mesma frequência (normalmente pontos de acesso e roteadores wi-

reless). Quando se utiliza a luz como meio de comunicação, a luz natural passa a ser uma fonte de interferência na comunicação, fato que cria diversos desafios, principalmente tratando-se de aplicações externas. Além da luz natural, outros fatores como luzes artificiais também causam interferência destrutiva na comunicação, podendo até saturar o receptor. Outro fator que atua como interferência no receptor é o problema do multicaminho. Diferente de uma comunicação cabeada, onde a propagação do sinal é restrita a um fio, na Comunicação por Luz Visível o sinal é propagado para o ambiente de acordo com a direção das lâmpadas LEDs, passando por fenômenos como refração e reflexão, podendo chegar ao receptor mais de uma vez.

A partir do momento em que a premissa de funcionamento de um sistema VLC engloba o uso restrito de lâmpadas LEDs em ambientes internos, interferências de lâmpadas incandescentes e fluorescentes pode ser desconsiderado. Entretanto, sistemas VLC que atuam em conjunto com outras fontes de luz terão problemas com os ruídos. Neste sentido, vale destacar o trabalho de Moreira *et al.*, onde foi realizada uma caracterização da interferência produzida por luzes artificiais, e além disso, foi proposto um modelo para descrever tal fenômeno. Os autores identificaram três classes de lâmpadas que oferecem interferência: incandescentes e dois tipos de lâmpadas fluorescentes [Moreira et al., 1997]. Apesar de ser um trabalho relativamente antigo (final dos anos 90), os tipos de lâmpadas estudadas ainda estão presentes nos dias de hoje.

A solução para mitigar problemas relacionados a interferências e ruídos é o uso de filtros óticos. Através do uso de filtros, ruídos de fontes naturais de luz podem ser removidos [Moreira et al., 1997]. Esta é uma abordagem comum, que pode ser encontrada em diversos trabalhos em que são implementados sistemas VLC [Yoo et al., 2016]. Amplificadores de sinais também foram utilizados para mitigar ruídos [Schmid et al., 2016a].

A escolha do receptor também é uma estratégia para evitar ruídos e interferências. Fotodiodos tendem a ser extremamente sensíveis, porque além de captarem o espectro visível, também captam ondas do infravermelho e ultravioleta. Por outro lado, quando se utiliza um LED como receptor, a situação muda. Como discutido na seção 6.3, os LEDs também são sensores, no entanto, captam somente ondas próximas às que eles transmitem, servindo como um filtro para as luzes naturais, por exemplo. Um exemplo muito interessante da diferença entre fotodiodo e LED na prática pode ser encontrada em [Wang et al., 2014], onde a performance foi avaliada em ambientes internos e externos, utilizando tanto o fotodiodo quanto o LED como receptores. De acordo com o autor, em ambientes externos, o fotodiodo ficava saturado rapidamente, enquanto o LED obteve sucesso na recepção de dados.

De maneira geral, independente da aplicação, problemas relacionados à interferência e ruídos vão ocorrer. Uma grande vantagem de VLC, nesse ponto, é a possibilidade de se enxergar a maioria das fontes de interferência, sejam elas naturais ou artificiais.

6.6.3. Canais de Uplink

Outra questão a ser tratada que surge com a ideia de VLC utilizando LEDs é o *uplink*. Sabe-se que o LED é um emissor de luz; apenas um dispositivo como um fotodiodo garante a recepção de dados. Neste sentido, o *downlink* é garantido. No entanto, considerando uma comunicação bidirecional, e.g. em que um notebook também desempenha

o papel de emissor, enviando dados para o LED, tem-se um problema com relação ao dispositivo que emitirá a luz.

De fato, a maioria das pesquisas em VLC tem seu foco em comunicações unidirecionais [Khalid et al., 2012, Cossu et al., 2012]. O *uplink* se torna um problema quando se tem, por exemplo, diversas células receptoras com características baseadas em *broadcast*. Aplicações VLC para *smartphones* e outros aparelhos de baixa potência também encontram muitos obstáculos relacionados ao envio de informações de volta ao transmissor. Em um ambiente interno não seria adequado o uso de luz visível para o *uplink* a partir do momento em que mais fontes de luz são adicionadas ao ambiente, em direções opostas, causando desconforto aos olhos humanos. Ainda assim, entre as propostas encontradas na literatura para mitigar este problema, figuram o uso de luz visível [Wang et al., 2013a], radiofrequência [Rahaim et al., 2011], infravermelho [Grobe et al., 2013] e *transceivers* retro-reflexivos [Komine et al., 2003].

Em primeiro lugar, o uso de radio frequência como alternativa para *uplink* oferece algumas vantagens, visto que não haverá uma fonte de luz saindo do dispositivo pessoal, como um *notebook*, por exemplo. Entretanto, *transceivers* RF terão que ser acoplados tanto no emissor quanto no receptor, indo de encontro a uma das propostas de VLC: o baixo custo. Além disso, há ambientes em que o uso de radiofrequência não é ideal, como em hospitais e aviões, por exemplo.

O uso de VLC para *downlink* e *uplink* também é encontrado na literatura. Para que a interferência por parte da reflexão do sinal seja eliminada, técnicas como *Time-division-duplex* (TDD) foram propostas. O uso de TDD diminui a taxa de dados, visto que são alocados *slots* de tempo específicos para *downlink* e *uplink*. Entretanto, mecanismos de modulação de alto nível podem aumentar a taxa de dados, como OFDM [Liu et al., 2012].

O uso do espectro ultravioleta (UV) como canal de *uplink* por causar pouca interferência com a luz visível, entretanto, os trabalhos que propõem UV têm seu foco em comunicação de longa distância, onda há NLOS (Non-Line of Sight). O uso de ondas do espectro infravermelho (IR) são interessantes, à medida em que também não interferem com a comunicação, e dispositivos IR são acessíveis. Um trabalho prático em que o *uplink* é feito através de IR durante um vôo foi feito por Perez-Jimenez *et al.*, onde foi alcançada uma velocidade de *uplink* de 512 Kbps.

Por fim, uma proposta interessante foi feita por Komine *et al.* em 2003. Em seu trabalho, os autores propuseram um sistemas VLC interno utilizando lâmpadas LEDs brancas. O diferencial de seu trabalho consiste na adoção de uma técnica para implementação de *uplink*. O mecanismo consiste no uso de uma superfície com alto poder de reflexão para modular a informação e retorná-la ao transmissor [Komine et al., 2003].

6.6.4. Níveis de iluminação

Considerando o uso de lâmpadas LED para troca de dados, deve-se considerar que a potência do sinal está diretamente relacionada com a intensidade da luz. Assim, teoricamente, quanto mais "escurecida" a luz, pior serão alcance e velocidade da comunicação. O escurecimento da luz pode ser definido como o controle do brilho percebido da fonte de luz, de acordo com os requisitos do usuário. Em muitos locais, o escurecimento é um

fator essencial, gerando benefícios como a geração de ambientes confortáveis e a economia de energia. Nestes casos, os sistemas VLC devem ter suporte para tal situação. O documento de padronização IEEE 802.15.7 define uma série de medidas para adequar as modulações ao escurecimento do LED. Todas as modulações apresentadas na seção 6.4 têm sua versão com controle de escurecimento.

De acordo com o padrão, a modulação ***On-Off Keying*** é sempre enviada com um símbolo simétrico devido à codificação Manchester, utilizada em conjunto. Para que esta modulação ofereça controle de escurecimento, um tempo de compensação deve ser inserido nos *frames* com o objetivo de ajustar a média de intensidade da luz. Isso é feito da seguinte forma: um *frame* é quebrado em *sub-frames*, onde o tempo de compensação é inserido junto com campos de re-sincronização, para manter a integridade dos *frames*.

A modulação VPPM (*Variable Pulse Position Modulation*) também é adaptada para escurecimento, além de oferecer proteção contra *flicker intra-frame*. Neste caso, o controle de escurecimento é feito através da manipulação da largura dos pulsos, pois como discutido na seção 6.4, a modulação VPPM faz uso de duas modulações: PPM (*pulse position modulation*) e PWM (*pulse width modulation*). A figura 6.15 apresenta o mecanismo básico de controle de escurecimento no VPPM. Como pode ser visto, a taxa de escurecimento é alcançada através da manipulação da largura de cada pulso.

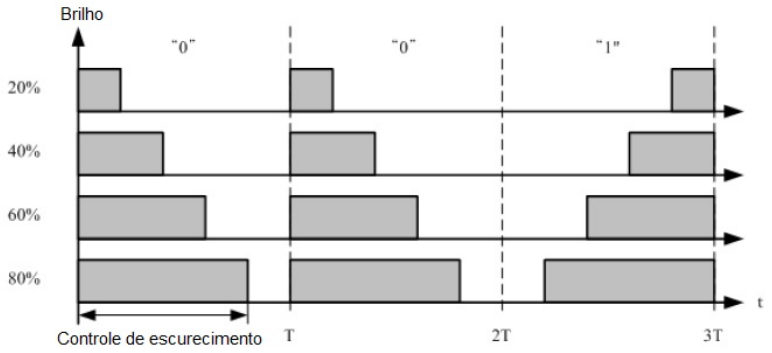


Figura 6.15: Mecanismo de modulação com controle de escurecimento [IEE, 2011].

Por fim, a modulação *Color-Shift Keying* também oferece controle de escurecimento através da manipulação da corrente que passa pelos LEDs. Entretanto, um problema que pode surgir é a mudança de cores devido à essa alteração de corrente.

6.6.5. Linha de visão

Em sistemas baseados em VLC para ambientes internos, pressupõe-se que o usuário está na linha de visão da fonte de luz (como uma lâmpada, por exemplo). No entanto, apesar das propriedades de reflexão da luz visível, ela não é capaz de penetrar a maioria dos objetos, podendo causar impactos na conexão e na experiência do usuário.

A maioria das aplicações envolvendo sistemas VLC internos propõem um mecanismo de funcionamento baseado em LOS (*Line-Of-Sight*) direto, como pode ser visto na figura 6.16a, ou seja, o receptor deve ter uma linha de visão clara do transmissor na maioria do tempo. A principal vantagem oferecida por esse mecanismo é a recepção de um

sinal mais forte. Na literatura, trabalhos que buscam maior velocidade da comunicação baseiam-se fortemente neste mecanismo, visto que, naturalmente, quanto menor o ângulo de abertura do transmissor, maior a resposta do canal, e consequentemente, maior a velocidade alcançada [Gomez et al., 2015]. Entretanto, se um receptor é capaz de captar os fótons, mesmo que fora da linha de visão, haverá possibilidade de transferência de dados, com uma taxa menor, devido à fraqueza do sinal, como é o caso da figura 6.16b e 6.16c. As propriedades da luz também trazem alguns aspectos interessantes, pois com a reflexão, por exemplo, existe a possibilidade de receber um sinal indiretamente.

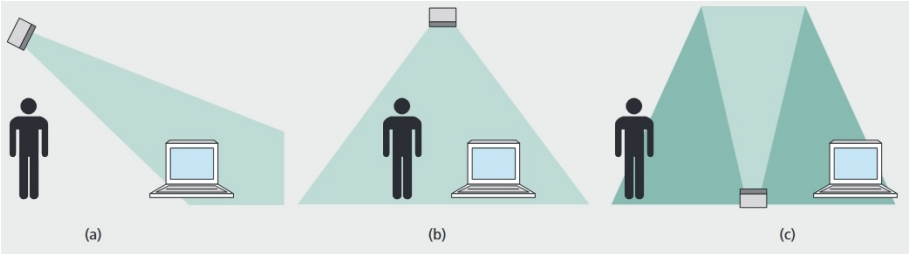


Figura 6.16: a) LOS direto b) LOS indireto c) Luz difusa [Grobe et al., 2013].

Outro fator essencial para a caracterização do canal é o constante movimento dos usuários. Em um sistema VLC interno em que o receptor é um *smartphone*, o usuário pode mudar de direção e orientação constantemente. Dessa forma, existe a necessidade de se desenvolver técnicas visando a comunicação em alta velocidade mesmo que o receptor não esteja em contato direto com o transmissor. Junto a isto, figura o problema criado pela presença de sombras, que diminuem drasticamente as ondas de luz que chegam ao receptor. Neste caso, deve-se explorar a luz refletida por obstáculos, de forma que o sistema reaja de maneira eficiente durante os momentos de bloqueio de luz.

Na prática, sistemas comerciais como o LiFi já possuem mecanismos de adaptação ao nível de iluminação. Um exemplo disso foi demonstrado pelo criador da tecnologia, Harald Haas durante um TED Talk, em 2015[14].

6.6.6. Mobilidade

Assim como no WiFi, alguns sistemas baseados em VLC devem permitir mobilidade ao usuário. Sendo assim, para que a Comunicação por Luz Visível se torne uma tecnologia ubíqua, são necessários mecanismos de forma a garantir uma conexão de alta velocidade de maneira ininterrupta, dentro da área de cobertura do sistema. Em outras palavras, o receptor deve ser capaz de detectar os sinais de luz do transmissor em qualquer lugar de uma sala, por exemplo, e para isso seria necessário um ângulo de emissão maior no transmissor e um FOV maior no receptor, o que pode causar uma maior interferência por parte das ondas refratadas [Karunatilaka et al., 2015].

A Comunicação por Luz Visível difere significativamente da radiofrequência no que diz respeito à propagação do sinal, por depender fortemente de LOS, além da orientação do receptor em relação ao transmissor. O SNR (*Signal-to-Noise Ratio*) da luz pode

[14]Harald Haas: Forget WiFi. Meet the new Li-Fi Internet - `https://www.ted.com/talks/harald_haas_a_breakthrough_new_kind_of_wireless_internet#t-308910`, 2015.

variar muito a partir do momento em que o receptor se movimenta, até mesmo dentro da área de cobertura da luz [Zhang et al., 2015]. Dessa forma, um grande desafio encontrado pela comunidade acadêmica é criar protocolos de adaptação de taxas, com o objetivo de serem eficazes em sistemas VLC.

Burton *et al.* apresentaram, em seu trabalho, uma avaliação do *design* ótico e geométrico de um sistema VLC interno utilizando LEDs brancas que possibilita mobilidade total, dentro da cobertura do transmissor. Para isso, é utilizado um receptor de diversidade angular (ADR), em que fotodiodos são acoplados geomatricamente, a fim de obter uma cobertura esférica do ambiente. Além disso, o receptor implementa combinação seletiva (SC), ou seja, o receptor com o maior sinal é utilizado para detecção de sinal. Em um ambiente fechado de 5 m^2, o sistema foi capaz de oferecer mobilidade completa, e tem funcionamento satisfatório para velocidades de até 55 Mbps.

Mecanismos de *handover* também são muito importantes para que a mobilidade seja garantida, sem perda de informação. Neste sentido, destaca-se a tecnologia LiFi, que é associada fortemente à integração de dispositivos WiFi e dispositivos VLC de alta velocidade. Os sistemas baseados em LiFi permitem múltiplos pontos de acesso que formam uma rede com *handover* integrado. Além disso, o LiFi possui um mecanismo de *handover* dinâmico para sistemas híbridos (WiFi/Lifi) [Haas et al., 2016], que pode reduzir a quantidade de *handovers* na rede e pode alcançar velocidades maiores.

Sistemas veiculares também baseiam-se fortemente em mecanismos de *handover*, visto que a maioria dos nós se movimenta constantemente. Zhu *et al.* realizaram simulações em que se utiliza o método de *handover* RSP (*receiving signal power*) em uma rede veicular, onde ocorre comunicação entre veículos que se movimentam e semáforos LEDs.

A mobilidade dos usuários e a cobertura dos sinais dos transmissores constituem um grande desafio para os sistemas VLC. Em especial, é um dos grandes desafios para a comercialização de tecnologias assim, visto que a mobilidade é essencial, e a ideia de perder o sinal do transmissor por causa de uma simples rotacionada no *smartphone*, por exemplo, não agrada os usuários.

6.7. Prática

Nesta seção é apresentado a OpenVLC [Wang et al., 2015a], uma plataforma para comunicação por luz visível que por ser flexível, de baixo custo e de código aberto é uma excelente opção para a prototipagem de sistemas VLC. A seguir é mostrado um tutorial detalhado de como instalar, configurar e utilizar esta plataforma.

6.7.1. Conceitos básicos sobre OpenVLC

A plataforma OpenVLC foi montada sobre o Beagle Bone Black (BBB) [Coley, 2013] sendo este um computador de placa única fácil de usar, versátil e de baixo custo. A plataforma consiste em uma placa de expansão (vide figura 6.17) contendo os elementos ópticos, como LEDs e fotodiodos, necessários para realizar a comunicação. A recepção do sinal pode ser feita atráves do LED ou do fotodiodo; já a transmissão é feita usando-se um LED comum ou um LED de alta potência.

[15] OpenVLC components - `http://www.openvlc.org/openvlc.html`

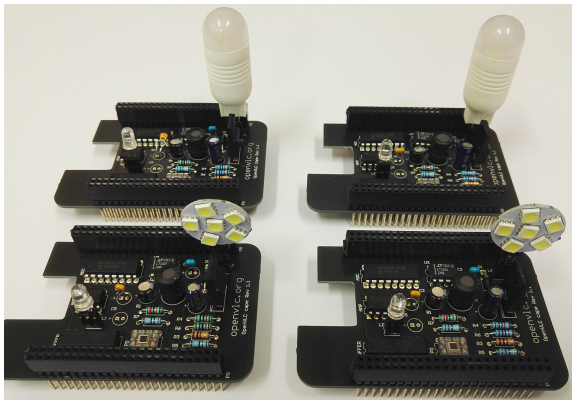


Figura 6.17: Placas de expansão OpenVLC.[15]

Os componentes de *software* que compõem a plataforma OpenVLC foram implementados nas camadas física e enlace. Como pode ser visto na figura 6.18, foram implementadas funções tais como amostragem de sinais, detecção de símbolos, codificação e decodificação, detecção e controle de colisões, e interoperabilidade com o protocolo Internet.

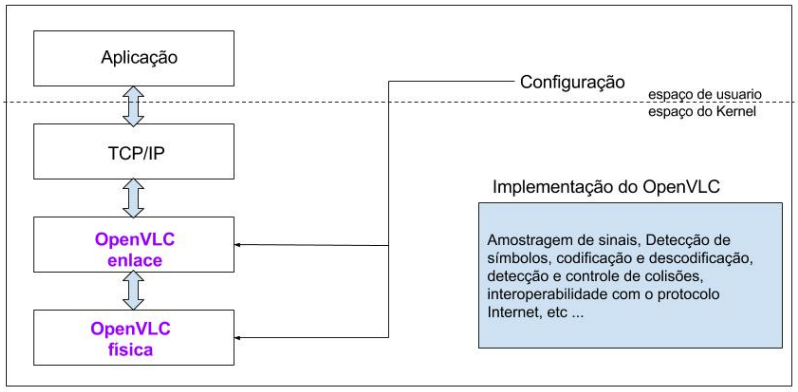


Figura 6.18: Componentes de software do sistema OpenVLC. [Wang et al., 2015b]

Na camada física estão implementados *On-off keying* [McMillin, 2006] para modulação e *Manchester* [Forster, 2000] para codificação. Para a correção de erros, é utilizado *Reed Solomom* [Wicker e Bhargava, 1999]. Na camada de enlace são utilizados CSMA/CD [Hortensius e Winbom, 1999] e CSMA/CD-HA [Wang e Giustiniano, 2014b] para controle de acesso ao meio.

O *software* é carregado na máquina através de um módulo para o Kernel do Linux que intermedia a comunicação entre a pilha de rede do Linux e a placa de expansão. Isto permite o acesso a um conjunto de *softwares* e recursos disponíveis no ambiente Linux, facilitando o uso e monitoramento de redes. Assim, fica claro como a OpenVLC se difere de outras ferramentas para comunicação com luz visível. Suas principais vantagens estão na facilidade de uso propiciada pela integração com a pilha de redes do Linux, seu baixo custo e sua flexibilidade.

6.7.2. Instalação da OpenVLC

O BBB executa uma versão do sistema operacional Debian. Para instalar a plataforma OpenVLC é necessário, primeiramente, adicionar ao sistema Debian o Xenomai [Gerum, 2004], um *framework* para desenvolvimento de sistemas de tempo real.

Existem dois métodos para instalar o Xenomai, que seguem listados abaixo:

1. Instale o Xenomai manualmente e baixe os arquivos do OpenVLC.
   - Compartilhe a internet do *host*[16] com o BBB através da conexão USB:
     - Com o BBB conectado a entrada USB, insira o seguinte comando para acessá-lo:

```
ssh 192.168.7.2 -l root
```

     - No console do BBB insira os comandos a seguir para configurar o IP e a rota, respectivamente:

```
ifconfig usb0 192.168.7.2
route add default gw 192.168.7.1
```

     - No *host* insira as configurações abaixo, responsáveis pelo compartilhamento da internet.
     - Verifique se as interfaces de rede do *host* são as mesmas do exemplo abaixo (wlan0 e eth5) e, caso necessário, modifique-as de acordo com as suas configurações.

```
ifconfig eth5 192.168.7.1
iptables --table nat --append POSTROUTING --out-interface wlan0 -j MASQUERADE
iptables --append FORWARD --in-interface eth5 -j ACCEPT
echo 1 > /proc/proc/sys/net/ipv4/ip_forward
```

   - Instale o Xenomai no BBB através do repositório abaixo:

```
sudo apt-get update
sudo apt-get install linux-image-3.8.13-xenomai-r72
sudo apt-get install linux-headers-3.8.13-xenomai-r72
sudo apt-get install linux-firmware-image-3.8.13-xenomai-r72
sudo apt-get install  linux-headers-3.8.13-bone79
cd xenomai-2.6.3
./configure
make
sudo make install
```

   - Baixe a última versão do OpenVLC[17] e copie os arquivos para o BBB.

[16] A expressão *host* será usada para se referir à máquina a qual o BBB está conectado.
[17] Disponível em `https://github.com/openvlc/openvlc`

2. Use uma imagem do sistema operacional já preparado.

   - Faça o download da imagem do sistema operacional[18].
   - Descompacte o arquivo.
   - Formate um cartão SD de pelo menos 4GB.
   - Copie os arquivos descompactados para o cartão SD.
   - Insira o cartão de memória no BBB.
   - Ligue a placa usando o conector USB.
   - O LED0[19] deve começar a piscar em um padrão regular.
   - Aguarde até que o LED0 pare de piscar.
   - Desligue o conector USB.
   - Retire o cartão de memória.
   - Ligue o BBB.
   - O acesso deve ser feito usando o usuário "openvlc"e a senha "openvlc".

Uma vez que o Xenomai estiver instalado e o código fonte do OpenVLC estiver copiado para o diretório OpenVLC1.1 do BBB é possível prosseguir com a instalação do driver:

- O código fonte do driver se encontra nos arquivos "openvlc.h"e "openvlc.c". Acesse o diretório onde estes arquivos estão localizados e insira os seguintes comandos para compilar o driver:

  ```
  make clean
  make
  ```

- Será gerado um arquivo chamado "vlc.ko". Este é o driver que compõe a plataforma OpenVLC.

- O driver pode ser inserido pelo comando *insmod* como no exemplo abaixo:

  ```
  insmod vlc.ko  frq=10
  ```

  A frequência é a taxa de amostragem utilizada na camada física expressa em KHz. A frequência máxima suportada pelo OpenVLC é de 50KHz.

- Uma interface de rede chamada vlc0 será gerada após inserir o driver.

[18] Disponível em `https://drive.google.com/open?id=0By9qjX2_K-rLNDMtUVFWMFFuR3c`

[19] O LED0 é um dos LEDs presentes no BBB. Ele estará identificado na placa com este rótulo.

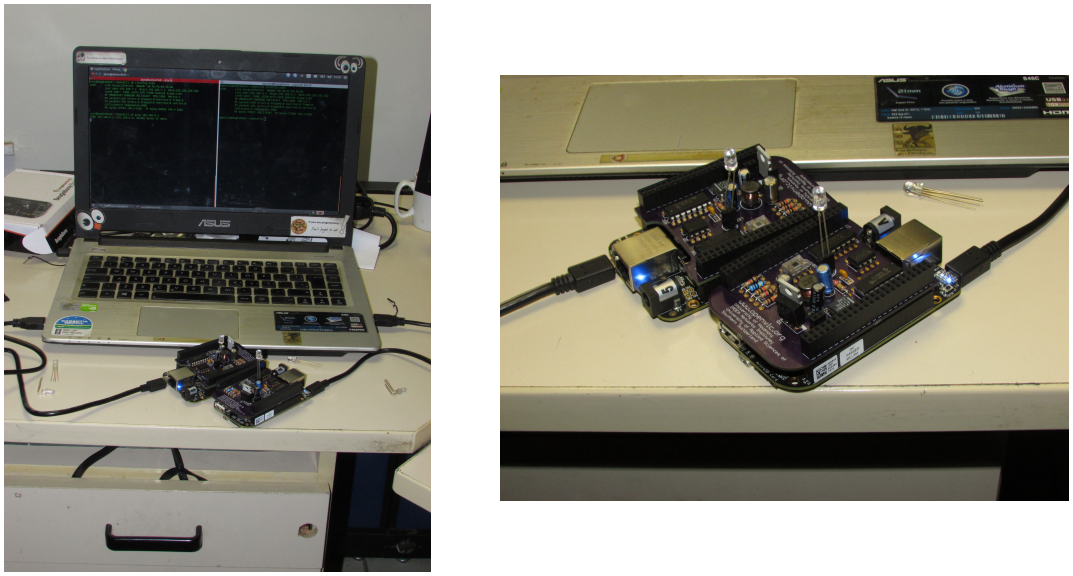


Figura 6.19: Ambiente de testes

6.7.3. Experimentos

Para realizar os experimentos descritos a seguir serão necessários dois BBBs, configurados conforme descrição anterior.

A fim de facilitar o acesso individual aos BBBs deve-se, inicialmente, modificar o IP de um dos BBBs:

- Acesse o BBB pelo seguinte comando:

```
ssh openvlc@192.168.7.2
```

- Altere as configurações do arquivo */etc/network/imterfaces* conforme abaixo.

```
iface usb0 inet static
    address 192.168.8.2
    netmask 255.255.255.252
    network 192.168.8.0
    gateway 192.168.8.1
```

- Reinicie o BBB.

Testando latência, taxa de transferência e conexão

- Conecte ambos os BBBs ao computador através das portas USB como pode ser visto na figura 6.19.
- O BBB de IP 192.168.7.2 será aqui referido como **servidor**, enquanto que o BBB de IP 192.168.8.2 será o **cliente**.

- Acesse o diretório do OpenVLC em ambos os BBBs.
- Para testar a latência:
  - No servidor execute o script *load_driver_server.sh*.
  - No cliente execute o script *load_driver_client.sh* e o comando abaixo:

    ```
    ping 192.168.8.1
    ```

  - Se o servidor estiver ativo, o cliente começará a receber respostas com as seguintes informações:
    - Endereço IP;
    - Número de *bytes* enviados;
    - Tempo gasto em milissegundos;
    - "TTL"ou "Time to Live".
  - Quanto maior for tempo em milissegundos, maior a latência da conexão.
- Para testar a taxa de transferência:
  - No servidor execute o script *iperf_server.sh*.
  - No cliente execute o script *iperf_server.sh*.
  - Os resultados serão exibidos no servidor.
  - Após a realização dos testes a saída na tela deverá estar semelhante a mostrada na figura 6.20

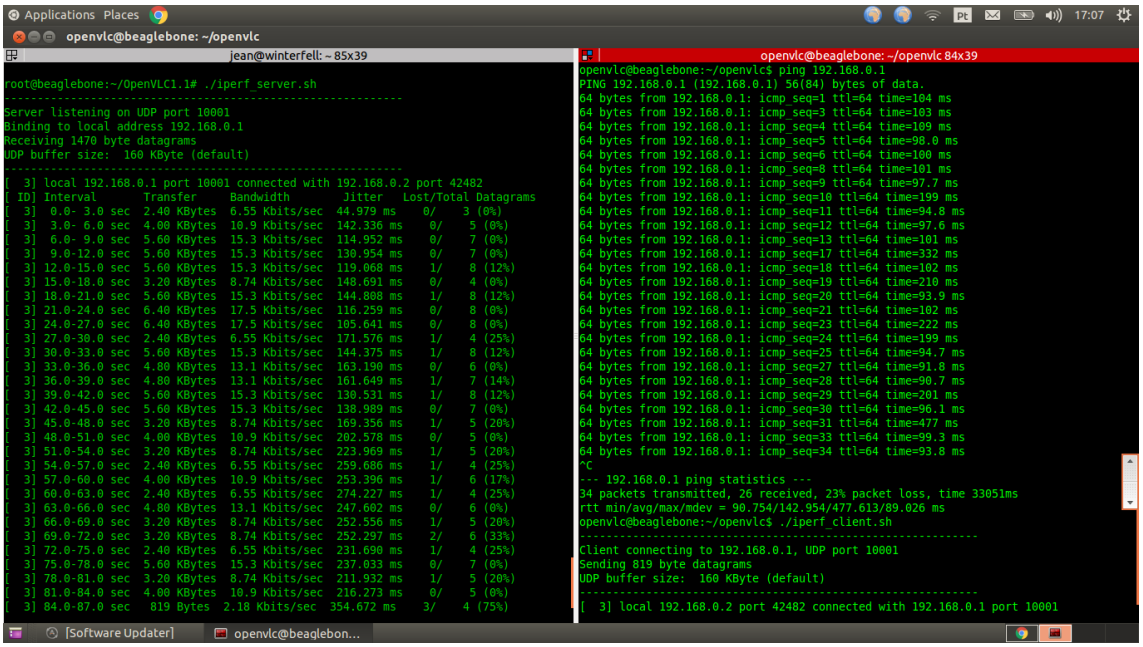


Figura 6.20: Saída de uma execução padrão.

- Para alternar os receptores (LED e fotodiodo) e transmissores (LED comum e LED de alta potência), os seguintes passos devem ser seguidos:
  - Remova o driver usando o comando abaixo:

    ```
    rmmod  vlc.ko
    ```

  - Modifique o arquivo "load_driver_server.sh" e o arquivo "load_driver_client.sh".
    - Para alternar os receptores: modifique a linha "echo X > /proc/vlc/rx". Use 'X'=0 para LED; e use 'X'=1 para fotodiodo;
    - Para alternar os transmissores: modifique a linha "echo X > /proc/vlc/tx". Use 'X'=0 para usar o LED comum; e use 'X'=1 para usar o LED de alta potência.
  - Repita os passos anteriores para realizar os experimentos (vide Seção 6.7.3).

6.8. Conclusões e discussões finais

Este minicurso apresentou os conceitos, aplicações e desafios envolvidos na Comunicação por Luz Visível, através de uma pesquisa bibliográfica extensa e compreensível, de forma a trazer aos participantes uma perspectiva da área de estudo, uma noção da evolução das tecnologias relacionadas e uma visão das possibilidades para o futuro de VLC.

VLC é, sem dúvidas, uma grande oportunidade para o cenário de comunicação sem fio do futuro. Hoje em dia, a demanda por recursos em redes sem fio é altíssima, e a tendência é que continue crescendo, exponencialmente. A popularização de *smartphones* e outros dispositivos *wireless* são exemplos de fatores que alavancam este aumento. Esta perspectiva levanta uma série de questionamentos relacionados à atual infraestrutura de redes sem fio. Neste sentido, figura a possível crise no espectro WiFi, em que a demanda por recursos passa a ser maior do que a capacidade oferecida pela rede.

Entre as alternativas propostas nos últimos anos para complementar a atual infraestrutura de redes sem fio, VLC se destaca, devido às grandes vantagens que proporciona. Espectro livre, altas frequências, disponibilidade de infraestrutura e lâmpadas LEDs são aspectos que chamam atenção na luz visível. Entretanto, muitos obstáculos ainda impedem que tecnologias e aplicações baseadas em VLC sejam comercializadas. Níveis de luminosidade, *uplink* e interferências são problemas que se destacam. Evidentemente, a comunidade acadêmica tem se esforçado na busca de soluções para estas condições.

A Comunicação por Luz Visível oferece uma grande oportunidade para complementação da atual infraestrutura de redes sem fio, pois oferece aumento de performance principalmente em ambientes como escritórios e residências, em que a distância é curta. Além disso, sistemas de localização interna, comunicação aquática e veicular são alguns exemplos de aplicações que podem fazer uso da luz visível para trazer melhorias.

De forma geral, o tema VLC é muito amplo e envolve diversas linhas de pesquisa, atraindo grande interesse por parte da indústria. Ainda assim, a área carece de maior exploração, o que deve acontecer nos próximos anos, considerando a popularização do tema e a crescente adoção de conceitos como Internet das Coisas e Lâmpadas Inteligentes.

Referências

[IEE, 2011] (2011). Ieee standard for local and metropolitan area networks–part 15.7: Short-range wireless optical communication using visible light. *IEEE Std 802.15.7-2011*, pages 1–309.

[Akanegawa et al., 2001] Akanegawa, M., Tanaka, Y., e Nakagawa, M. (2001). Basic study on traffic information system using led traffic lights. *IEEE Transactions on Intelligent Transportation Systems*, 2(4):197–203.

[Arnon, 2015] Arnon, S. (2015). *Visible light communication*. Cambridge University Press.

[Ayyash et al., 2016] Ayyash, M., Elgala, H., Khreishah, A., Jungnickel, V., Little, T., Shao, S., Rahaim, M., Schulz, D., Hilt, J., e Freund, R. (2016). Coexistence of wifi and lifi toward 5g: Concepts, opportunities, and challenges. *IEEE Communications Magazine*, 54(2):64–71.

[Azhar et al., 2013] Azhar, A. H., Tran, T., e O'Brien, D. (2013). A gigabit/s indoor wireless transmission using mimo-ofdm visible-light communications. *IEEE Photonics Technology Letters*, 25(2):171–174.

[Azhar et al., 2010] Azhar, A. H., Tran, T.-A., e O'Brien, D. (2010). Demonstration of high-speed data transmission using mimo-ofdm visible light communications. In *IEEE GLOBECOM Workshops (GC Wkshps)*, pages 1052–1056.

[Baylis et al., 2014] Baylis, C., Fellows, M., Cohen, L., e Marks II, R. J. (2014). Solving the spectrum crisis: Intelligent, reconfigurable microwave transmitter amplifiers for cognitive radar. *IEEE Microwave Magazine*, 15(5):94–107.

[Bell, 1880] Bell, A. G. (1880). The photophone. *Journal of the Franklin Institute*, 110(4):237–248.

[Boroson et al., 2012] Boroson, D. M., Robinson, B., Burianek, D., Murphy, D., e Biswas, A. (2012). Overview and status of the lunar laser communications demonstration. In *SPIE*, volume 8246, pages 82460C–82460C.

[Burchardt et al., 2014] Burchardt, H., Serafimovski, N., Tsonev, D., Videv, S., e Haas, H. (2014). Vlc: Beyond point-to-point communication. *IEEE Communications Magazine*, 52(7):98–105.

[Chan, 2006] Chan, V. W. (2006). Free-space optical communications. *Journal of Lightwave Technology*, 24(12):4750–4762.

[Cheng et al., 2016] Cheng, L., Tsai, H.-M., Viriyasitavat, W., e Boban, M. (2016). Comparison of radio frequency and visible light propagation channel for vehicular communications. In *Proc. of ACM International Workshop on Smart, Autonomous, and Connected Vehicular Systems and Services*, pages 66–67.

[Classen et al., 2016] Classen, J., Steinmetzer, D., e Hollick, M. (2016). Opportunities and pitfalls in securing visible light communication on the physical layer. In *Proc. of ACM Workshop on Visible Light Communication Systems*, pages 19–24.

[Coley, 2013] Coley, G. (2013). Beaglebone black system reference manual. *Texas Instruments, Dallas*.

[Cossu et al., 2012] Cossu, G., Khalid, A., Choudhury, P., Corsini, R., e Ciaramella, E. (2012). 3.4 gbit/s visible optical wireless transmission based on rgb led. *Optics express*, 20(26):B501–B506.

[Cui et al., 2010] Cui, K., Chen, G., Xu, Z., e Roberts, R. D. (2010). Line-of-sight visible light communication system design and demonstration. In *IEEE CSNDSP*.

[Cui et al., 2012] Cui, K., Chen, G., Xu, Z., e Roberts, R. D. (2012). Traffic light to vehicle visible light communication channel characterization. *Applied optics*, 51(27).

[Dambul et al., 2011] Dambul, K. D., O'Brien, D. C., e Faulkner, G. (2011). Indoor optical wireless mimo system with an imaging receiver. *IEEE photonics technology letters*, 23(2):97–99.

[Dang e Zhang, 2012] Dang, J. e Zhang, Z. (2012). Comparison of optical ofdm-idma and optical ofdma for uplink visible light communications. In *IEEE WCSP*, pages 1–6.

[De Vries et al., 2014] De Vries, J. P., Simić, L., Achtzehn, A., Petrova, M., e Mähönen, P. (2014). The wi-fi "congestion crisis": Regulatory criteria for assessing spectrum congestion claims. *Telecommunications Policy*, 38(8):838–850.

[Dilhac, 2001] Dilhac, J. (2001). The telegraph of claude chappe-an optical telecommunication network for the xviiith century.

[Dimitrov e Haas, 2015] Dimitrov, S. e Haas, H. (2015). *Principles of LED Light Communications: Towards Networked Li-Fi*. Cambridge University Press.

[Duque et al., 2016] Duque, A., Stanica, R., Rivano, H., e Desportes, A. (2016). Unleashing the power of led-to-camera communications for iot devices. In *Proceedings of the 3rd Workshop on Visible Light Communication Systems*, pages 55–60. ACM.

[Elgala et al., 2011] Elgala, H., Mesleh, R., e Haas, H. (2011). Indoor optical wireless communication: potential and state-of-the-art. *IEEE Communications Magazine*, 49(9).

[Forster, 2000] Forster, R. (2000). Manchester encoding: opposing definitions resolved. *Engineering Science and Education Journal*, 9(6):278–280.

[Garratt, 1994] Garratt, G. R. M. (1994). *TheEarly History of Radio: From Faraday to Marconi*. Number 20. Iet.

[Gerum, 2004] Gerum, P. (2004). Xenomai-implementing a rtos emulation framework on gnu/linux. *White Paper, Xenomai*, page 81.

[Gfeller e Bapst, 1979] Gfeller, F. R. e Bapst, U. (1979). Wireless in-house data communication via diffuse infrared radiation. *Proceedings of the IEEE*, 67(11):1474–1486.

[Gomez et al., 2015] Gomez, A., Shi, K., Quintana, C., Sato, M., Faulkner, G., Thomsen, B. C., e O'Brien, D. (2015). Beyond 100-gb/s indoor wide field-of-view optical wireless communications. *IEEE Photon. Technol. Lett.*, 27(4):367–370.

[Grobe et al., 2013] Grobe, L., Paraskevopoulos, A., Hilt, J., Schulz, D., Lassak, F., Hartlieb, F., Kottke, C., Jungnickel, V., e Langer, K.-D. (2013). High-speed visible light communication systems. *IEEE Communications Magazine*, 51(12):60–66.

[Grubor et al., 2007] Grubor, J., Jamett, O., Walewski, J., Randel, S., e Langer, K.-D. (2007). High-speed wireless indoor communication via visible light. *ITG-Fachbericht-Breitbandversorgung in Deutschland-Vielfalt für alle?*

[Haas, 2013] Haas, H. (2013). High-speed wireless networking using visible light. *SPIE Newsroom*.

[Haas, 2015] Haas, H. (2015). Visible light communication. In *Optical Fiber Communication Conference*. Optical Society of America.

[Haas et al., 2016] Haas, H., Yin, L., Wang, Y., e Chen, C. (2016). What is lifi? *Journal of Lightwave Technology*, 34(6):1533–1544.

[Hansen, 2011] Hansen, C. J. (2011). Wigig: Multi-gigabit wireless communications in the 60 ghz band. *IEEE Wireless Communications*, 18(6).

[Haruyama, 2013] Haruyama, S. (2013). Visible light communication using sustainable led lights. In *ITU Kaleidoscope: Building Sustainable Communities (K-2013), 2013 Proceedings of*, pages 1–6. IEEE.

[Hewage et al., 2016] Hewage, K., Varshney, A., Hilmia, A., e Voigt, T. (2016). modbulb: a modular light bulb for visible light communication. In *Proceedings of the 3rd Workshop on Visible Light Communication Systems*, pages 13–18. ACM.

[Heydariaan et al., 2016] Heydariaan, M., Yin, S., Gnawali, O., Puccinelli, D., e Giustiniano, D. (2016). Embedded Visible Light Communication: Link Measurements and Interpretation . In *Proceedings of the MadCom: New Wireless Communication Paradigms for the Internet of Things Workshop (MadCom 2016)*.

[Hortensius e Winbom, 1999] Hortensius, P. D. e Winbom, H. B. (1999). Transceiver for extending a csma/cd network for wireless communication. US Patent 5,917,629.

[Hou et al., 2015] Hou, R., Chen, Y., Wu, J., e Zhang, H. (2015). A brief survey of optical wireless communication. In *Proc. Australas. Symp. Parallel Distrib. Comput.(AusPDC 15)*, volume 163, pages 41–50.

[Hou et al., 2013] Hou, W., Jarosz, E., Woods, S., Goode, W., e Weidemann, A. (2013). Impacts of underwater turbulence on acoustical and optical signals and their linkage. *Optics express*, 21(4):4367–4375.

[Hu et al., 2016] Hu, P., Pathak, P. H., Das, A. K., Yang, Z., e Mohapatra, P. (2016). Plifi: hybrid wifi-vlc networking using power lines. In *Proceedings of the 3rd Workshop on Visible Light Communication Systems*, pages 31–36. ACM.

[Jamali et al., 2016] Jamali, M. V., Salehi, J. A., e Akhoundi, F. (2016). Performance studies of underwater wireless optical communication systems with spatial diversity: Mimo scheme. *IEEE Transactions on Communications*.

[Jovicic et al., 2013] Jovicic, A., Li, J., e Richardson, T. (2013). Visible light communication: opportunities, challenges and the path to market. *IEEE Communications Magazine*, 51(12):26–32.

[Karunatilaka et al., 2015] Karunatilaka, D., Zafar, F., Kalavally, V., e Parthiban, R. (2015). Led based indoor visible light communications: State of the art. *IEEE Communications Surveys and Tutorials*, 17(3):1649–1678.

[Kaushal e Kaddoum, 2016] Kaushal, H. e Kaddoum, G. (2016). Underwater optical wireless communication. *IEEE Access*, 4:1518–1547.

[Khalid et al., 2012] Khalid, A., Cossu, G., Corsini, R., Choudhury, P., e Ciaramella, E. (2012). 1-gb/s transmission over a phosphorescent white led by using rate-adaptive discrete multitone modulation. *IEEE Photonics Journal*, 4(5):1465–1473.

[Khalighi e Uysal, 2014] Khalighi, M. A. e Uysal, M. (2014). Survey on free space optical communication: A communication theory perspective. *IEEE Communications Surveys & Tutorials*, 16(4):2231–2258.

[Khan, 2016] Khan, L. U. (2016). Visible light communication: Applications, architecture, standardization and research challenges. *Digital Communications and Networks*.

[Kim et al., 2012] Kim, D.-R., Yang, S.-H., Kim, H.-S., Son, Y.-H., e Han, S.-K. (2012). Outdoor visible light communication for inter-vehicle communication using controller area network. In *Communications and Electronics (ICCE), 2012 Fourth International Conference on*, pages 31–34. IEEE.

[Komine et al., 2003] Komine, T., Haruyama, S., e Nakagawa, M. (2003). Bidirectional visible-light communication using corner cube modulator. *Proc. Wireless and Optical Communication (WOC)*.

[Komine e Nakagawa, 2004] Komine, T. e Nakagawa, M. (2004). Fundamental analysis for visible-light communication system using led lights. *IEEE trans. on Consumer Electronics*, 50(1):100–107.

[Kuo et al., 2014] Kuo, Y.-S., Pannuto, P., Hsiao, K.-J., e Dutta, P. (2014). Luxapose: Indoor positioning with mobile phones and visible light. In *Proceedings of the 20th annual international conference on Mobile computing and networking*, pages 447–458. ACM.

[Lanbo et al., 2008] Lanbo, L., Shengli, Z., e Jun-Hong, C. (2008). Prospects and problems of wireless communication for underwater sensor networks. *Wireless Communications and Mobile Computing*, 8(8):977–994.

[Lee et al., 2011] Lee, K., Park, H., e Barry, J. R. (2011). Indoor channel characteristics for visible light communications. *IEEE Communications Letters*, 15(2):217–219.

[Li et al., 2015] Li, J., Liu, A., Shen, G., Li, L., Sun, C., e Zhao, F. (2015). Retro-vlc: Enabling battery-free duplex visible light communication for mobile and iot applications. In *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications*, pages 21–26. ACM.

[Li et al., 2014] Li, L., Hu, P., Peng, C., Shen, G., e Zhao, F. (2014). Epsilon: A visible light based positioning system. In *NSDI*, pages 331–343.

[Lin et al., 2017] Lin, B., Ye, W., Tang, X., e Ghassemlooy, Z. (2017). Experimental demonstration of bidirectional noma-ofdma visible light communications. *Optics Express*, 25(4):4348–4355.

[Liu et al., 2011] Liu, C. B., Sadeghi, B., e Knightly, E. W. (2011). Enabling vehicular visible light communication (v2lc) networks. In *Proceedings of the Eighth ACM international workshop on Vehicular inter-networking*, pages 41–50. ACM.

[Liu et al., 2007] Liu, H., Darabi, H., Banerjee, P., e Liu, J. (2007). Survey of wireless indoor positioning techniques and systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 37(6):1067–1080.

[Liu et al., 2012] Liu, Y., Yeh, C., Chow, C., Liu, Y., Liu, Y., e Tsang, H. (2012). Demonstration of bi-directional led visible light communication using tdd traffic with mitigation of reflection interference. *Optics express*, 20(21):23019–23024.

[Luo et al., 2014] Luo, P., Ghassemlooy, Z., Le Minh, H., Bentley, E., Burton, A., e Tang, X. (2014). Fundamental analysis of a car to car visible light communication system. In *Communication Systems, Networks & Digital Signal Processing (CSNDSP), 2014 9th International Symposium on*, pages 1011–1016. IEEE.

[Luzhanskiy et al., 2016] Luzhanskiy, E., Edwards, B., Israel, D., Cornwell, D., Staren, J., Cummings, N., Roberts, T., e Patschke, R. (2016). Overview and status of the laser communication relay demonstration. In *SPIE LASE*, pages 97390C–97390C. International Society for Optics and Photonics.

[Marsh e Kahn, 1996] Marsh, G. W. e Kahn, J. M. (1996). Performance evaluation of experimental 50-mb/s diffuse infrared wireless link using on-off keying with decision-feedback equalization. *IEEE Transactions on Communications*, 44(11):1496–1504.

[McMillin, 2006] McMillin, B. K. (2006). On/off keying node-to-node messaging transceiver network with dynamic routing and configuring. US Patent 7,027,773.

[Medina et al., 2015] Medina, C., Zambrano, M., e Navarro, K. (2015). Led based visible light communication: Technology, applications and challenges-a survey. *International Journal of Advances in Engineering & Technology*, 8(4):482.

[Monteiro e Hranilovic, 2014] Monteiro, E. e Hranilovic, S. (2014). Design and implementation of color-shift keying for visible light communications. *Journal of Lightwave Technology*, 32(10):2053–2060.

[Moreira et al., 1997] Moreira, A. J., Valadas, R. T., e de Oliveira Duarte, A. (1997). Optical interference produced by artificial light. *Wireless Networks*, 3(2):131–140.

[O'Brien et al., 2008] O'Brien, D., Le Minh, H., Zeng, L., Faulkner, G., Lee, K., Jung, D., Oh, Y., e Won, E. T. (2008). Indoor visible light communications: challenges and prospects. In *Optical Engineering+ Applications*, pages 709106–709106. International Society for Optics and Photonics.

[Oh, 2013] Oh, M. (2013). A flicker mitigation modulation scheme for visible light communications. In *Advanced Communication Technology (ICACT), 2013 15th International Conference on*, pages 933–936. IEEE.

[Okada et al., 2009] Okada, S., Yendo, T., Yamazato, T., Fujii, T., Tanimoto, M., e Kimura, Y. (2009). On-vehicle receiver for distant visible light road-to-vehicle communication. In *Intelligent Vehicles Symposium, 2009 IEEE*, pages 1033–1038. IEEE.

[Oubei et al., 2015] Oubei, H. M., Li, C., Park, K.-H., Ng, T. K., Alouini, M.-S., e Ooi, B. S. (2015). 2.3 gbit/s underwater wireless optical communications using directly modulated 520 nm laser diode. *Optics express*, 23(16):20743–20748.

[Papadimitratos et al., 2009] Papadimitratos, P., De La Fortelle, A., Evenssen, K., Brignolo, R., e Cosenza, S. (2009). Vehicular communication systems: Enabling technologies, applications, and future outlook on intelligent transportation. *IEEE Communications Magazine*, 47(11).

[Pathak et al., 2015] Pathak, P. H., Feng, X., Hu, P., e Mohapatra, P. (2015). Visible light communication, networking, and sensing: A survey, potential and challenges. *ieee communications surveys & tutorials*, 17(4):2047–2077.

[Pohlmann, 2010] Pohlmann, C. (2010). Visible light communication. In *Seminar Kommunikationsstandards in der Medizintechnik*, pages 1–14.

[Prince e Little, 2012] Prince, G. B. e Little, T. D. (2012). A two phase hybrid rss/aoa algorithm for indoor device localization using visible light. In *Global Communications Conference (GLOBECOM), 2012 IEEE*, pages 3347–3352. IEEE.

[Qiu et al., 2016] Qiu, K., Zhang, F., e Liu, M. (2016). Let the light guide us: Vlc-based localization. *IEEE Robotics & Automation Magazine*, 23(4):174–183.

[Rahaim et al., 2011] Rahaim, M. B., Vegni, A. M., e Little, T. D. (2011). A hybrid radio frequency and broadcast visible light communication system. In *GLOBECOM Workshops (GC Wkshps), 2011 IEEE*, pages 792–796. IEEE.

[Roberts et al., 2011] Roberts, R. D., Rajagopal, S., e Lim, S.-K. (2011). Ieee 802.15. 7 physical layer summary. In *GLOBECOM Workshops (GC Wkshps), 2011 IEEE*, pages 772–776. IEEE.

[Rohner et al., 2015] Rohner, C., Raza, S., Puccinelli, D., e Voigt, T. (2015). Security in visible light communication: Novel challenges and opportunities. *Sensors & Transducers*, 192(9):9.

[Rust e Asada, 2012] Rust, I. C. e Asada, H. H. (2012). A dual-use visible light approach to integrated communication and localization of underwater robots with application to non-destructive nuclear reactor inspection. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 2445–2450. IEEE.

[Schanda, 2007] Schanda, J. (2007). *Colorimetry: understanding the CIE system*. John Wiley & Sons.

[Schill et al., 2004] Schill, F., Zimmer, U. R., e Trumpf, J. (2004). Visible spectrum optical communication and distance sensing for underwater applications. In *Proceedings of ACRA*, pages 1–8.

[Schmid et al., 2016a] Schmid, S., Arquint, L., e Gross, T. R. (2016a). Using smartphones as continuous receivers in a visible light communication system. In *Proceedings of the 3rd Workshop on Visible Light Communication Systems*, pages 61–66. ACM.

[Schmid et al., 2015] Schmid, S., Bourchas, T., Mangold, S., e Gross, T. R. (2015). Linux light bulbs: Enabling internet protocol connectivity for light bulb networks. In *Proceedings of the 2nd International Workshop on Visible Light Communications Systems*, pages 3–8. ACM.

[Schmid et al., 2013] Schmid, S., Corbellini, G., Mangold, S., e Gross, T. R. (2013). Led-to-led visible light communication networks. In *Proceedings of the fourteenth ACM international symposium on Mobile ad hoc networking and computing*, pages 1–10. ACM.

[Schmid et al., 2016b] Schmid, S., Richner, T., Mangold, S., e Gross, T. R. (2016b). Enlighting: An indoor visible light communication system based on networked light bulbs. In *Sensing, Communication, and Networking (SECON), 2016 13th Annual IEEE International Conference on*, pages 1–9. IEEE.

[Schmid et al., 2014] Schmid, S., Ziegler, J., Corbellini, G., Gross, T. R., e Mangold, S. (2014). Using consumer led light bulbs for low-cost visible light communication systems. In *Proceedings of the 1st ACM MobiCom workshop on Visible light communication systems*, pages 9–14. ACM.

[Sevincer et al., 2013] Sevincer, A., Bhattarai, A., Bilgi, M., Yuksel, M., e Pala, N. (2013). Lightnets: Smart lighting and mobile optical wireless networks—a survey. *IEEE Communications Surveys & Tutorials*, 15(4):1620–1641.

[Simpson et al., 2012] Simpson, J. A., Hughes, B. L., e Muth, J. F. (2012). Smart transmitters and receivers for underwater free-space optical communication. *IEEE Journal on selected areas in communications*, 30(5):964–974.

[Singh et al., 2013] Singh, R., O'Farrell, T., e David, J. P. (2013). Performance evaluation of ieee 802.15. 7 csk physical layer. In *Globecom Workshops (GC Wkshps), 2013 IEEE*, pages 1064–1069. IEEE.

[Sung et al., 2015] Sung, J.-Y., Yeh, C.-H., Chow, C.-W., Lin, W.-F., e Liu, Y. (2015). Orthogonal frequency-division multiplexing access (ofdma) based wireless visible light communication (vlc) system. *Optics Communications*, 355:261–268.

[Takai et al., 2014] Takai, I., Harada, T., Andoh, M., Yasutomi, K., Kagawa, K., e Kawahito, S. (2014). Optical vehicle-to-vehicle communication system using led transmitter and camera receiver. *IEEE Photonics Journal*, 6(5):1–14.

[Takai et al., 2013] Takai, I., Ito, S., Yasutomi, K., Kagawa, K., Andoh, M., e Kawahito, S. (2013). Led and cmos image sensor based optical wireless communication system for automotive applications. *IEEE Photonics Journal*, 5(5):6801418–6801418.

[Tanaka et al., 2003] Tanaka, Y., Komine, T., Haruyama, S., e Nakagawa, M. (2003). Indoor visible light data transmission system utilizing white led lights. *IEICE transactions on communications*, 86(8):2440–2454.

[Tanenbaum et al., 2003] Tanenbaum, A. S. et al. (2003). Computer networks, 4-th edition. *ed: Prentice Hall*.

[Tian et al., 2016a] Tian, Z., Wright, K., e Zhou, X. (2016a). The darklight rises: visible light communication in the dark: demo. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, pages 495–496. ACM.

[Tian et al., 2016b] Tian, Z., Wright, K., e Zhou, X. (2016b). Lighting up the internet of things with darkvlc. In *Proceedings of the 17th International Workshop on Mobile Computing Systems and Applications*, pages 33–38. ACM.

[Tsonev et al., 2013] Tsonev, D., Videv, S., e Haas, H. (2013). Light fidelity (li-fi): towards all-optical networking. In *SPIE OPTO*, pages 900702–900702. International Society for Optics and Photonics.

[Uysal e Nouri, 2014] Uysal, M. e Nouri, H. (2014). Optical wireless communications—an emerging technology. In *Transparent Optical Networks (ICTON), 2014 16th International Conference on*, pages 1–7. IEEE.

[Wang e Giustiniano, 2014a] Wang, Q. e Giustiniano, D. (2014a). Communication networks of visible light emitting diodes with intra-frame bidirectional transmission. In *Proceedings of the 10th ACM International on Conference on emerging Networking Experiments and Technologies*, pages 21–28. ACM.

[Wang e Giustiniano, 2014b] Wang, Q. e Giustiniano, D. (2014b). Communication networks of visible light emitting diodes with intra-frame bidirectional transmission. In *Proceedings of the 10th ACM International on Conference on emerging Networking Experiments and Technologies*, pages 21–28. ACM.

[Wang e Giustiniano, 2016] Wang, Q. e Giustiniano, D. (2016). Intra-frame bidirectional transmission in networks of visible leds. *IEEE/ACM Transactions on Networking (TON)*, 24(6):3607–3619.

[Wang et al., 2015a] Wang, Q., Giustiniano, D., e Gnawali, O. (2015a). Low-Cost, Flexible and Open Platform for Visible Light Communication Networks . In *Proceedings of the 2nd ACM Workshop on Hot Topics in Wireless (HotWireless 2015)*.

[Wang et al., 2015b] Wang, Q., Giustiniano, D., e Gnawali, O. (2015b). Low-cost, flexible and open platform for visible light communication networks. In *Proceedings of the 2nd International Workshop on Hot Topics in Wireless*, pages 31–35. ACM.

[Wang et al., 2014] Wang, Q., Giustiniano, D., e Puccinelli, D. (2014). Openvlc: software-defined visible light embedded networks. In *Proceedings of the 1st ACM MobiCom workshop on Visible light communication systems*, pages 15–20. ACM.

[Wang e Chi, 2014] Wang, Y. e Chi, N. (2014). Demonstration of high-speed $2\times$ 2 non-imaging mimo nyquist single carrier visible light communication with frequency domain equalization. *Journal of Lightwave Technology*, 32(11):2087–2093.

[Wang et al., 2013a] Wang, Y., Shao, Y., Shang, H., Lu, X., Wang, Y., Yu, J., e Chi, N. (2013a). 875-mb/s asynchronous bi-directional 64qam-ofdm scm-wdm transmission over rgb-led-based visible light communication system. In *Optical Fiber Communication Conference*, pages OTh1G–3. Optical Society of America.

[Wang et al., 2013b] Wang, Y., Wang, Y., Chi, N., Yu, J., e Shang, H. (2013b). Demonstration of 575-mb/s downlink and 225-mb/s uplink bi-directional scm-wdm visible light communication using rgb led and phosphor-based led. *Optics express*, 21(1):1203–1208.

[Wicker e Bhargava, 1999] Wicker, S. B. e Bhargava, V. K. (1999). *Reed-Solomon codes and their applications*. John Wiley & Sons.

[Wilkins et al., 2010] Wilkins, A., Veitch, J., e Lehman, B. (2010). Led lighting flicker and potential health concerns: Ieee standard par1789 update. In *Energy Conversion Congress and Exposition (ECCE), 2010 IEEE*, pages 171–178. IEEE.

[Wu et al., 2014] Wu, S., Wang, H., e Youn, C.-H. (2014). Visible light communications for 5g wireless networking systems: from fixed to mobile communications. *IEEE Network*, 28(6):41–45.

[Xu et al., 2016] Xu, J., Kong, M., Lin, A., Song, Y., Yu, X., Qu, F., Han, J., e Deng, N. (2016). Ofdm-based broadband underwater wireless optical communication system using a compact blue led. *Optics Communications*, 369:100–105.

[Yang et al., 2014] Yang, S.-H., Kim, H.-S., Son, Y.-H., e Han, S.-K. (2014). Three-dimensional visible light indoor localization using aoa and rss with multiple optical receivers. *Journal of Lightwave Technology*, 32(14):2480–2485.

[Yin e Gnawali, 2016] Yin, S. e Gnawali, O. (2016). Towards Embedded Visible Light Communication Robust to Dynamic Ambient Light. In *Proceedings of the IEEE Global Communications Conference (GLOBECOM 2016)*.

[Yoo et al., 2016] Yoo, J.-H., Jang, J.-S., Kwon, J., Kim, H.-C., Song, D.-W., e Jung, S.-Y. (2016). Demonstration of vehicular visible light communication based on led headlamp. *International journal of automotive technology*, 17(2):347–352.

[Zhang et al., 2015] Zhang, J., Zhang, X., e Wu, G. (2015). Dancing with light: Predictive in-frame rate selection for visible light networks. In *Computer Communications (INFOCOM), 2015 IEEE Conference on*, pages 2434–2442. IEEE.

[Zhu e Kahn, 2002] Zhu, X. e Kahn, J. M. (2002). Free-space optical communication through atmospheric turbulence channels. *IEEE Transactions on communications*, 50(8):1293–1300.

Realização

Apoio Fomento

Apoio Institucional

Patrocinador Diamante

Núcleo de Informação e Coordenação do Ponto BR

Comitê Gestor da Internet no Brasil

Patrocinador Ouro

Patrocinador Bronze