Capítulo

4

Aspectos de Segurança na Interconexão de Redes Celulares e WLANs

Silas Leite Albuquerque, Paulo Roberto de Lira Gondim e Cláudio de Castro Monteiro

Departamento de Engenharia Elétrica, Faculdade de Tecnologia, Universidade de Brasília

Abstract

Wireless communication is extremely present in our everyday lives. Among the standards currently on the market two stand out: the cellular networks and WLAN. These standards are considered complementary from the moment that one enables high transmission rates and other large coverage areas. Thus, the integration of these patterns is very attractive. However this is not straightforward and presents many challenges. One in particular regards to security aspects considered in a transition between two networks of different standards (inter-technology handover). In this context, this work explores exactly the aspects of security (more precisely the authentication and authorization) observed during a handover between WLANs and cellular networks (focus on the 2G and 3G).

Resumo

A comunicação sem fio é algo extremamente presente no nosso cotidiano. Dentre os padrões existentes atualmente no mercado destacam-se dois: as redes celulares e as WLAN. Esses padrões são considerados complementares a partir do momento em que um viabiliza altas taxas de transmissão e outro, grandes áreas de cobertura. Dessa forma, a integração desses padrões é algo muito atrativo. Entretanto isso não simples e apresenta muitos desafios. Um em especial diz respeito aos aspectos de segurança considerados em uma transição entre duas redes de padrões distintos (handover inter-tecnologias). Nesse contexto o presente trabalho explora exatamente os aspectos de segurança (mais precisamente a autenticação e a autorização) observados durante um handover entre WLANs e redes celulares (foco nas redes 2G e 3G).

4.1. Aspectos Introdutórios

As redes de comunicação sem fio (*wireless*) são uma realidade inequívoca nos dias de hoje. Seguindo diferentes padrões e técnicas, elas existem em quase todos os lugares por onde passamos no nosso cotidiano. São WWANs (*Wireless Wide Area Network* – Redes Sem Fio de Área Ampla), WMANs (*Wireless Metropolitan Area Networks* - Redes Sem Fio de Área Metropolitana), WLANs (*Wireless Local Area Networks* - Redes Sem Fio de Área Local) e até WPANs (*Wireless Personal Area Network* - Redes Sem Fio de Área Pessoal) espalhadas por todos os lugares. Essas redes englobam residências, ambientes de trabalho, ruas, *shoppings centers*, aeroportos, rodoviárias, estações de metrô, *cyber-cafés*, cidades e áreas de campo, enfim, quase todos os lugares habitados ou pelos quais o ser humano passa.

Do conjunto de redes sem fio existentes destacam-se dois tipos: as redes móveis celulares e as WLAN baseadas no padrão [IEEE 802.11]. Esses tipos são, de certa forma, complementares, pois quando são comparados, percebe-se que o primeiro fornece ampla área de cobertura com taxa de transmissão reduzida e o segundo viabiliza altas taxas de transmissão em pequenas áreas de cobertura.

Dessa forma, a interconexão entre essas tecnologias mostra-se algo promissor e tem sido alvo de amplos estudos na comunidade acadêmica.

Nesse contexto, também a continuidade de conexão e de serviço em redes sem fio móveis vem se tornando uma necessidade latente dos usuários, que cada vez mais exigem ubiqüidade em seus acessos a serviços considerados críticos (voz e vídeo, por exemplo).

As redes celulares tem sido uma opção importante nesse cenário, considerando sua evolução e seu recente suporte a serviços comutados por pacotes e a altas taxas de transmissão.

Por outro lado, como dito anteriormente, as redes sem fio do tipo WLAN têm sido uma alternativa de baixo custo, oferecendo taxas de transmissão bem mais elevadas, típicas desse tipo de rede.

A questão então é integrar essas duas redes, de forma que a conexão do usuário e o serviço oferecido a ele possam ter continuidade independente da rede a qual o usuário esteja enlaçado.

Para isso, questões inerentes ao controle de acesso ao meio, às arquiteturas, tecnologias e protocolos para a integração WLAN-3G vem sendo estudadas. No entanto, todos esses estudos passam por um ponto comum: o suporte à segurança da informação.

Paradoxalmente aliado a esse aspecto, as tentativas de uso de arquiteturas e tecnologias para a integração de redes focam suas atenções na redução do tempo envolvido com os processos de transferência do móvel de uma rede para outra.

Como solucionar esse problema? De um lado a necessidade de diminuir o tempo de um *handover* com o objetivo de viabilizar a manutenção de uma sessão de uma

determinada aplicação. Do outro lado a necessidade de realizar um *handover* de forma segura, pautado em processos de autenticação e autorização confiáveis e que certamente geram um consumo de tempo relevante.

São muitos os aspectos que devem ser considerados para a resolução desse binômio formado por partes aparentemente contraditórias.

Nesse contexto, focalizando especificamente esses desafios, o objetivo geral do curso é apresentar aspectos teóricos e práticos envolvidos no fornecimento de segurança para a interconexão (*handover*) entre as redes do padrão [IEEE 802.11] e as redes móveis celulares (focalizando 2ª e 3ª gerações). Além disso, também têm-se a intenção de explorar algumas formas de diminuir o custo temporal total de processos de segurança com o objetivo de minimizar o impacto sobre as sessões de aplicações em execução.

Cabe salientar que serão priorizados os problemas de segurança que envolvem os mecanismos de autenticação e autorização entre as partes envolvidas na interconexão, além de aspectos ligados ao gerenciamento de chaves criptográficas. Serão apresentados alguns protocolos e propostas aplicados a diversas situações onde a autenticação e a autorização são necessárias.

Para atingir esses objetivos, o restante do texto deste curso está organizado da seguinte forma:

- A seção 4.2 abordará assuntos ligados à garantia de confiança entre as diversas partes envolvidas em um processo de comunicação de dados baseado em redes sem fio. Serão feitas algumas considerações gerais e será descrito um modelo e níveis de confiança que devem existir para que as partes possam ser consideradas confiáveis frente às outras entidades componentes do processo.
- A seção 4.3 versará sobre os serviços de AAA (*Authentication, Authorization and Accounting* – Autenticação, Autorização e Contabilização) e focalizará dois importantes protocolos para AAA que estão relacionados ao problema abordado neste curso: o RADIUS (*Remote Authentication Dial In User Service* – Serviço de Autenticação e Contabilização Remota de Usuários) e o Diameter.
- A seção 4.4 apresentará o protocolo EAP (*Extensible Authentication Protocol* – Protocolo de Autenticação Estensível), que, atuando em nível de enlace, é a base que viabiliza os processos de autenticação que envolvem as WLANs e redes celulares de 2ª e 3ª gerações. Também serão abordados alguns de seus métodos criados posteriormente com o objetivo de resolver problemas específicos de determinados ambientes e tecnologias. Será tratado mais detalhadamente, além do próprio *framework* EAP, o ERP (*EAP Re-authentication Protocol* – Protocolo de re-autenticação EAP), pelo fato de ser um protocolo ligado diretamente ao processo de re-autenticação que é um dos focos de um processo de *handover*.
- Na seção 4.5, semelhante àquilo que foi mostrado na seção anterior, serão apresentados os principais protocolos de autenticação utilizados em redes celulares de 2ª e 3ª gerações, o EAP-SIM (*EAP Method for Global System for Mobile Communications Subscriber Identity Module* – Método EAP para

Módulo de Identidade do Assinante do Sistema Global para Comunicações Móveis) e o EAP-AKA (EAP *Method for 3rd Generation Authentication and Key Agreement* – Método EAP para Autenticação e Acordo de Chave de 3ª Geração), respectivamente.

- A seção 4.6 abordará o gerenciamento de chaves e tratará alguns aspectos gerais com a hierarquia de chaves, distribuição e reuso de material usado para a criação de chaves.
- Na seção 4.7 serão descritas as métricas, as técnicas e as fases do processo de *handover*, caracterizando cada uma delas no contexto das redes envolvidas e destacando os aspectos de segurança observados. Além disso, serão apresentados também os aspectos gerais do padrão IEEE 802.21, que foi proposto para tentar resolver (ou atenuar) o problema dos *handovers* entre redes heterogêneas.
- Na seção 4.8 serão tratados alguns protocolos de autenticação envolvidos nas operações de *handover* inter-tecnologias. Dessa forma, serão apresentados os protocolos MPA (*Media Independent Pre-Authentication* – Pré-autenticação Independente do Meio), que será priorizado tendo em vista sua importância no contexto tratado, o HOKEY (*Handover Keying* – "Chaveamento" de Handover) e o PANA (*Protocol for Carrying Authentication for Network Access* – Protocolo para Transporte de Autenticação para Acesso a Rede).
- A seção 4.9 finalizará o presente trabalho apresentando algumas conclusões

4.2. Garantia de confiança

4.2.1. Premissas e conceitos básicos

Antes de tratar especificamente do modelo adequado ao problema abordado neste trabalho, serão enunciadas algumas definições que viabilizarão um entendimento mais adequado daquilo que está sendo explorado.

A primeira e mais relevante no contexto desta seção diz respeito àquilo que vem a ser **confiança**. A definição clássica encontrada no dicionário Houaiss da Língua Portuguesa é:

> *"...crença na probidade moral, na sinceridade afetiva, nas qualidades profissionais, etc., de outrem, que torna incompatível imaginar um deslize, uma traição, uma demonstração de incompetência de sua parte; crédito, fé..."*

Percebe-se nitidamente que o escritor focaliza, em sua definição, relacionamentos humanos. Assim, apesar de colaborar para o entendimento do termo, o texto ressaltado não é totalmente adequado àquilo que está sendo explorado, afinal as entidades que fazem parte do contexto abordado neste trabalho extrapolam o ser humano, sendo consideradas, de forma resumida, sistemas complexos formados por pessoas, equipamentos (*hardware*) e programas (*software*).

Muitas outras definições existem, e as mais adequadas tratam confiança sob dois enfoques: o da esperança e o da dependência [Josang ET AL, 2007]. No primeiro caso,

confiança está ligada ao fato de uma entidade ter esperança (dar crédito), com certa probabilidade, que outra entidade aja em seu favor (da primeira). No segundo caso, confiança diz respeito às situações nas quais uma entidade está disposta, com certa probabilidade, a depender de outra para atingir algum objetivo. Os dois enfoques são muito próximos e em ambos os casos, quanto maior a probabilidade, maior será a confiança.

Outro conceito importante no contexto desta seção é o de **reputação**. O mesmo dicionário utilizado na definição passada indica:

"... conceito de que alguém ou algo goza num grupo humano..."

Como no caso passado, o escritor focaliza relacionamentos humanos. De forma mais adequada, pode-se entender reputação como o que é geralmente dito ou acreditado (em quantidade e qualidade) sobre uma entidade, ou seja, é algo derivado de um comportamento frente a outras entidades [Josang ET AL, 2007].

Confiança e reputação parecem conceitos proporcionais (boa reputação implica em maior confiabilidade, por exemplo), entretanto isto é um equívoco facilmente percebido quando examinamos as afirmações:

- "Confio em você por causa da sua boa reputação";
- "Confio em você, independente da sua má reputação";
- "Não confio em você, independente da sua boa reputação".

Na primeira afirmação, a confiança deriva diretamente da reputação, entretanto nas duas seguintes, ela deriva de outra coisa que provavelmente é mais relevante que a própria reputação (um conjunto de leis e normas, por exemplo, que previna danos causados a uma entidade - que confia - por outra entidade - em quem se confia – pode ser suficiente para alguém confiar em outrem, independente de sua reputação).

Outro conceito importante para o contexto deste trabalho é o de **modelo de confiança**. Este pode ser entendido como o arcabouço formado por entidades, seus relacionamentos e seus comportamentos, que viabiliza, para as diversas partes interessadas, confiança em níveis adequados para a consecução de seus objetivos. No item 4.2.2 será discutido um modelo de confiança adequado para o problema considerado neste trabalho.

Para que seja possível definir modelos e soluções que forneçam confiabilidade aos processos de comunicações que envolvem, em particular, usuários, redes celulares e WLANs, é necessário preocupar-se com questões do tipo [Koien & Haslestad, 2003]:

- Em que entidades deve-se confiar?
- De que forma deve-se confiar nessas entidades?
- Que tipos de características de segurança são necessárias para justificar confiança?

Essas questões podem ser respondidas de diversas formas, dependendo da situação analisada, e indicarão as premissas básicas que devem ser atendidas por um modelo de confiança que esteja em conformidade com o problema explorado.

4.2.2. Modelo e níveis de confiança

Um modelo de confiança adequado às situações tratadas neste trabalho considera a presença de três entidades básicas [ETSI TS 133 234]: A rede celular, a rede sem fio (WLAN) e o usuário. Salienta-se que para a implementação de soluções específicas, cada caso deve ser analisado separadamente e demanda a geração de modelos de confiança apropriados.

Pode-se definir mais detalhadamente as três entidades básicas da seguinte forma:

- Rede celular – integrante da parte fixa do modelo e formada por todos os componentes da rede de comunicações que provê serviços de telefonia móvel celular, podendo ser composta por uma ou mais operadoras, um ou mais domínios e uma ou mais tecnologias de acesso (GSM e UMTS, por exemplo). Também são consideradas, como parte da rede celular, as outras redes celulares que podem ser acessadas pelo usuário quando em situação de "*roamming*". Note que, apesar de não estarem sendo tratados nesta seção, também deverão existir modelos de confiança internos, entre os diversos integrantes da rede celular (Estações Rádio Base, Centros de Autenticação, Servidores diversos e etc.) para que a rede como um todo seja considerada confiável.
- WLAN – também integrante da parte fixa do modelo, essa entidade é composta por todos os elementos que se encontram "à retaguarda" dos pontos de acesso sem fio (incluindo os próprios) e que fornecem acesso público à rede do padrão [IEEE 802.11]. Tal qual na rede celular, também deverão existir modelos de confiança entre seus elementos para que a WLAN seja considerada confiável.
- Usuário - é elemento móvel do modelo que gera, dessa forma, a necessidade de *handovers*. Considera-se neste trabalho que o usuário é formado pelo próprio assinante das redes celulares e WLANs, pelo equipamento com capacidade de processamento e de comunicação (telefone, PDA, laptop e etc.) e pelos dispositivos de segurança vinculados, particularmente os módulos de identificação do usuário (SIM/USIM/UICC) ligados ao equipamento e que embutem o material criptográfico a ser utilizado na segurança dos processos. Note que para que seja possível a realização de *handovers* inter-tecnologias (usuário saindo de uma WLAN e entrando em uma rede celular, por exemplo), o equipamento com capacidade de comunicação indicado anteriormente deverá ser multi-modo, ou seja, deverá suportar comunicações na rede celular e na WLAN (simultaneamente ou não, dependendo da necessidade).

Cabe salientar que a WLAN e a rede celular podem fazer parte de uma arquitetura muito acoplada, pouco acoplada ou não acoplada, como será visto com mais detalhes na seção 4.7 deste trabalho. Cada tipo de arquitetura gerará peculiaridades à relação de confiança existente entre essas duas entidades.

A Figura 4.1 representa as entidades abordadas e as relações entre elas.

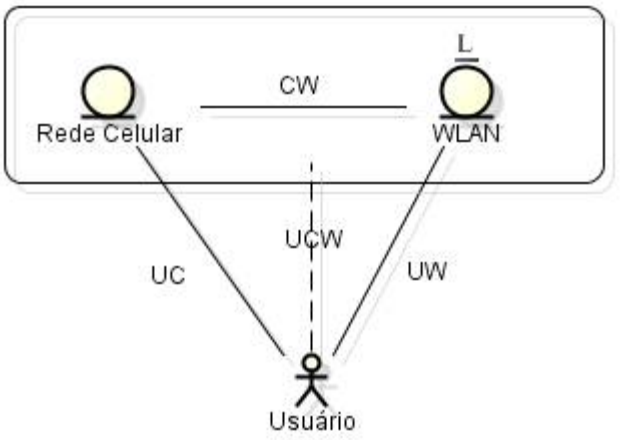

Figura 4.1 – modelo de confiança

As relações de confiança existentes entre as partes envolvidas são representadas, na figura, por CW (relação de confiança existente entre a rede celular e a WLAN), UC (usuário e rede celular), UW (usuário e WLAN) e UCW (usuário, rede celular e WLAN quando estas duas últimas estão muito acopladas). Cada uma possui suas características particulares.

A relação CW está intimamente ligada ao grau de acoplamento entre as entidades. Em uma arquitetura muito acoplada, como a WLAN está praticamente embutida da rede celular (fazer parte da mesma entidade legal), essa relação passa a ser considerada interna à rede celular e é viabilizada por meio dos modelos de confiança internos da rede celular (e WLAN). Nesta situação, as relações UC e UW podem ser resumidas à relação UCW. Já em situações de pouco ou nenhum acoplamento, a relação CW é controlada por meio de tratados formais de parceria entre rede celular e WLAN e é implementada, geralmente, por protocolos de AAA (seção 4.3).

A relação UC é regida formalmente por um acordo de fornecimento de serviços de telefonia móvel celular seguros firmado, no ato da contratação, entre o assinante e a operadora de telefonia (ou operadoras). Entretanto, quem confere confiabilidade prática à relação (particularmente aos processos de comunicação que ocorrem entre as duas partes) são os mecanismos de segurança física e lógica (neste último, ressalta-se o papel da criptografia) desencadeados por ambas as partes e que focalizam certos componentes das entidades (no usuário, o SIM/USIM/UICC, e na rede celular, o centro de autenticação, por exemplo).

A relação UW, semelhante àquilo que ocorre na UC, é criada a partir acordos entre assinantes e provedores de acesso a redes sem fio (um usuário pode ter um contrato com uma empresa que fornece acesso a uma WLAN existente em um aeroporto, por exemplo) e também é viabilizada por meio de processos e protocolos de segurança da informação. Essa relação torna-se bastante complexa a partir do momento em que acessos a WLANs públicas (que não demandam acordos prévios entre o usuário e a rede) passam a ser considerados. Na primeira situação (WLANs contratadas), deve ser papel dos provedores (inerente ao modelo de confiança) viabilizarem segurança inter-usuários, ou seja, evitar, por exemplo, que dados de um usuário possam ser capturados

indevidamente por outros usuários. Já na segunda situação (WLANs públicas), geralmente o provedor abstém-se dessa responsabilidade. Note que essa lacuna no modelo de confiança poderia ser preenchida caso fosse considerada uma relação de confiança entre usuários (UU), o que não será tratado neste trabalho.

Além dos elementos e de suas relações, também é interessante que um modelo de confiança considere os níveis de confiança tolerados entre as diversas partes. No caso em análise, o usuário e a rede celular, desde que autenticados mutuamente, devem ter total confiança um no outro (o que deve ser garantido pelo acordo firmado). Entretanto a WLAN, como única entidade que foi considerada em vários estados (WLANs públicas, contratadas, embutidas na rede celular, etc.), apresenta níveis de confiabilidade distintos. Nesse contexto, e focalizando as relações que consideram a WLAN, percebem-se três níveis de confiança distintos [ETSI TS 133 234]:

1. A WLAN pode ser totalmente não confiável para o usuário e para a rede celular;
2. Alguns elementos da WLAN são confiáveis para o usuário e para a rede celular (alguns servidores em uma arquitetura pouco acoplada, por exemplo), outros elementos não são confiáveis;
3. Todos os elementos da WLAN são totalmente confiáveis para o usuário e a rede celular.

Cada nível de confiança implicará em comportamentos específicos das diversas partes e interferirá diretamente sobre o fornecimento dos serviços para os usuários (determinados serviços que demandam muita segurança não poderão ser viabilizados para os usuários por uma WLAN caso esta seja considerada totalmente não confiável, por exemplo).

4.3.Protocolos Genéricos para AAA

A presente seção versará sobre os serviços de AAA (*Authentication, Authorization and Accounting* - Autenticação, Autorização e Contabilização) e focalizará dois importantes protocolos para AAA que estão relacionados ao problema abordado neste curso: o RADIUS (*Remote Authentication Dial In User Service* – Serviço de Autenticação e Contabilização Remota de Usuários) e o Diameter (diâmetro – passando a idéia de que é duas vezes o raio – RADIUS).

Cabe definir, antes de abordar os protocolos propriamente ditos, o que vem a ser cada um dos serviços de AAA:

- Autenticação – serviço que viabiliza a verificação da autenticidade da identidade de uma entidade, ou seja, constata que ela é quem diz ser;
- Autorização – uma vez identificada a entidade, esse serviço verifica suas prerrogativas (seu direitos em termos de acesso a serviços e recursos) e autoriza o acesso;
- Contabilização – tendo sido autenticada e autorizada, a entidade começa a utilizar o serviço solicitado; essa utilização passa a ser, então, medida ou

contabilizada (em termos de tempo de uso, por exemplo) para fins específicos (cobrança financeira por tempo de uso do serviço, por exemplo).

4.3.1. RADIUS

O RADIUS, caracterizado pela [RFC 2865], foi definido originalmente para prover AAA para sessões SLIP [RFC 1055] e PPP [RFC 1661]. Ele viabiliza uma diminuição de sobrecarga sobre o NAS (*Network Access Server* - Servidor de Acesso à Rede) a partir do momento em que reúne, sob sua responsabilidade (RADIUS), as listas de usuários autorizados e seus respectivos parâmetros de segurança (senhas e etc.) que em situações anteriores eram armazenadas pelos próprios NAS. Esta arquitetura torna possível a criação um banco de dados de usuários centralizado que permite a gerência unificada e o suporte a chamadas vindas de NAS fisicamente distribuídos.

O RADIUS está fundamentado em uma arquitetura cliente/servidor e opera na camada de aplicação utilizando, para troca de mensagens, o protocolo de transporte UDP (*User Datagram Protocol* – Protocolo de Datagramas de Usuário) [RFC 768]. De forma simplificada, o protocolo é utilizado na conferência dos parâmetros "nome de usuário" e "senha". Estas informações, uma vez disponibilizadas para o NAS, são repassadas para o servidor RADIUS (ou servidor de AAA), que verifica a veracidade das informações e libera (ou não) o acesso aos serviços solicitados pelo usuário (caso o acesso seja liberado, também é feita a contabilização do uso dos recursos).

A Figura 4.2 representa uma típica troca de mensagens de autenticação e autorização entre um cliente e um servidor RADIUS (entidades intermediárias são omitidas por questão de simplicidade).

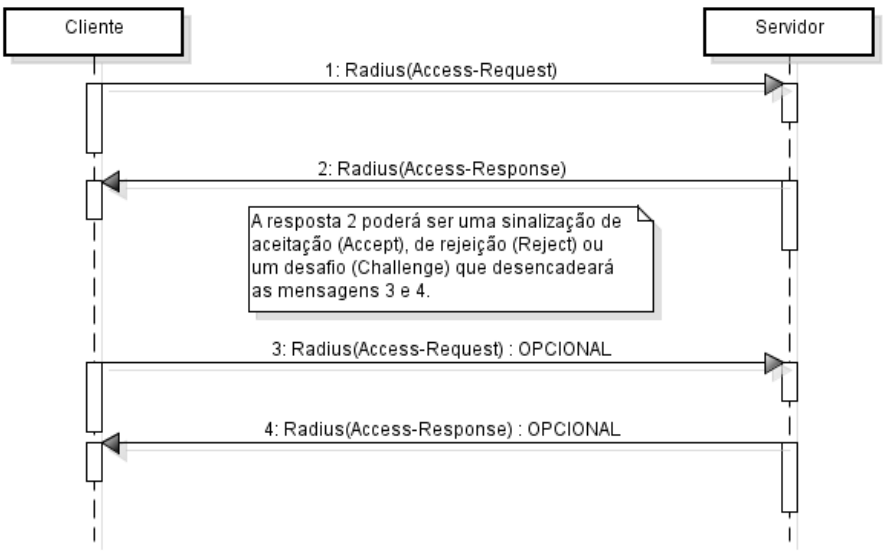

Figura 4.2 – Troca de mensagens de Autenticação e Autorização RADIUS

Quando um usuário solicita uma conexão, o NAS envia uma mensagem de pedido de acesso RADIUS ao servidor de AAA, transmitindo o nome do usuário e a senha ofuscada (um derivado da senha calculado a partir da função *hash* MD5), tipo de conexão (porta), identidade do NAS e o campo autenticador.

Em resposta, o servidor de AAA usa a fonte do pacote, identidade do NAS e o campo autenticador para determinar se o NAS pode enviar requisições de autenticação. Nesse caso, o servidor de AAA tenta achar o nome do usuário em seu banco de dados e aplica a senha ofuscada e outros atributos constantes no pedido de acesso para decidir se deve ser concedido acesso a este usuário.

Dependendo do método de autenticação utilizado, o Servidor de AAA pode responder à solicitação devolvendo uma mensagem-desafio (*access-challenge* – desafio de acesso) que contem, como parâmetro, um valor aleatório. O NAS retransmite o desafio ao usuário remoto que deverá, por sua vez, responder com o valor correto para provar sua identidade (a resposta pode ser, por exemplo, calculada por meio da codificação do desafio utilizando um derivado da senha). Essa resposta é encaminhada para o NAS que retransmite ao servidor de AAA dentro de outra mensagem de solicitação de acesso RADIUS.

Se o servidor de AAA verificar que o usuário é autêntico, ele o autoriza a usar o serviço solicitado (acesso a um recurso de rede, por exemplo) e devolve uma mensagem de permissão de acesso RADIUS. Caso a autenticidade na seja verificada, o Servidor de AAA devolve uma mensagem de rejeição de acesso RADIUS e o NAS desconecta o usuário.

Quando uma mensagem de permissão de acesso é recebida, instantaneamente é iniciado o processo de contabilização RADIUS (não representado na Figura 4.2). Para isso o NAS envia uma mensagem de requisição de contabilização RADIUS para o Servidor de AAA. O Servidor, tendo recebido a requisição, passa a realizar um registro contábil vinculado ao usuário recém autorizado. Ao mesmo tempo, o NAS ativa a sessão desse usuário. Terminando a sessão, uma nova mensagem de requisição de contabilização do RADIUS é enviada pelo NAS para o servidor, o que sinaliza que deverão ser gravadas, nos registros contábeis do servidor, a razão da desconexão e a duração da sessão encerrada. Salienta-se que poderão ser solicitadas alterações antes do término da sessão. Essas solicitações também gerarão mensagens de requisição e de resposta de contabilização.

As mensagens do RADIUS são transportadas por meio de datagramas UDP e possuem os seguintes campos, conforme Figura 4.3: código (tipo da mensagem), identificador (geralmente um número seqüencial), tamanho (quantidade de octetos), autenticador (desafio do servidor RADIUS – *access-challenge* - ou resposta a esse desafio) e atributos (opcionalmente requeridos para tipos específicos de serviços).

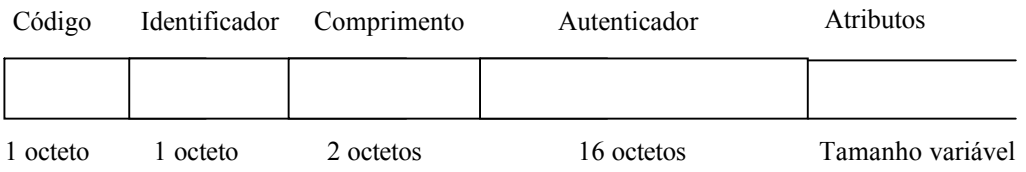

Figura 4.3 – Mensagem RADIUS – formato genérico

Salienta-se que o propósito do campo autenticador é prover segurança às transmissões. Ele pode ser de dois tipos: requisição ou resposta. O autenticador-requisição segue do usuário para o servidor AAA (passando pelo NAS) e contém um valor totalmente aleatório. Já o autenticador-resposta segue no sentido oposto e contém um valor *hash* MD5 calculado a partir do pacote de requisição RADIUS recebido seguido dos atributos de resposta e do segredo compartilhado entre servidor AAA e usuário.

O campo autenticador também fornece ajuda para o servidor AAA descobrir falsificação de respostas RADIUS, além de ser usado para obscurecer a senha do usuário, inibindo revelação ao NAS ou qualquer outra entidade intermediária que poderia bisbilhotar as mensagens RADIUS.

O uso do campo autenticador intimida ataques passivos, mas um intruso que consiga capturar as mensagens de requisição de acesso RADIUS e as mensagens de resposta de acesso pode executar um ataque de dicionário e descobrir a senha. A [RFC 3580] documenta diversas vulnerabilidades, bem como diretrizes de uso para reduzir riscos na utilização do RADIUS.

A filosofia de transmissão, pelo RADIUS, de pares atributo-valor (AVP – *Attribute-Value Pair*) facilita o uso do protocolo com uma grande variedade de tecnologias de acesso e métodos de autenticação. O padrão original definiu vários pares atributo-valor comuns, como usuário-nome, usuário-senha, NAS-IP-endereço, porta NAS e tipo de serviço. Entretanto novos pares atributo-valor podem ser definidos para que sejam criadas extensões proprietárias, o que pode ser observado, por exemplo, na [RFC 2548], que define os pares atributo-valor da Microsoft associados ao protocolo MS-CHAP v1 [RFC 2433].

Além disso, considerando o contexto específico deste trabalho, muitos pares atributo-valor padrão foram definidos para dar suporte ao EAP (seção 4.4.1 deste trabalho), como pode ser constatado na [RFC3579].

4.3.2. Diameter

O Diameter, tal qual o RADIUS, foi criado para prover os serviços de AAA para viabilizar, por exemplo, o controle de acesso a serviços e recursos de redes. Ele está definido na [RFC 3588] e pode ser utilizado tanto em redes locais quanto em ambientes totalmente distribuídos (situações onde o cliente está em situação de *roaming*, por exemplo).

O protocolo é considerado por muitos como o sucessor do RADIUS (apesar de não serem totalmente compatíveis) e foi criado para atender a demandas, no contexto de AAA, que não eram mais atendidas pelo seu antecessor. A tabela a seguir mostra algumas das principais diferenças entre os dois protocolos:

Tabela 4.1– Comparação Diameter X RADIUS (adaptado de [Liu ET AL, 2006]).

Característica analisada	Diameter	RADIUS
Protocolo de transporte	Orientado à conexão (TCP e SCTP)	Não orientado à conexão (UDP)
Segurança	Ponto-a-ponto e fim-a-fim	Ponto-a-ponto
Suporte a agentes	Agentes de retransmissão (*relay*), de procuração (*proxy*), de redirecionamento (*redirect)* e de tradução (*translation*)	Não suporta
Capacidade de negociação	Negocia aplicações suportadas e níveis de segurança	Não suporta
Descobrimento de parceiro	Configuração estática e busca dinâmica	Configuração estática
Mensagem iniciada pelo servidor	Suportada (por exemplo, mensagens de re-autenticação e de término da sessão)	Não suporta
Tamanho máximo dos dados de atributos	16.777.215 octetos	255 octetos

As bases que fundamentaram a criação do Diameter podem ser encontradas na [RFC 2989] que cataloga uma série de requisitos que devem ser atendidos por protocolos de AAA de uma forma geral. Esses requisitos incluem, por exemplo, capacidade de controle de falhas e mecanismos de auditoria que não foram viabilizados por meio da criação do RADIUS mas que foram considerados para o desenvolvimento do Diameter.

O protocolo em si aproveita as idéias fundamentais do RADIUS e acrescenta uma série de mensagens e procedimentos genéricos que permitem o atendimento à maior parte dos requisitos desejados. Uma das grandes evoluções observadas no Diameter está ligada à capacidade de viabilizar interfaces que sejam compatíveis com várias aplicações que necessitem de AAA (que já existem ou que venham a existir). Isso é possível devido à criação de aplicações Diameter que funcionam sobre o protocolo base de AAA e sob as aplicações específicas (camada superior).

O Diameter possui uma arquitetura ponto-a-ponto, o que permite o encadeamento de vários nós para o fornecimento de serviços de AAA. Cada nó (o termo usado é nó Diameter) pode agir, dependendo do contexto, como cliente, servidor ou agente de AAA. Os nós com os dois primeiros papéis agem como seus correspondentes no protocolo antecessor, entretanto último papel (agente) tem características diferentes e pode assumir quatro diferentes configurações:

- retransmissão (*relay*) – agente que recebe requisições e as retransmite para outros nós Diameter;
- procuração (*proxy*) – agente semelhante ao anterior, entretanto, ao invés de retransmitir as próprias requisições recebidas, cria novas mensagens derivadas da requisição e de políticas (controle de admissão, utilização de recursos, etc.) definidas previamente;
- redirecionamento (*redirect*) – parecido com o primeiro, entretanto não retransmite a mensagem recebida, ao invés disso, responde ao remetente com informações suficientes para que este transmita sua mensagem diretamente para o destinatário adequado;
- tradução (*translation*) – agente utilizado na interface com outros protocolos (Diameter-RADIUS, por exemplo).

As mensagens Diameter são síncronas, ou seja, a cada mensagem enviada por um nó deverá ser criada uma correspondente resposta. O formato geral de um pacote Diameter pode ser observado na Figura 4.4:

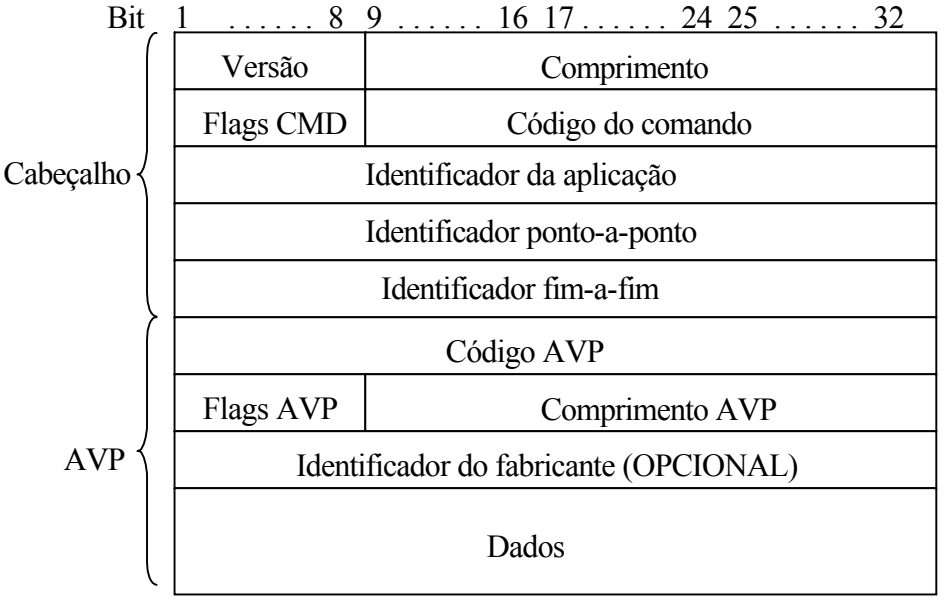

Figura 4.4 – Mensagem Diameter – formato genérico

Os campos são assim definidos:

- Versão – 1 octeto – versão do protocolo Diameter;
- Comprimento – 3 octetos – comprimento de toda a mensagem, em número de octetos, incluindo o próprio tamanho do cabeçalho;

- *Flags* CMD – 1 octeto – 8 bits de controle do comando que podem sinalizar se a mensagem é uma requisição ou uma resposta, se pode ser repassada ou não, se é uma mensagem de erro ou não, etc.;
- Código do comando – 3 octetos – comando associado à mensagem;
- Identificador da aplicação – 4 octetos – identifica a aplicação para qual a mensagem está sendo direcionada;
- Identificador ponto-a-ponto – 4 octetos – auxilia na conferência do sincronismo das mensagens entre os nós (uma resposta deve ter o mesmo identificador que a requisição que a originou, esse identificador, por sua vez, não pode fazer parte de outras respostas ou requisições que transitam entre esses nós);
- Identificador fim-a-fim – 4 octetos – usado para a detecção de mensagens duplicadas entre servidor e cliente (agentes intermediários não podem alterar esse valor);
- Código AVP – 4 octetos – código do par atributo-valor (os códigos de 1 a 255 são reservados para manter certa compatibilidade com o RADIUS);
- *Flags* AVP – 1 octeto – informa ao recebedor como o atributo deve ser manipulado (se há necessidade de cifração fim-a-fim, se é um dado que contem informações do fabricante, etc.);
- Comprimento AVP – 3 octetos – indica o tamanho do par atributo-valor, em número de octetos, incluindo o cabeçalho do AVP (código, *flags*, etc.);
- Identificador do fabricante (OPCIONAL) – 4 octetos – valor que identifica um fabricante específico (necessidade criada pelo fato do Diameter permitir que fabricantes criem seus próprios tipos de AVP, independentes dos definidos pelo próprio padrão);
- Dados – tamanho variável – zero ou mais octetos que contêm as informações específicas do atributo considerado.

Antes de qualquer transmissão ser feita, há a necessidade de se configurar, no NAS, a localização do servidor AAA que deve ser utilizado. Essa configuração pode ser feita manualmente ou pode ser auxiliada por servidores que contêm informações dos diversos nós Diameter existentes em um dado ambiente (domínio, etc.). Esse segundo tipo de configuração é chamado de "descoberta do par", pois ao ser gerada a requisição, a busca pelo nó que poderá atendê-la da melhor forma é feita automaticamente.

Tendo sido feito o direcionamento coerente (par direcionado para par adequado), é estabelecida, então, a conexão em nível de transporte entre os pares, que representa uma ligação entre dois nós Diameter. Essa ligação, como visto na

Tabela 4.1, deve ser baseada em TCP ou SCTP, que são protocolos de transporte mais confiáveis que o UDP, e viabiliza a troca de mensagens.

Além das conexões, o Diameter estabelece sessões entre nós. Diferente das conexões, as sessões são ligações lógicas entre clientes (usuários) e servidores, são

consideradas em nível de aplicação e podem embutir várias conexões. Exemplificando: em sua situação hipotética na qual o cliente esteja separado do servidor por meio de um agente de retransmissão, este agente fará conexões (uma com o cliente e outra com o servidor) que pertencerão à única sessão criada entre o cliente e o servidor.

A troca de mensagens Diameter, como é possível inferir da complexidade da arquitetura até agora vista, não é tão simples quanto a equivalente no RADIUS. Por esse motivo, a [RFC 3588] define uma série de máquinas de estado que representam os procedimentos e trocas de mensagens que devem ser desencadeados para que os serviços de AAA sejam fornecidos a contento.

De forma simplificada, um fluxo de mensagens entre dois nós Diameter quaisquer (cliente-servidor, cliente-agente, agente-agente e agente-servidor) pode ser representado pela Figura 4.5.

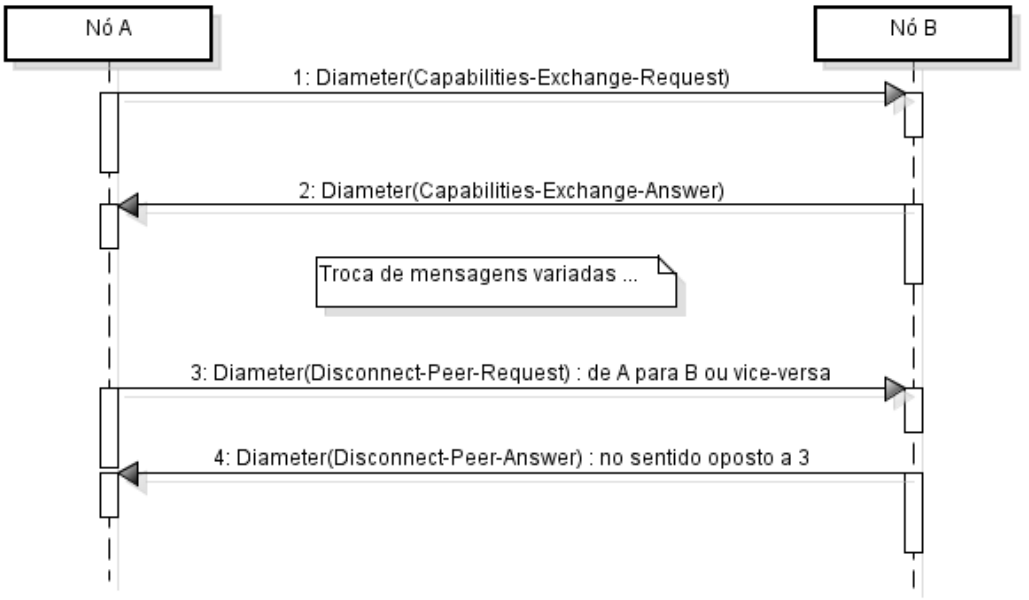

Figura 4.5 – Trocas de mensagens - Diameter

Como pode ser visto, antes de focalizar um serviço de AAA específico, o nó inicial (A) solicita informações sobre as possibilidades do seu parceiro (B). Tendo recebido essa informação, os nós A e B começam a troca de mensagens que realmente viabilizarão o serviço desejado. Após terem sido atingidos os objetivos (ou em situações de encerramento precoce planejado), as partes finalizam a conexão (dependendo do contexto, tanto A quanto B podem solicitar esta finalização).

4.4. Protocolos de Autenticação utilizados nas redes WLAN

Nesta seção será apresentado o protocolo EAP, que é a base que viabiliza os processos de autenticação que envolvem as WLAN. Também serão abordados alguns de seus métodos criados posteriormente com o objetivo de resolver problemas específicos de determinados ambientes e tecnologias. Será tratado mais detalhadamente, além do próprio *framework* EAP, o ERP, pelo fato de ser um protocolo ligado diretamente ao processo de re-autenticação que é um dos focos de um processo de *handover*.

4.4.1. EAP

O EAP (*Extensible Authentication Protocol* – Protocolo de Autenticação Extensível) é um framework que suporta vários métodos de autenticação e que tipicamente é executado diretamente na camada de enlace, não requerendo, por exemplo, conectividade IP. Ele foi projetado para autenticações de acesso a redes e, apesar de fornecer funções que servem de base para diversos métodos de autenticação, não pode ser considerado um mecanismo de autenticação específico.

O EAP é bastante flexível, podendo ser usado em enlaces dedicados ou em circuitos comutados, da mesma forma funcionando em redes com ou sem fio. O processo de encapsulamento para meios com fio (os vinculados especificamente ao padrão IEEE-802) é descrito no padrão [IEEE-802.1X]. Já no caso das redes IEEE sem fio, a descrição do processo encontra-se no padrão [IEEE-802.11-2007], que embute a emenda IEEE-802.11i.

Diferente de muitos protocolos de autenticação existentes, as autenticações baseadas em EAP sempre são iniciadas pelo autenticador (quem autentica), e não pelo par ou suplicante (quem deseja autenticar-se). Além disso, o framework viabiliza a utilização de servidores de autenticação remotos. Em situações nas quais um autenticador não suporta algum dos métodos de autenticação escolhidos para o processo, este poderá delegar ao servidor de autenticação remoto a execução do método. Nesse caso, o autenticador servirá apenas de passagem para o fluxo de informações entre o servidor e o suplicante [RFC3748]. Cabe salientar que essa última situação é a que ocorre mais freqüentemente quando se utiliza o EAP.

Apesar de fornecer suporte à retransmissão de pacotes, o EAP não suporta fragmentação (alguns métodos vinculados, como o EAP-TLS, por exemplo, fornecem esse suporte) nem ordenação de pacotes (utiliza esse serviço de camadas inferiores).

Os pacotes EAP têm, de uma forma geral, o formato observado na Figura 4.6, contendo os seguintes componentes:

- Código – ocupa 1 octeto e identifica o tipo de pacote EAP, que pode ser *Request* (requisição), *Response* (resposta), *Success* (sucesso) e *Failure* (falha);
- Identificador – ocupa 1 octeto e tem a finalidade de sincronizar respostas com suas respectivas requisições;
- Comprimento – ocupa 2 octetos e indica o tamanho (comprimento), em número de octetos, do pacote EAP como um todo (incluindo o "cabeçalho" formado pelo código, identificador e comprimento);

- Dados – tem seu formato definido com base no código do pacote e pode ocupar 0 ou mais octetos.

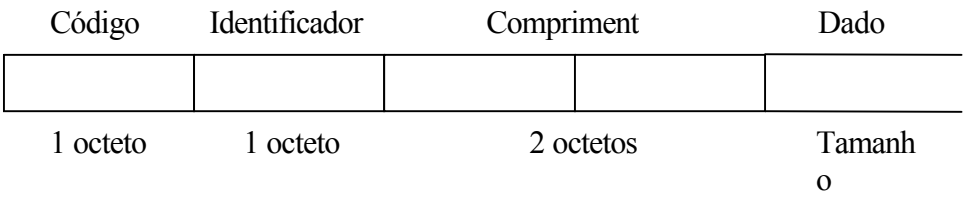

Figura 4.6 – Pacote EAP – formato genérico

Conforme aquilo que é visto na Figura 4.7, o processo de troca de mensagens de autenticação do EAP ocorre seguindo os seguintes passos:

1. O autenticador envia uma solicitação de autenticação de determinado tipo (identidade, desafio-MD5, etc.) para o suplicante;
2. O suplicante responde com um pacote adequado ao tipo de autenticação solicitado;
3. O autenticador envia um pacote de requisição adicional e o suplicante responde com pacotes adequados; esse passo deve ser repetido quantas vezes forem necessárias; salienta-se que uma nova requisição somente poderá ser feita após a resposta adequada ter sido recebida, entretanto, caso não sejam recebidas respostas, poderão ser feitas retransmissões de requisições já enviadas; caso proceda-se dessa forma e, mesmo assim, as respostas continuem sem chegar, o autenticador deverá finalizar a troca de mensagens;
4. Após o número adequado de mensagens trocadas, o autenticador concluirá sobre o sucesso ou a falha da autenticação e, em ambos os casos, enviará para o suplicante o pacote final sinalizando o resultado (sucesso ou falha).

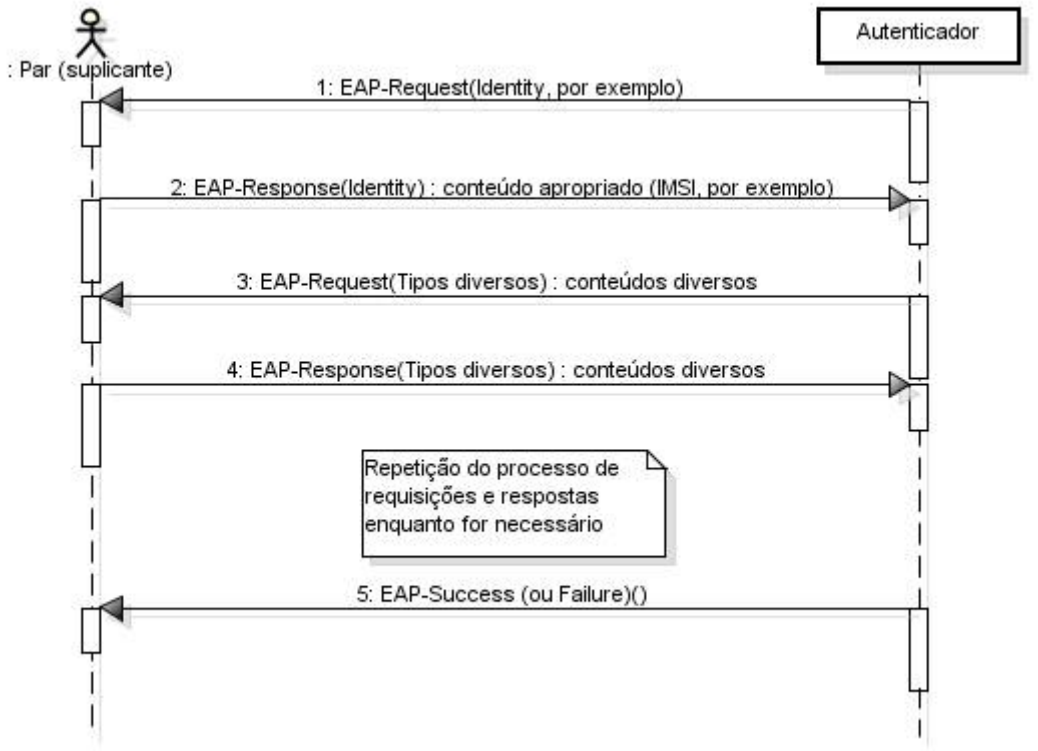

Figura 4.7 – EAP – troca de mensagens genéricas

A hierarquia de chaves do EAP (maior detalhamento na seção4.6) está baseada na criação inicial de duas chaves: a MSK (*Master Session Key* - Chave Mestra da Sessão) e a EMSK (*Extended Master Session Key* - Chave Mestra Estendida da Sessão). Após a troca de mensagens entre o servidor e o suplicante (tendo com intermediário o autenticador) e o sucesso do processo de autenticação, o servidor remete a MSK para o autenticador (utilizando protocolos de AAA como os enunciados na seção 4.3) que, por sua vez, estabelece, em conjunto com o suplicante e utilizando a MSK, as TSKs (*Transient Session Keys* - Chaves Transientes de Sessão), que são utilizadas na proteção dos pacotes de dados [RFC 5296].

Para viabilizar suas funcionalidades, o EAP obedece a um modelo conceitual composto pelas camadas definidas a seguir [RFC 3748]:

- Camada do método EAP – responsável pela execução dos algoritmos de autenticação e pela fragmentação dos dados, visto que esta última tarefa não é suportada pelo framework EAP;
- Camada do suplicante ou do autenticador (conforme o local onde esteja sendo executado o determinado passo do protocolo) – responsável por diferenciar a execução das funções do método EAP (camada acima) entre o suplicante e o autenticador;
- Camada EAP – recebe e transmite os pacotes EAP por meio da camada inferior, implementa a detecção de pacotes duplicados e o processo de retransmissão de

pacotes, além de demultiplexar os pacotes EAP para a camada superior (do suplicante ou do autenticador, conforme o caso);

- Camada inferior – tem a função de receber e transmitir os *frames* EAP entre o suplicante e o autenticador; são exemplos de camadas inferiores o PPP [RFC 1661], o UDP [RFC 768], o TCP [RFC 793], o IKEv2 [RFC 4306], o [IEEE 802.1X] e o [IEEE 802.11], dentre outros; o EAP assume que a camada inferior apresenta características como transporte não confiável (solicitação de retransmissões é papel do EAP), detecção de erros, garantia de ordenação de pacotes, possibilidade de duplicação de pacotes, dentre outras.

Na seção a seguir serão descritos alguns métodos de autenticação baseados no EAP e que podem ser utilizados em redes do padrão [IEEE 802.11]. Esses métodos (e quaisquer outros que venham a ser utilizados com este padrão) devem atender obrigatoriamente a alguns requisitos para que sejam considerados seguros [RFC4017]:

1. Capacidade de geração de material para chaves simétricas;
2. Utilização de chaves fortes (mínimo de 128 bits);
3. Suporte a autenticação mútua;
4. Equivalência de estado compartilhado (estados equivalentes no autenticador e no suplicante);
5. Resistência a ataques de dicionário;
6. Proteção contra ataques MITM (*man-in-the-middle* – homem no meio) [Schneier, 2006];
7. Negociação protegida de suítes criptográficas.

Além desses, há alguns outros requisitos recomendados e características opcionais que devem ser levados em consideração para os métodos EAP a serem utilizados em WLAN: fragmentação, ocultação da identidade dos usuários finais, canal de ligação, re-conexão rápida, etc.

4.4.2. ERP

O ERP (*EAP Re-authentication Protocol* – Protocolo de Re-autenticação EAP) [RFC 5296] será abordado com um nível de detalhamento maior que os demais métodos tendo em vista a sua importância no contexto da diminuição do atraso gerado nos processos de autenticação.

Ele foi criado para minimizar os problemas de re-autenticação para a passagem de suplicantes (para o caso explorado neste curso, equipamentos móveis) entre autenticadores ou para a extensão da autenticação entre um suplicante e um autenticador específico.

Nos métodos que implementam o EAP, quando um suplicante deseja autenticar-se para um novo autenticador, geralmente é executado um novo processo de autenticação EAP completo (há métodos, como o EAP-SIM, o EAP-AKA e o EAP-TTLS que são exceções a essa regra), independente do fato do suplicante já estar autenticado em outro ambiente e haver material para a criação de chaves cuja validade ainda não expirou. Esse novo processo de autenticação envolve uma nova troca de mensagens entre as partes envolvidas (como visto na seção 4.4.1), o que gera um atraso bastante relevante ao processo de passagem (*handover*) do suplicante de um autenticador para outro, quer no contexto da necessidade de aumento do tempo de processamento, quer no tocante ao tempo necessário para o trânsito das novas mensagens [RFC 5169].

Percebendo esse problema, alguns esforços anteriores ao ERP foram feitos no sentido de reduzir esse atraso, entretanto os mecanismos paliativos criados incorriam em problemas relacionados, por exemplo, à dependência de métodos EAP específicos ou à impossibilidade de re-autenticação entre autenticadores diferentes. Dessa forma, permanecia uma lacuna a ser solucionada por um protocolo genérico que pudesse ser executado independentemente do método EAP utilizado.

Nesse contexto surge o ERP, cuja idéia, apesar de poder ser utilizada no âmbito de *handovers* inter-domínios e inter-tecnologias, focaliza as re-autenticações dentro de um mesmo domínio administrativo (*handover* intra-domínio).

Os principais objetivos a serem atingidos pela implementação do ERP são [RFC 5169]:

- Execução de operações com baixa latência (ou pelo menos com latência inferior à observada em um processo de autenticação EAP completo);
- Independência relativa das camadas inferiores, o que viabiliza a execução do protocolo sobre qualquer tecnologia de acesso;
- Independência e compatibilidade com os métodos EAP utilizados para o processo de autenticação, permitindo o uso do ERP vinculado a quaisquer desses métodos;
- Compatibilidade com protocolos de AAA, particularmente com o RADIUS e o Diameter (seção 4.3);
- Agilidade criptográfica, entendida nesse contexto como a capacidade de negociação dinâmica de parâmetros criptográficos sem afetar o desempenho global do processo;
- Minimização das mudanças necessárias nas diversas entidades partícepes dos processos de re-autenticação (suplicante, autenticador e servidor EAP), desde que o desempenho geral do processo não seja afetado.

Além dos descritos acima, especificamente no contexto da segurança, também devem ser seguidos os preceitos definidos para o gerenciamento de chaves utilizadas nos processos de AAA (seção 4.6).

É válido ressaltar que, apesar da independência relativa das camadas inferiores ser um dos objetivos citados, algumas especificações de camadas inferiores, como o [IEEE 802.1X] e o IKEv2 [RFC 4306], por exemplo, devem ser revisadas para permitir o funcionamento coerente do ERP.

O funcionamento do ERP permite que um suplicante autentique-se para um servidor de re-autenticação EAP, e vice-versa, utilizando material criptográfico gerado a partir de execuções prévias de autenticações EAP, não sendo possível qualquer processo de re-autenticação sem uma autenticação completa anterior. No melhor caso, será necessária apenas uma viagem de ida e volta (*round trip*) entre o suplicante e o servidor de re-autenticação EAP. Tal qual nessas execuções prévias, no ERP, geralmente, o autenticador serve apenas de intermediário entre o suplicante e o servidor.

A Figura 4.8 detalha uma típica troca de mensagens ERP.

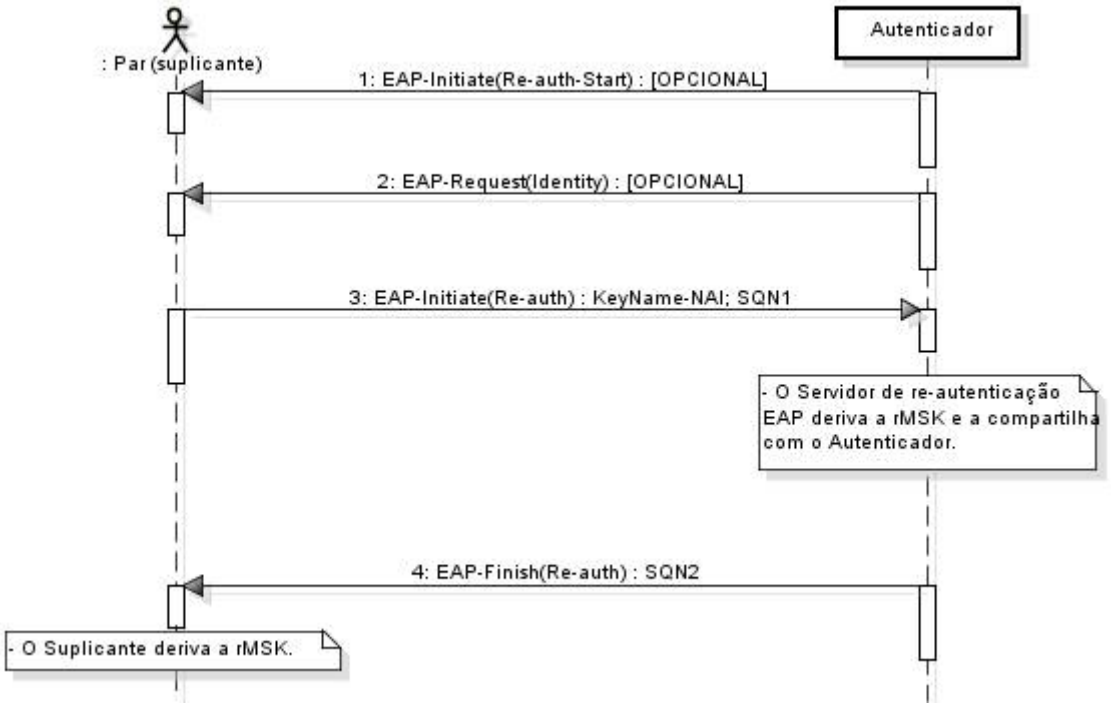

Figura 4.8 – ERP – troca de mensagens genéricas

As quatro situações a seguir são possíveis, e para cada uma delas haverá uma troca de mensagens diferente:

1. O processo é iniciado pelo suplicante (caso normal para o ERP) e o autenticador suporta o ERP e possui material não expirado para criação de chaves:
    - O suplicante, em situações normais, é o responsável por iniciar a troca de mensagens ERP utilizando um novo código EAP, o "EAP-Initiate" (este código e o "EAP-Finish" foram criados especificamente para re-autenticações EAP) – passo 3 da figura. Nesta mensagem, seguem os parâmetros "KeyName-NAI", que identifica o domínio do servidor de re-autenticação EAP e a "rIK" (chave de integridade da re-autenticação) usada para proteger as mensagens de re-autenticação, e um número

seqüencial usado na proteção contra ataques de repetição (*Replay Attacks*).

- O autenticador, tendo recebido a solicitação, encaminha a mensagem ao servidor de re-autenticação apropriado que foi deduzido a partir do "KeyName-NAI". Esse servidor deriva, então, a "rMSK" (Chave Mestra da Sessão de Re-autenticação) a partir da "rRK" (Chave Raiz de Re-autenticação, que foi deduzida a partir da EMSK criada por ocasião do processo de autenticação EAP completo prévio) e do número seqüencial recebido. Após isso, o servidor envia uma mensagem "EAP-Finish" para o autenticador o qual a repassa para o suplicante – passo 4 da figura. Nessa mensagem segue, como parâmetro, o número seqüencial recebido e incrementado pelo servidor. Seguirá também, apenas do servidor até o autenticador, a "rMSK" derivada no servidor.
- Tendo recebido a mensagem do autenticador, o suplicante deriva a "rMSK" a partir da "rRK" (que já possui) e do número seqüencial recebido.
- Salienta-se que as comunicações ocorridas entre o autenticador e o servidor são viabilizadas por meio de protocolos de AAA (seção 4.3) e são omitidas da Figura 4.8 por questões de simplicidade.

2. O processo é iniciado pelo suplicante (caso normal para o ERP) e o autenticador não suporta o ERP ou não possui material não expirado para criação de chaves:
   - O suplicante envia um "EAP-Initiate" – passo 3 da figura – para o autenticador;
   - O autenticador despreza a mensagem do suplicante por não reconhecê-la;
   - O suplicante permanece insistindo no passo 3 da figura até receber um "EAP-Request" do autenticador, o que faz com que o suplicante conclua que o autenticador não suporta o ERP;
   - Desse ponto em diante, tudo ocorrerá como visto na Figura 4.7 – EAP – troca de mensagens genéricas.
3. O processo é iniciado pelo autenticador e o suplicante suporta o ERP e possui material não expirado para criação de chaves:
   - O autenticador envia o "EAP-Initiate" solicitando ao suplicante o início da re-autenticação – passo 1 da figura;
   - O suplicante responde com outro "EAP-Initiate" – passo 3 da figura – e o processo prossegue seguindo os passos da situação 1;
   - Salienta-se que, caso o suplicante demore a responder à solicitação do autenticador, este poderá inferir, de forma equivocada, que o suplicante não suporta o ERP. Caso isso ocorra, o autenticador enviará um "EAP-Request" para o suplicante (passo 2 da figura) que, por sua vez, deverá ignorar essa requisição do autenticador e prosseguir com o passo 3 da figura. O autenticador, tão logo receba o "EAP-Initiate" atrasado do

suplicante, deverá desprezar o "EAP-Request" enviado por ele mesmo e dar prosseguimento ao processo de re-autenticação.

4. O processo é iniciado pelo autenticador e o suplicante não suporta o ERP ou não possui material não expirado para criação de chaves:
   - O autenticador envia o "EAP-Initiate" solicitando ao suplicante o início da re-autenticação – passo 1 da figura;
   - O suplicante ignora e descarta as mensagens recebidas por não reconhecê-las;
   - Após um número configurável de tentativas sem sucesso, o autenticador envia um "EAP-Request" para o suplicante solicitando o início de um processo de autenticação EAP completo – passo 2 da figura;
   - Desse ponto em diante, tudo ocorrerá como visto na Figura 4.7 – EAP – troca de mensagens genéricas.

4.4.3. Outros métodos EAP

Existem vários outros métodos EAP definidos para a solução de problemas específicos e para atender a demandas particulares de certos fabricantes. Pode-se citar, dentre vários outros, os seguintes:

1. EAP-TLS [RFC 5216]

Esse método considera a utilização do TLS (*Transport Layer Security* – Segurança da Camada de Transporte) [RFC 5246] vinculado ao funcionamento do *framework* EAP representa uma solução bastante segura, apesar de ser de difícil implementação. O TLS pode ser usado, por exemplo, para promover a autenticação mútua dos elementos que participam de uma troca de mensagens EAP. O protocolo oferece um túnel fim-a-fim criptografado para a transferência de dados, inclusive material utilizado na derivação de chaves, baseado na utilização de certificados digitais e na existência de uma ICP (infra-estrutura de chaves públicas).

2. EAP-TTLS [RFC 5281]

É um método EAP que encapsula uma sessão TLS que viabiliza, na fase de aperto de mãos (*handshake*) a autenticação do servidor para o cliente (ou a autenticação mútua) e a geração de material para a criação de chaves criptográficas. Essas chaves serão utilizadas posteriormente, na fase de troca de dados, para o fechamento de um túnel criptográfico que proverá segurança às informações em trânsito entre as partes. Uma vez fechado o túnel seguro, uma nova autenticação será feita (cliente para o servidor ou autenticação mútua) de forma sigilosa utilizando o próprio mecanismo de autenticação do EAP (é possível utilizar-se outros métodos de autenticação).

Uma característica especialmente interessante deste método é o fato dele prever a possibilidade de recomeço de uma sessão já encerrada, o que diminui significativamente o tempo necessário para o processo de autenticação e geralmente é empregado quando um cliente necessita realizar uma re-autenticação para um ponto de acesso de uma rede sem fio.

3. LEAP

O LEAP (*Lightweight* EAP – EAP Leve) consiste em um método EAP proprietário desenvolvido pela Cisco. Por ser um padrão fechado, a sua escolha como método EAP deve ser feita com cautela, pois o seu uso está relacionado ao tipo de equipamento utilizado, ou seja, a utilização do LEAP implica na necessidade de que os equipamentos da infra-estrutura (placas de rede e *Access Points*, por exemplo) sejam compatíveis com o mesmo. O LEAP foi desenvolvido para funcionar no topo da camada da infra-estrutura de autenticação do modelo [IEEE 802.1X] e viabiliza autenticação mútua e autenticação baseado no equipamento.

4. PEAP

Desenvolvido pela Microsoft, O PEAP (*Protected* EAP – EAP Protegido) consiste em um método de autenticação com dois estágios. O primeiro estágio estabelece uma sessão TLS para o servidor e permite que o cliente autentique o servidor usando o certificado digital do servidor. O segundo estágio requer um segundo método EAP encapsulado na sessão PEAP para autenticar o cliente a um servidor RADIUS. Isso permite que o PEAP use uma variedade de métodos de autenticação de clientes, incluindo o uso de senhas no protocolo MS-CHAP v2 [RFC 2759] e certificados usando o EAP-TLS encapsulado no PEAP.

4.5. Principais Protocolos de Autenticação utilizados nas redes móveis celulares

Semelhante àquilo que foi mostrado na seção anterior, aqui serão apresentados os principais protocolos de autenticação utilizados em redes celulares de 2ª e 3ª gerações, o EAP-SIM e o EAP-AKA, respectivamente.

4.5.1. Redes 2G - EAP-SIM

O EAP-SIM (EAP *Method for Global System for Mobile Communications Subscriber Identity Module* – Método EAP para Módulo de Identidade do Assinante do Sistema Global para Comunicações Móveis) [RFC4186] foi desenvolvido pelo 3GPP (*3rd Generation Mobile System* – Sistema Móvel de Terceira Geração) e é usado para autenticação e distribuição de chaves de sessão utilizando o SIM do GSM (*Global System for Mobile Communications* - Sistema Global para Comunicações Móveis).

A autenticação na rede GSM é baseada em mecanismos de Desafio-Resposta (*Challenge-Response*) e utiliza os algoritmos criptográficos A3 e A8, cujas seguranças interferem diretamente na confiabilidade do próprio EAP-SIM.

A Figura 4.9 retrata um processo de autenticação completa executado no contexto do EAP-SIM. Cabe salientar que na figura estão omitidos, por questão de simplicidade, o servidor EAP e outras entidades da arquitetura GSM que serão utilizadas na autenticação (a comunicação entre o Autenticador e as demais entidades é feita por meio de protocolos AAA). Dessa forma o Autenticador, que na maioria das vezes serve apenas como passagem para as mensagens de autenticação, está aglutinando as funções do Servidor EAP e das demais entidades.

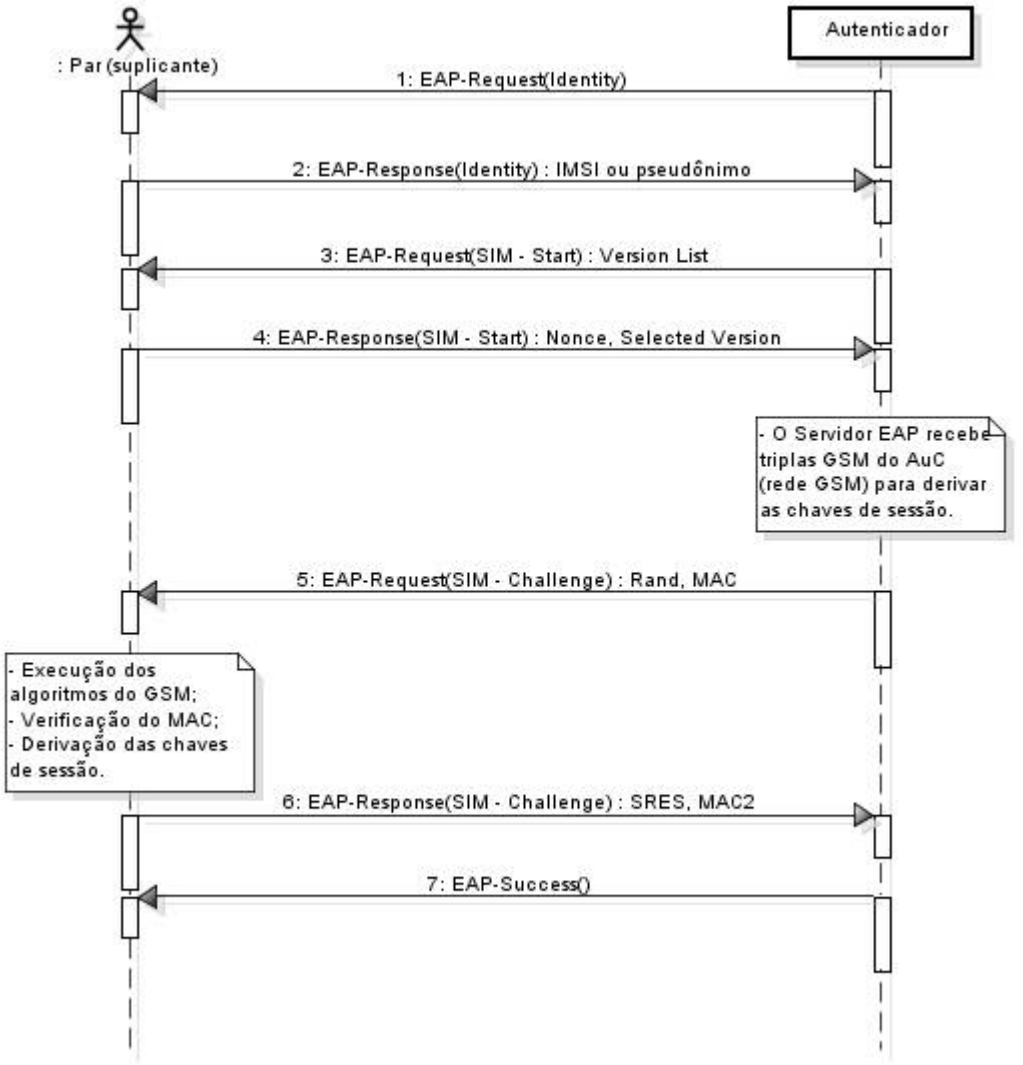

Figura 4.9 – processo de autenticação completa do EAP-SIM

Em uma primeira mensagem, o Autenticador solicita a identidade do Suplicante. Este, por sua vez, responde com parâmetros vinculados ao IMSI (*International Mobile Subscriber Identity* – Identidade Internacional do Assinante Móvel) ou, caso deseje-se preservar a identidade do assinante, com um pseudônimo temporário. Após isso, o Suplicante recebe do Autenticador a solicitação de que as operações em seu SIM sejam iniciadas. Vinculada a essa solicitação, segue a lista de versões do EAP-SIM suportadas pelo Servidor EAP (representado, na figura, pelo Autenticador). O Suplicante responde com a versão EAP-SIM selecionada e com um valor aleatório (Nonce) que será utilizado em operações futuras. Tendo recebido esse conteúdo, o Autenticador entra em contato com o AuC (*Authentication Center* – Centro de Autenticação) da rede GSM e obtém algumas triplas GSM (vetores de autenticação) que contêm material a ser utilizado na autenticação do Suplicante (RAND – número aleatório que servirá de desafio; SRES – valor esperado de resposta do suplicante; Kc – chave de cifração). Cabe salientar que, apesar desse material pertencer à rede GSM, o Autenticador poderá armazená-lo com a

finalidade de utilizar posteriormente em caso de erros. O Autenticador envia, então, para o Suplicante, um Desafio contendo um valor aleatório (RAND) protegido por um MAC (*Message Authentication Code* – Código de Autenticação de Mensagem). Ao receber o Desafio, o Suplicante executa alguns algoritmos do padrão GSM (tal qual o fez o Autenticador) para validar o MAC e para derivar chaves criptográficas que serão utilizadas no restante do processo. Ele responde para o Autenticador, então, com um segundo MAC que protege os valores resposta criados no ambiente do SIM. Ao receber a resposta, o Autenticador valida esse segundo MAC e, caso todo o processo tenha ocorrido dentro do esperado, finaliza o tráfego de mensagens sinalizado para o Suplicante o sucesso do processo de autenticação.

A hierarquia de chaves do EAP-SIM está baseada na geração da "MK" (*Master Key* – Chave mestra) a partir dos parâmetros secretos encontrados na rede GSM (chaves secretas localizadas nos Centros de Autenticação e nos Módulos de Identificação de Assinantes) e dos valores trocados por ocasião do processo de autenticação. A partir daí, a "MK" viabiliza a criação de "TEKs" (*Transient EAP Keys* – Chaves EAP Transientes), que protegerão os pacotes EAP-SIM, da "MSK" e da "EMSK" (comentadas na seção 4.4.1 deste trabalho).

Com o objetivo de diminuir o tempo empregado nos processo de autenticação que utilizam o EAP-SIM, este método viabilizou um mecanismo conhecido como re-autenticação rápida (*fast re-authentication*) que é particularmente interessante em situações nas quais autenticações EAP-SIM são uma necessidade freqüente. Salienta-se que, tal qual na situação do ERP apresentado na seção 4.4.2 deste trabalho, para que haja uma re-autenticação, obrigatoriamente deverá ter ocorrido uma autenticação completa anteriormente. O ganho temporal observado na comparação da autenticação completa com a re-autenticação rápida advém do fato desta não utilizar os algoritmos A3/A8 (do padrão GSM) e não precisar de novas triplas GSM obtidas a partir do Centro de Autenticação (*AuC*). Dessa forma, a re-autenticação rápida pode ser realizada por meio de uma quantidade menor de viagens de ida e volta (*Roundtrips*) do que a autenticação completa.

Apesar de ser algo bastante interessante, esse mecanismo é opcional tanto nas implementações de suplicantes quanto nas de servidores. Assim, caso uma das partes deseje realizar uma re-autenticação rápida e a outra não forneça suporte para tal (ou apenas opte por não fazê-lo), o processo será obrigatoriamente revertido para uma autenticação completa.

As chaves utilizadas no processo de re-autenticação rápida são derivadas daquelas utilizadas na autenticação completa realizada previamente (o que caracteriza a REUTILIZAÇÃO de MATERIAL CRIPTOGRÁFICO). As chaves de autenticação e de cifração utilizadas ("TEKs" derivadas da "MK") são as mesmas do processo de autenticação prévio, entretanto novas "MSK" e "EMSK" serão criadas a partir da "MK" e de novos parâmetros trocados durante a re-autenticação.

4.5.2. Redes 3G - EAP-AKA

O EAP-AKA (EAP *Method for 3rd Generation Authentication and Key Agreement* – Método EAP para Autenticação e Acordo de Chave de 3ª Geração) [RFC4187] é

semelhante ao EAP-SIM, também tendo sido desenvolvido pelo 3GPP e sendo usado para autenticação e distribuição de chaves de sessão. Utiliza, para tal, o mecanismo de Autenticação e Acordo de Chave (AKA) encontrado em redes móveis de 3ª geração UMTS (*Universal Mobile Telecommunications System* – Sistema de Telecomunicações Móveis Universal) e CDMA2000 (*Code Division Multiple Access* 2000 - Acesso Múltiplo por Divisão de Código 2000).

A Figura 4.10 retrata um processo de autenticação completa executado no contexto do EAP-AKA. Também estão omitidos, tal qual na figura relativa ao EAP-SIM, o servidor EAP e outras entidades da arquitetura 3G (UMTS/CDMA2000) que serão utilizadas na autenticação (a comunicação entre o Autenticador e as demais entidades é feita por meio de protocolos AAA). Dessa forma o Autenticador, que na maioria das vezes serve apenas como passagem para as mensagens de autenticação, está aglutinando as funções do Servidor EAP e das demais entidades.

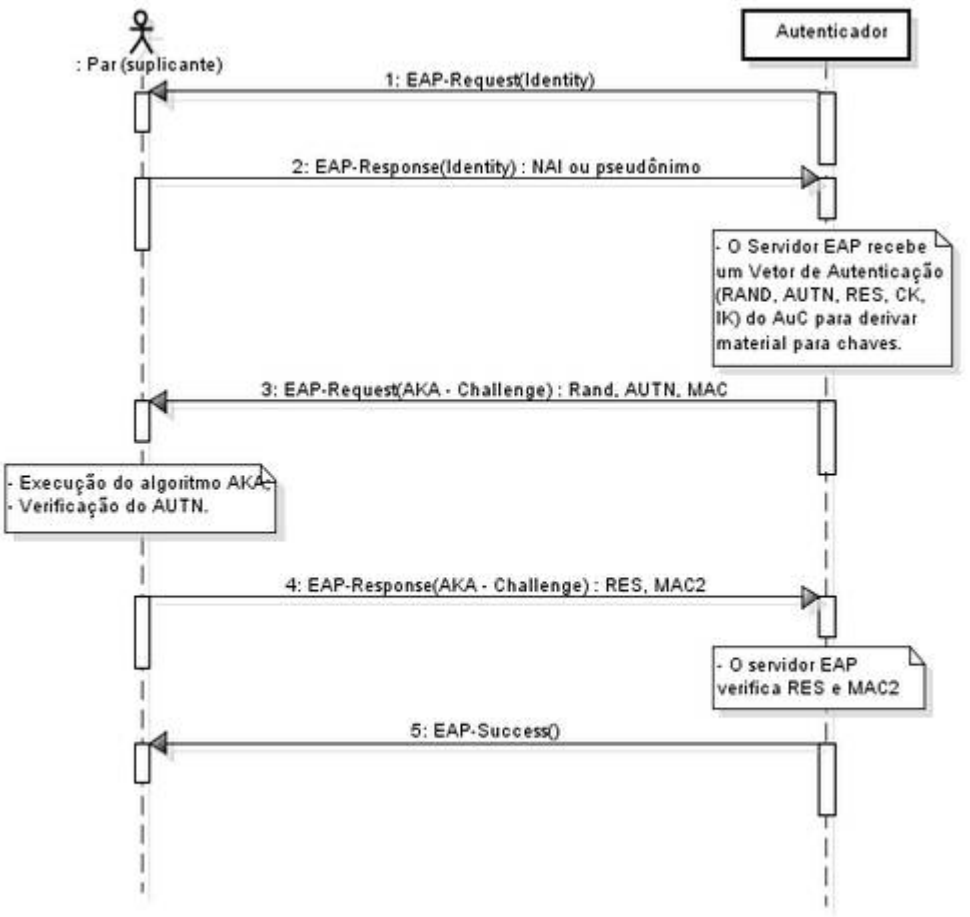

Figura 4.10 – processo de autenticação completa do EAP-AKA

O EAP-AKA utiliza, no mínimo, duas viagens de ida e volta (*round trip*) para a autenticação das partes. Inicialmente, o Autenticador solicita a identidade do Suplicante. Este, por sua vez, responde com o NAI (*Network Access Identifier* - Identificador de

Acesso à Rede), parâmetro vinculado ao IMSI, ou, caso deseje-se preservar a identidade do assinante, com um pseudônimo temporário (ao invés de transmitir a identidade do suplicante neste passo, é possível deixar para comunicá-la em comunicações EAP-AKA futuras). Após isso, o Servidor EAP obtém do Centro de Autenticação (AuC) da rede 3G um vetor de autenticação composto por um número aleatório (RAND), um *token* de autenticação da rede (AUTN), um valor esperado que autenticará o suplicante (RES ou XRES), e duas chaves de sessão, uma de cifração ("CK") e outra para garantia de integridade ("IK"). Salienta-se que esse vetor de autenticação poderá ser armazenado pelo servidor EAP para utilizações posteriores, tornando desnecessárias novas comunicações entre o servidor EAP e o núcleo da rede 3G. A seguir, o servidor EAP inicia o protocolo AKA enviando um Desafio contendo os parâmetros RAND, AUTN e um código de autenticação de mensagem MAC. O suplicante, a partir da execução do algoritmo AKA (utilizando parâmetros secretos contidos em seu *smartcard*), verifica o *token* de autenticação da rede (AUTN) e, em caso de sucesso, responde ao servidor com o parâmetro RES, que permitirá sua autenticação perante esse servidor, e um MAC que viabiliza a garantia de integridade dessa mensagem. Tendo recebido a resposta, o servidor atesta a autenticidade do suplicante por meio da verificação dos parâmetros RES e MAC2. Finalizado o procedimento, o servidor envia uma mensagem para o suplicante sinalizando o sucesso de todo o processo. Cabe salientar que o autenticador poderá receber, do servidor EAP, material para chaves que na será repassado para o suplicante tendo em vista o fato deste já possuir esse material em seu meio.

A hierarquia de chaves do EAP-AKA é parecida à do EAP-SIM e está baseada na geração da "MK" (*Master Key* – Chave mestra) a partir dos parâmetros AKA encontrados na rede 3G ("IK" e "CK" discutidos anteriormente) e da identidade trocada por ocasião do processo de autenticação. A partir daí, a "MK" viabiliza a criação de "TEKs" (*Transient EAP Keys* – Chaves EAP Transientes), que protegerão os pacotes EAP-AKA, da "MSK" e da "EMSK" (comentadas na seção 4.4.1 deste trabalho).

Semelhante àquilo que ocorre no EAP-SIM e com o objetivo de diminuir o tempo empregado nos processo de autenticação que utilizam o EAP-AKA, este método viabilizou um mecanismo conhecido como re-autenticação rápida (*fast re-authentication*) que é particularmente interessante em situações nas quais autenticações EAP-AKA são uma necessidade freqüente. Salienta-se que, tal qual na situação do ERP apresentado na seção 4.4.2 deste trabalho, para que haja uma re-autenticação, obrigatoriamente deverá ter ocorrido uma autenticação completa anteriormente. O ganho temporal observado na comparação da autenticação completa com a re-autenticação rápida advém do fato desta não utilizar os algoritmos AKA (do padrão 3G) e não precisar de novos vetores obtidos a partir do Centro de Autenticação (*AuC*). Dessa forma, a re-autenticação rápida pode ser realizada por meio de uma quantidade menor de viagens de ida e volta (*Roundtrips*) do que a autenticação completa.

Apesar de ser algo bastante interessante, esse mecanismo é opcional tanto nas implementações de suplicantes quanto nas de servidores. Assim, caso uma das partes deseje realizar uma re-autenticação rápida e a outra não forneça suporte para tal (ou apenas opte por não fazê-lo), o processo será obrigatoriamente revertido para uma autenticação completa.

As chaves utilizadas no processo de re-autenticação rápida são derivadas daquelas utilizadas na autenticação completa realizada previamente (o que caracteriza, tal qual no EAP-SIM, a REUTILIZAÇÃO de MATERIAL CRIPTOGRÁFICO). As chaves de autenticação e de cifração utilizadas ("TEKs" derivadas da "MK") são as mesmas do processo de autenticação prévio, entretanto novas "MSK" e "EMSK" serão criadas a partir da "MK" e de novos parâmetros trocados durante a re-autenticação.

4.6.Mecanismos de gerenciamento de chaves

Com o objetivo de apresentar alguns desafios e soluções para o problema do gerenciamento de chaves criptográficas utilizadas nas redes envolvidas, serão abordados nesta seção alguns aspectos gerais do processo, hierarquia de chaves, distribuição e reuso de material usado para a criação de chaves.

4.6.1. Aspectos gerais

A atividade de gerenciamento de chaves é algo complexo e geralmente envolve vários elementos (entidades e procedimentos). Analisando, por exemplo, o gerenciamento de chaves a serem utilizadas em um processo de AAA (seção 4.3), percebe-se que estão envolvidas entidades como servidores de AAA e vários elementos do núcleo das redes consideradas, além de serem utilizados procedimentos para viabilizar os diversos controles necessários e protocolos de comunicação seguros para permitir a troca adequada de informações. Dessa forma, constata-se facilmente que um mecanismo que se propõe a gerenciar chaves deve ser, no mínimo, tão complexo quanto os problemas que ele tenta resolver.

O êxito em operações com sistemas criptográficos depende intimamente de um bom gerenciamento de chaves que engloba atividades como criação de chaves, armazenamento, transmissão segura para as partes autorizadas, utilização (e re-utilização) segura de material criptográfico, troca (reposição) de chaves e sua destruição, dentre outras.

Muitos têm sido os trabalhos realizados sobre essas diversas atividades (separadamente ou em conjunto). Nesse contexto a presente seção não pretende exaurir o assunto, mas tão somente dar noções sobre alguns problemas existentes e algumas de suas possíveis soluções, particularmente no que diz respeito às inter-conexões tratadas neste curso.

Especificamente relacionado às chaves tratadas nas seções passadas (EAP), a [RFC 5247] especifica uma hierarquia de chaves EAP e sugere um *framework* para o transporte e utilização do material a ser utilizado na geração das chaves e dos parâmetros gerados pelos métodos EAP.

Para um entendimento mais detalhando do processo de derivação de chaves EAP, faz-se necessária uma visão mais geral das fases de uma conversação que embute a utilização de uma autenticação EAP. Pode-se, então, considerar a ocorrência de três fases distintas [RFC 5247]:

1. Descobrimento – suplicante (par) encontra o autenticador e descobre suas possibilidades e compatibilidades;

2. Autenticação – tem início após o suplicante e o autenticador terem se reconhecido mutuamente e viabiliza a garantia da identificação das partes envolvidas;
    a. Autenticação EAP – sub-fase obrigatória e que viabiliza a autenticação por meio do protocolo EAP. Nessa sub-fase o material de criação de chaves é derivado tanto no suplicante quanto no servidor EAP envolvido (que pode ou não estar embutido no autenticador) a partir de valores armazenados previamente nessas entidades;
    b. Transporte de chaves por protocolo de AAA (opcional) – essa sub-fase ocorre caso haja a necessidade de comunicação entre o autenticador e um servidor EAP que se encontra à retaguarda. Nesse caso é utilizado um protocolo de AAA (tipicamente o RADIUS ou o Diameter) para viabilizar essa comunicação e permitir que o material de criação de chaves adequado (tipicamente a MSK) seja repassado do servidor para o autenticador (que é quem de fato entrará em contato direto com o suplicante);
3. Protocolo de associação segura - tendo ocorrido um processo de autenticação com êxito, deverá ser estabelecida uma associação de segurança entre as partes extremas (suplicante e autenticador) para viabilizar a comunicação de informações (não mais de dados ligados à autenticação) entre suplicante e autenticador.

É interessante observar que a maioria das fases descritas anteriormente não são executadas pelo protocolo EAP em si. As fases 1, 2.b e 3 são desencadeadas além do próprio escopo do método EAP considerado (1 e 3 dependem das camadas inferiores e 2.b depende de um protocolo de AAA externo).

4.6.2. Hierarquia de chaves

Como percebido no item anterior, as derivações de chaves a serem utilizadas pelos métodos EAP sempre são feitas a partir de conteúdo existente nas entidades envolvidas no processo. Esse conteúdo pode estar no formato de segredos pré-compartilhados ou de chaves assimétricas (públicas e privadas) e dá origem a dois tipos de material para a criação de chaves EAP: material calculado localmente pelo método EAP e não exportado (TEKs, por exemplo) e material calculado e exportado pelo método EAP (MSK e EMSK, por exemplo).

Cada chave criada no âmbito do *framework* de gerenciamento de chaves EAP tem um nome e um escopo bem definidos. Podem ser citadas, por exemplo:

- MSK (*Master Session Key* – Chave Mestra de Sessão) - chave derivada, entre o suplicante e o servidor, gerada a partir de material previamente existente e que é exportada pelo método EAP;

- EMSK (*Extended* MSK – MSK Estendida) - chave adicional criada no mesmo âmbito da MSK e utilizada para outros propósitos; é exportada pelo método EAP, mas é de conhecimento apenas do suplicante e do servidor;
- TEK (*Transient EAP Key* – Chave EAP Transiente) - chaves de sessão que são utilizadas para o estabelecimento de um canal protegido entre suplicante e servidor durante o processo de autenticação EAP.
- TSK (*Transient Session Keys* - Chaves Transientes de Sessão) – chaves utilizadas na proteção dos pacotes de dados (canal protegido) que trafegam entre suplicante e autenticador após o processo de autenticação EAP.

Cabe salientar que a [RFC 5295] define que a EMSK poderá ser utilizada para a derivação de chaves de aplicações específicas, a saber:

- USRK (*Usage Specific Root Key* – Chave Raiz de Uso Específico);
- DSRK (*Domain Specific Root Key* – Chave Raiz de Domínio Específico);
- DSUSRK (*Domain Specific Usage Specific Root Key* – Chave Raiz de Domínio e Uso Específico).

Essas chaves, como indicado pelos próprios nomes, têm aplicações específicas e propõe-se a resolver problemas particularizados de certos métodos (ou melhorias de métodos) e também podem ter seus escopos delimitados para certos domínios de gerenciamento de chaves.

A Figura 4.11 ilustra a dependência existente entre as diversas chaves abordadas.

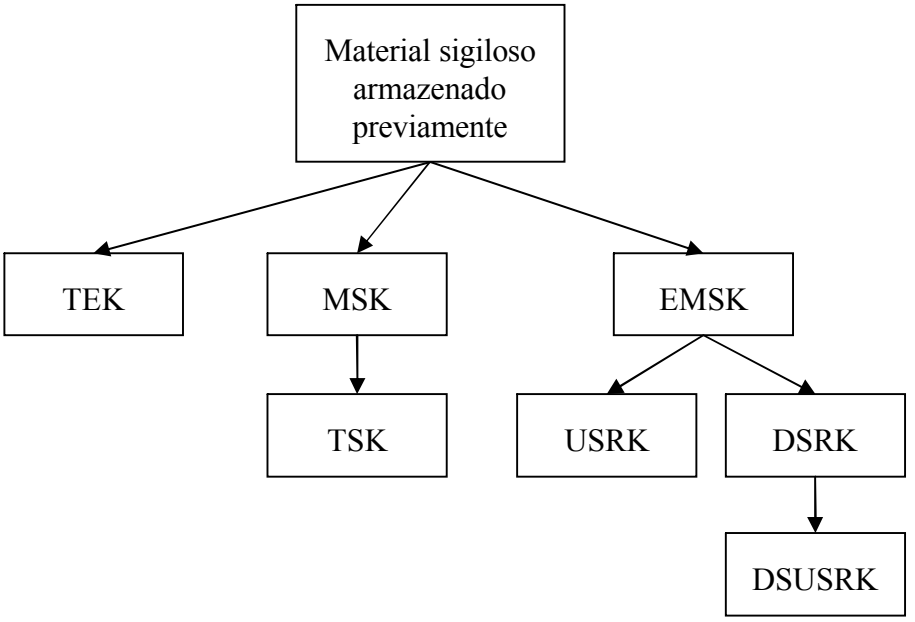

Figura 4.11 – Hierarquia de chaves EAP

4.6.3. Distribuição e reuso de material para chaves

O processo de geração de chaves criptográficas a serem utilizadas em protocolos de autenticação, como pôde ser visto, é algo complexo e demorado. Apesar dessas chaves serem criadas prioritariamente para viabilizar segurança entre um determinado suplicante e um autenticador, tem-se percebido que pelo menos parte do material criptográfico pode ser re-utilizado em novas parcerias (o mesmo suplicante e um novo autenticador), o que certamente permite uma diminuição relevante do computo do tempo global da nova autenticação. Entretanto, além do ganho gerado por esse novo processo, há também o ônus relacionado à necessidade de novos mecanismos que viabilizem a distribuição e o reuso do material de forma segura.

No contexto das chaves EAP que podem ser utilizadas em *handovers*, a [RFC 5749] propõe um mecanismo que viabiliza a distribuição controlada da chave raiz de um servidor EAP para outro servidor de rede que necessita dessa chave (ou de suas derivadas). O documento também propõe um modelo geral para um protocolo de intercâmbio de distribuição de chaves (KDE – *Key Distribution Exchange*) que se propõe a distribuir diferentes tipos de chaves utilizando protocolos de AAA.

O mecanismo de distribuição de chaves comentado anteriormente fundamenta-se na existência de dois servidores: o KDS (*Key Delivering Server* – Servidor de Distribuição de Chave), que é um servidor de rede que possui uma EMSK (ou uma DSRK) e a utiliza para criar e distribuir RKs (*Root Key* – Chave Raiz) para outros servidores, e o KRS (*Key Requesting Server* – Servidor de Requisição de Chave), que se comunica com o KDS solicitando as RKs. Uma possível forma de funcionamento da arquitetura de distribuição de chaves é, por exemplo:

1. Um suplicante solicita conexão a um servidor;
2. O servidor age como um KRS e solicita a um KDS (que de alguma forma já tenha compartilhado material de criação de chaves com o suplicante) uma RK que viabilize o fornecimento do serviço solicitado pelo suplicante;
3. A partir desse momento (e enquanto a RK não expirar) o KRS poderá atender a novas solicitações de mesmo suplicante utilizando a RK obtida e utilizadas na derivação das chaves adequadas.

Salienta-se que as EMSK, por definição, não podem sair do servidor EAP original (que contém os parâmetros adequados compartilhados com os suplicantes considerados), dessa forma os KDS obrigatoriamente serão servidores EAP originais ou então conterão apenas DSRKs derivadas de EMSKs localizadas em servidores EAP originais.

Além da arquitetura ilustrada, a [RFC 5749], como dito anteriormente, propõe o KDE. O papel desse protocolo é viabilizar a distribuição segura e controlada das RKs entre KRSs e KDSs. Para tal utiliza, além de seus próprios mecanismos internos, protocolos de AAA como os vistos na seção 4.3. A troca de mensagens gerenciadas pelo KDE é simples e baseia-se no envio, do KRS para o KDS, de uma "KDE-Request", e do KDS para o KRS, de uma "KDE-Response". Uma "KDE-Request" é um KRT (*Key Request Token – Token* de Requisição de Chave) encapsulado em uma mensagem de

AAA. Já uma "KDE-Response" é um KDT (*Key Delivery Token* – Token de Entrega de Chave) que contém a chave solicitada e também é encapsulado em uma mensagem de AAA.

O KDE é um protocolo que pode ser utilizado, por exemplo, para viabilizar os processos de re-autenticação propostos pelo ERP visto na seção 4.4.2.

4.7.Processos de *Handover* e o padrão IEEE 802.21

Nessa seção serão descritas as métricas, as técnicas e as fases do processo de *handover*, caracterizando cada uma delas no contexto das redes envolvidas e destacando os aspectos de segurança observados. Serão apresentados os tipos de *handover* caracterizados na literatura e as métricas de qualidade de serviço (QoS) que são utilizadas para sinalizar os impactos da execução de processos de segurança para a consecução do *handover*. Além disso, serão apresentados também os aspectos gerais do padrão IEEE 802.21 (*Media Independent Handover – Handover* Independente do Meio) – padrão aplicado aos *handover*s entre redes heterogêneas que são alvo deste trabalho - detalhando sua arquitetura e seu funcionamento a fim de que fiquem claras as suas possibilidades e a sua aplicabilidade dentro do contexto do problema explorado.

4.7.1. Tipos de *handover*

O processo de *handover* possui algumas características de acordo com suas classificações. A primeira e mais básica delas é feita em relação ao seu tipo, podendo ser [IEEE 802.21]:

1. *soft handover* - em que os recursos para suporte a fluxos de tráfego estão continuamente disponíveis durante as transferências de conexão do nó móvel, em nível de enlace, de um ponto de acoplamento origem para um ponto de acoplamento alvo; neste caso, são alocados recursos para o ponto de acoplamento alvo antes da ocorrência do evento de troca de enlace (correspondente a *make-before-break* – faça antes de quebrar);
2. *seamless handover* - *handover* no qual há mudança do ponto de acoplamento, em nível da camada de enlace, no entanto o nó móvel não sofre degradação dos parâmetros de serviço (qualidade do sinal, por exemplo) ou, caso sofra, essa degradação é tolerável tanto para o nó móvel quanto para a própria rede servidora;
3. *hard handover* - onde o *handover* é feito de forma abrupta, ocorrendo a perda da conexão anterior antes da atual ser estabelecida; neste sentido, recursos para suportar fluxos de tráfego são sujeitos a completa indisponibilidade entre a perda da conexão e a sua restauração (correspondente a *break-before-make* – quebre antes de fazer).

Outra classificação do *handover* é feita baseada nos tipos de tecnologia de acesso (RAT - *Radio Access Tecnologies*) das redes envolvidas. Durante a migração de um terminal móvel entre redes com a mesma RAT, o *handover* é dito horizontal (HHO – *Horizontal Handover*) ou intra-tecnologia (*Intratechnology Handover*). Por outro lado, em se tratando de redes heterogêneas, as RATs deverão ser diferentes, caracterizando o

handover dito vertical (VHO – *Vertical Handover*) ou inter-tecnologias (*Intertechnology Handover*) que, sem dúvida necessita de maiores cuidados em sua gerência [Dutta ET AL, 2008].

Há também a classificação que se preocupa com os domínios envolvidos no processo de *handover*. Nesse caso, um *handover* pode ser considerado Intra-domínio (*Intradomain*), para situações nas quais as redes envolvidas fazem parte do mesmo domínio administrativo, ou Inter-domínios (*Interdomain*), situação observada quando o móvel passa de uma rede pertencente a um domínio administrativo para outra rede, de outro domínio administrativo.

De outra maneira, o *handover* pode ser também classificado em relação à sua execução. Nessa classificação, encontramos:

1. o *handover* assistido pelo móvel (MAHO – *Mobile Assisted Handover*), que, embora executado pela rede, é baseado em informações de qualidade da conexão medidas no móvel;
2. o *handover* iniciado pela rede (NIHO – *Network Initiated Handover*), no qual o móvel não possui nenhuma participação quer seja na identificação, quer seja na execução do *handover*, ficando todas as decisões a critério da rede; e
3. o *handover* controlado pelo móvel (MCHO – *Mobile Controlled Handover*), que, por sua vez, é identificado e executado integralmente pelo móvel.
4. o *handover* controlado pela rede (NCHO – *Network Controlled Handover*), em que a rede toma a decisão a respeito do handover e instrui o móvel com relação à sua execução.

Em relação ao momento de execução das diversas atividades componentes de um *handover*, é possível dividir o procedimento em três fases distintas, conforme a Figura 4.12.

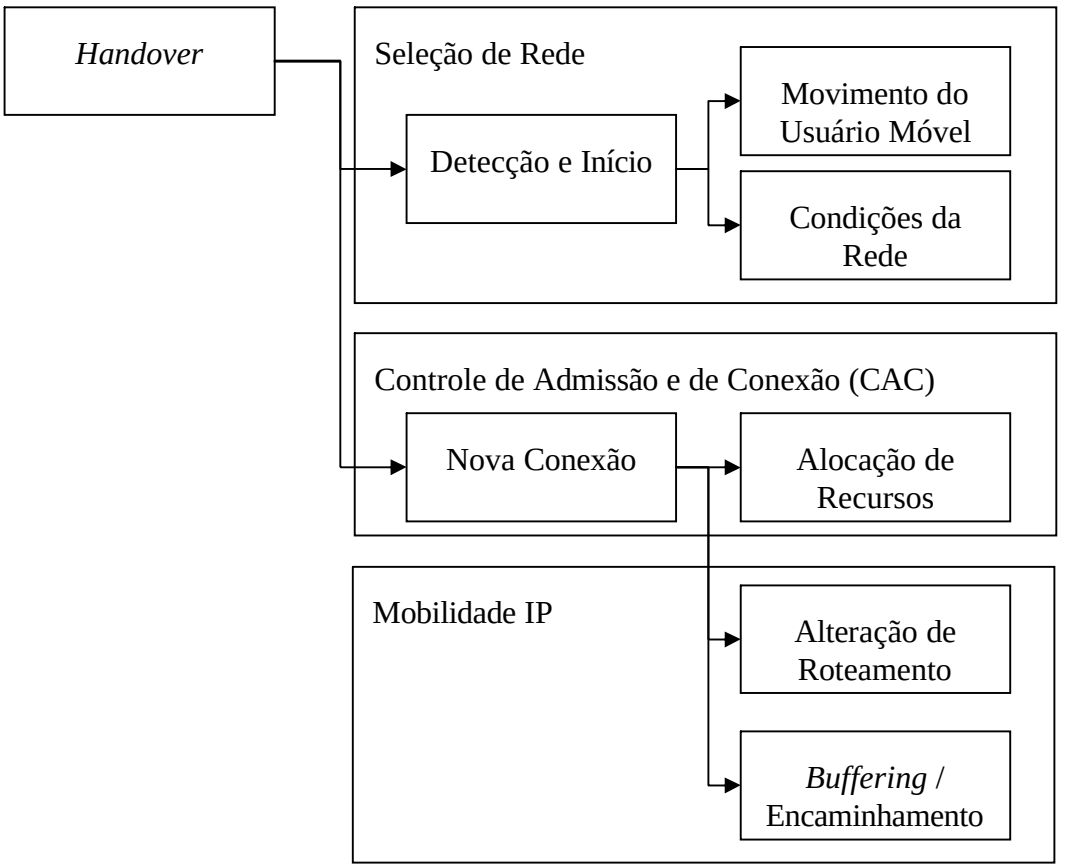

Figura 4.12 – fases do *Handover*

Percebe-se que a integração entre as soluções apresentadas para os problemas de seleção de redes (*Network Selection*), controle de admissão de conexão (CAC) e gerência de mobilidade IP (*Mobility* IP) é essencial para o sucesso da arquitetura de integração entre redes sem fio, visto que juntas elas tentam atender aos requisitos de um *handover* eficiente, a saber:

- A detecção e o início do *handover* devem acontecer antes da perda da conexão atual do móvel (*soft handover*);
- A melhor rede deve ser sempre selecionada antes da execução do *handover*;
- O *handover* não pode deixar de ser finalizado devido à falta de recursos na nova rede;
- A latência do *handover* deve ser a menor possível, visando atender aos requisitos de aplicações com altos requisitos de qualidade de serviço (QoS).

Cabe salientar que esse último requisito é o ponto fundamental no qual os processos de segurança discutidos neste trabalho interferem. Daí a necessidade de prover-se segurança gerando o mínimo de latência possível.

4.7.2. Garantia de QoS durante o *handover* X requisitos de segurança

Os estudos feitos no sentido de integrar redes wireless podem ser categorizados de três formas: os esforços feitos visando à padronização de tecnologias que possam integrar, de forma transparente ao usuário, essas redes; os esforços feitos no sentido de reduzir a latência dos efeitos da mobilidade nessa integração; e os esforços feitos no sentido de garantir ao usuário uma transição transparente entre redes, preservando todos os seus requisitos de qualidade de serviço.

O segundo e o terceiro problemas estão intimamente ligados à execução de processos de segurança durante a passagem de uma rede para outra. A latência, como comentado anteriormente, é o ponto principal aonde a segurança interfere, pois é patente que quanto maior a preocupação com a segurança, maior será o tempo dedicado à execução de processos de segurança, o que poderá gerar prejuízos para o desempenho (segurança X desempenho) geral do processo de *handover*. Além disso, a transição transparente pressupõe a manutenção de sessões de aplicações que porventura estejam sendo realizadas, o que pode ser prejudicado pelo atraso gerado pelos processos de segurança ao processo geral de *handover*.

O problema de garantir qualidade de serviço em redes wireless pode ainda ser sub-dividido em três outros:

1. o problema da seleção de redes;
2. o problema do controle de admissão de conexões; e
3. o problema da reserva e escalonamento de recursos.

Normalmente, estes problemas estão intimamente ligados e suas soluções se fundem quando a intenção é integrar redes sem fio, minimizando os impactos do processo de migração entre as redes envolvidas na integração.

Sendo assim, os aspectos ligados à seleção de redes têm sido abordados, de forma eficiente, levando em consideração as informações medidas diretamente no móvel, o que normalmente fornece uma posição mais real e efetiva sobre a qualidade de cada rede, sob o ponto de vista de quem está efetivamente recebendo o serviço.

No entanto, em uma visão sistêmica do processo de *handover*, não faz muito sentido selecionar uma rede com bons níveis globais de QoS e, ao tentar efetuar o *handover* para a mesma, os mecanismos de CAC e alocação de recursos recusarem tal conexão. Isso implicaria em aumento na latência global do processo, visto que uma nova rede teria que ser selecionada.

Por essa razão, o mais interessante é que a admissão da rede seja um dos parâmetros considerados na função de custo usada pelo mecanismo de seleção de rede, mostrando a interligação necessária entre as soluções para cada uma dessas fases do processo de *handover*.

Por outro lado, considerando os aspectos de segurança, as avaliações mostram-se mais complexas, pois é inviável, por exemplo, mensurar previamente a probabilidade de um móvel obter êxito em um processo de autenticação para permitir acesso à nova rede. Salienta-se que o fracasso nesse processo inviabilizará a nova conexão e gerará a

necessidade de seleção de uma nova rede, aumentando consideravelmente a latência e provavelmente gerando a queda das sessões de aplicações em execução.

Em um contexto onde temos redes sem fio heterogêneas (WLANs e redes celulares, por exemplo), podemos contar com duas situações distintas:

1. o móvel possui interfaces distintas que permitam o mesmo estar conectado às duas redes ao mesmo tempo (multimodo); e
2. o móvel está sempre conectado usando apenas uma de suas interfaces.

No primeiro caso, as redes teriam coberturas sobrepostas, podendo acontecer o *soft handover* com característica vertical, tendo em vista as diferenças entre as tecnologias de acesso. No segundo caso, independentemente das redes possuírem áreas de cobertura sobrepostas, a única possibilidade de passagem é o *hard handover*.

Assim, em ambos os casos, os dispositivos móveis, usando seus procedimentos padrão de seleção de redes, tendem a selecionar àquela cujo nível de sinal é julgado melhor. No entanto, apesar dessa avaliação poder ser eficiente para *handovers* horizontais visto que as grandezas medidas em relação ao nível de sinal são as mesmas, isso não ocorre para situações nas quais são necessários *handovers* verticais.

4.7.3. Arquiteturas e tecnologias de integração de redes heterogêneas

Muitos estudos têm adotado formas de reduzir a latência do *handover* entre redes heterogêneas usando MIH [IEEE 802.21], MIP [RFC 3344] e SIP [RFC 3261], esquemas de Controle de Admissão de Conexão e alocação de recursos.

Esses estudos se baseiam em arquiteturas para a integração categorizadas como muito acoplada, pouco acoplada e não acoplada (também conhecida como *peer-to-peer*) [Munasinghe & Jamalipour, 2008] [Song ET AL., 2007]. Os conceitos que norteiam essa categorização podem ser resumidos como segue.

Na arquitetura muito acoplada, a WLAN está diretamente ligada ao núcleo da rede celular, fazendo com que tanto o tráfego de dados quanto o de sinalização, sejam roteados através do núcleo da rede celular antes de chegar à rede IP externa. Portanto, as técnicas de gerência de mobilidade usadas em redes celulares podem ser aplicadas diretamente sobre a WLAN, considerando que a mesma é parte integrante da rede celular.

Por outro lado, em arquiteturas pouco acopladas, a troca de sinalização entre a WLAN e a rede celular é feita através do núcleo da rede celular, enquanto que os fluxos de dados são encaminhados diretamente à rede IP externa.

Uma vez que o tráfego de dados é encaminhado diretamente a uma rede IP externa, este método pode ajudar a evitar um gargalo de tráfego no núcleo da rede celular, fazendo com que esse método seja mais eficiente nas entregas de dados e, portanto, a mobilidade de sessões de conexões que exijam mais QoS possa ser garantida mais facilmente [Yang and Deng, 2007].

Finalmente, as arquiteturas *peer-to-peer* consideram as redes celulares e WLAN como redes independentes, interligadas normalmente através de gateways. Essa arquitetura pode ser vista como uma variante da arquitetura pouco acoplada [Yusof ET AL., 2007]. Neste caso, a gerência de mobilidade deve ser realizada por um protocolo de camada superior como Mobile IP (MIP) [Munasinghe & Jamalipour, 2007].

4.7.4. O padrão IEEE 802.21

O padrão IEEE 802 não suporta o *handover* vertical entre os diferentes tipos de redes. Ele também não fornece mecanismos para facilitar o *handover* vertical em um ambiente de redes heterogêneas. Portanto, um novo padrão, chamado IEEE 802.21, foi proposto com o objetivo de preencher essa lacuna bastante atual.

Ele possui algoritmos para permitir o *seamless handover* (*handover* suave) entre diferentes tipos de rede, através do fornecimento de informações que permitam a entrega de dados entre redes celulares, GSM, GPRS, WLAN, Bluetooth, e WiMaX através de diferentes mecanismos de transmissão. Este padrão também é chamado de MIH (*Media Independent Handover – Handover* Independente do Meio) [da Silva, 2009].

Os objetivos do padrão IEEE 802.21 podem ser resumidos como [Machan ET AL., 2008]:

- Habilitar mobilidade e *seamless handover* entre redes heterogêneas sem fio;
- Incluir definições de objetos gerenciados que são compatíveis com as normas de gerência do SNMP (*Simple Network Management Protocol* – Protocolo Simples de Gerenciamento de Rede);
- Incluir a definição de um nova abstração de serviço na camada de enlace para fornecer uma tecnologia com interface comum e independente da camada física;
- Definir um conjunto de funções que, interagindo com as camadas superiores possam executar, de forma eficiente, *handover*s verticais;
- Fornecer suporte para autenticação, autorização e detecção/seleção de rede.

Esse último objetivo está totalmente ligado ao tema abordado neste curso e tem sido alvo de estudos recentes realizados por um grupo de trabalho vinculado ao padrão MIH e que preocupa-se exclusivamente com os aspectos de segurança a serem considerado pelo padrão.

Quando um usuário se movimenta entre dois pontos de acessos utilizando a mesma tecnologia, o *handover* geralmente pode ser realizado usando métodos nativos da própria tecnologia sem fio, sem envolver a função MIH (MIHF – *MIH Fuction*). Dessa forma, durante a realização de uma sessão de uma determinada aplicação (uma chamada de voz sobre IP, por exemplo), um nó móvel conectado a uma rede (WLAN ou rede celular) e envolvido nessa sessão pode se movimentar de um ponto de acesso para outro na mesma rede (*handover* intra-tecnologia) e utilizar mecanismos internos ao padrão considerado (IEEE 802.11, GSM, UMTS, etc.) para tratar esse *handover*.

Contudo, se o *handover* ocorre motivado pela movimentação de um equipamento entre um ponto de acesso WLAN localizado em uma rede corporativa e outro ponto de acesso de uma rede celular pública (tipicamente uma Estação Rádio Base – ERB), por exemplo, então o MIH é necessário, pois os dois pontos de acesso pertencem a tecnologias distintas.

As principais funcionalidades fornecidas pelo MIH residem na possibilidade de comunicação entre redes sem fio com tecnologias de enlace diferentes e entre elas e a camada de rede (IP). Este procedimento de *handover* utiliza as informações dos MN (*Mobile Node* - Nó Móvel) e das infra-estruturas das redes.

O padrão IEEE 802.21 informa a disposição das próximas redes ao MN e o ajuda a detectar e selecionar a melhor rede. Esta informação inclui dados sobre a camada de enlace de cada rede. O MIH também pode se comunicar com vários protocolos das camadas superiores que utilizam o IP (SIP e MIP, por exemplo).

A MIHF engloba três tipos de serviços que têm o objetivo de facilitar o *handover* entre redes heterogêneas:

- O MIES (*Media Independent Event Service* – Serviço de Evento Independente do Meio): dá suporte à transferência, filtragem e classificação de alterações dinâmicas da camada de enlace para a camada de rede. Isto gera eventos que informam o estado do enlace (“up/down”) e/ou a disponibilidade de um novo enlace.
- O MICS (*Media Independent Command Service* - Serviço de Comando Independente do Meio): possibilita o controle e gerência das características do enlace da rede que possam contribuir para a decisão do *handover*. O MICS oferece as funções para que a camada de rede possa gerenciar e controlar a camada de enlace.
- O MIIS (*Media Independent Information Service* - Serviço de Informação Independente do Meio): oferece as informações que são necessárias para realizar o *handover*. Ele define um serviço que fornece informações contendo uma lista de redes disponíveis, a versão do protocolo IP utilizado e os dados sobre as operadoras dessas redes, visando à realização de *handover*s mais rápidos. Usando estas informações, o terminal móvel é capaz de tomar uma decisão sobre o *handover*.

As mensagens criadas a partir desses serviços são retransmitidas pela MIHF que está localizada entre a camada de enlace e a camada de rede. O MIHF fornece interfaces homogêneas e independentes de tecnologias de enlace utilizadas pelas redes. Estas interfaces manipulam a comunicação entre as camadas 3 e 2, conforme ilustrado na Figura 4.13.

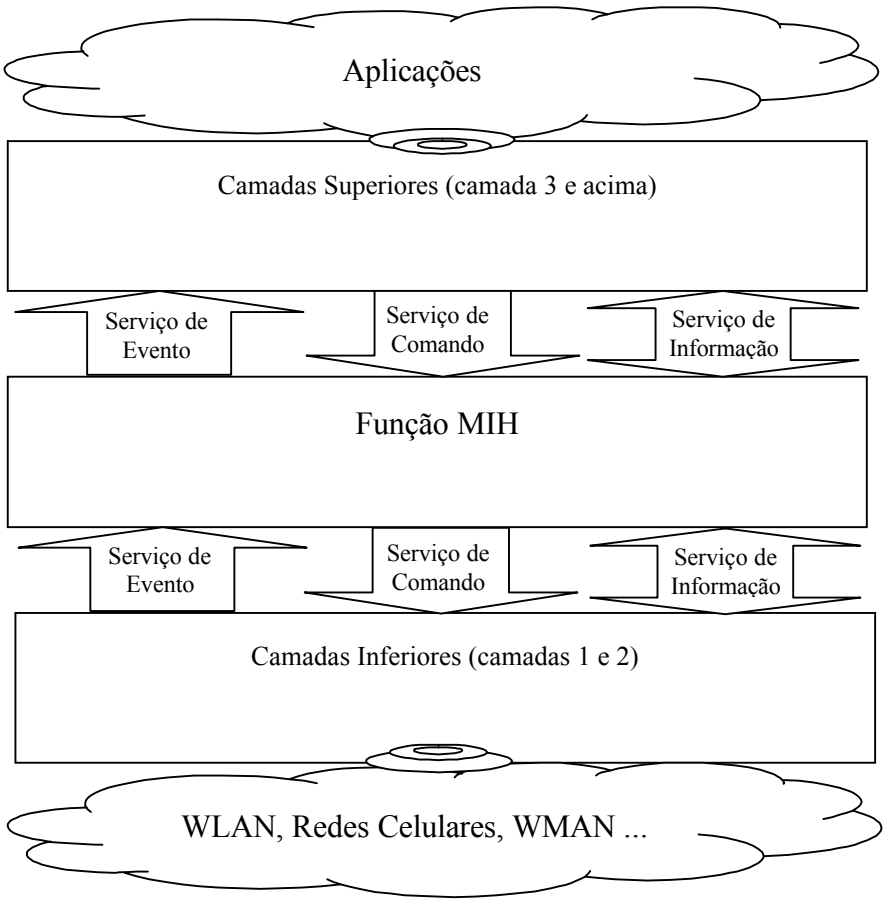

Figura 4.13 – arquitetura do padrão IEEE 802.21 - MIH

Apesar de viabilizar uma série de ferramentas a serem utilizadas para a consecução de *handovers* verticais, a especificação base do padrão [IEEE 802.21] não trata de aspectos de segurança ligados, por exemplo, ao problema da autenticação dos equipamentos móveis. Essa lacuna, entretanto, tem sido alvo de um novo grupo de trabalho (802.21-a *Task Group*) vinculado ao padrão e que foi criado especificamente para tratar dos problemas de segurança.

Esse grupo considera, de uma forma geral, os seguintes passos como necessários para viabilizar segurança no acesso e utilização das redes envolvidas nos *handovers* verticais:

1. Autenticação para acesso à rede – passo fundamental que permite a identificação, por parte de um servidor de autenticação, do usuário (nó móvel) a fim de que sejam verificados seus privilégios para que seja viabilizado um acesso adequado;
2. Criação de uma associação segura – compartilhamento e derivação de parâmetros que permitirão a troca segura de informações entre as partes envolvidas na comunicação aérea (nó móvel e ponto de acesso) do processo;

3. Controle de acesso e cifração de dados – o controle de acesso permite que somente usuários autenticados acessem aos serviços adequados, já a cifração viabiliza a proteção das informações que trafegarão entre as partes envolvidas.

Alguns outros detalhes sobre algumas propostas que têm sido feitas por esse grupo (e por outros acadêmicos que estão preocupados com os aspectos de segurança vinculados ao padrão IEEE 802.21) poderão ser observadas na seção 4.8 deste trabalho.

4.8.Protocolos de segurança aplicados no *Handover* Inter-tecnologias

Nesta seção serão tratados alguns protocolos de autenticação envolvidos nas operações de *handover* inter-tecnologias. Dessa forma, serão apresentados os protocolos MPA (*Media Independent Pre-Authentication* – Pré-autenticação Independente do Meio), que será priorizado tendo em vista sua importância no contexto tratado, o HOKEY (*Handover Keying* – "Chaveamento" de *Handover*) e o PANA (*Protocol for Carrying Authentication for Network Access* – Protocolo para Transporte de Autenticação para Acesso a Rede). Cabe salientar que a apresentação desses protocolos tem o objetivo de ilustrar os diversos esforços que têm sido feitos no contexto da integração de padrões existentes para a resolução do problema tratado nesse curso.

4.8.1. MPA

O MPA é um *framework* que encontra-se atualmente em estudo por um grupo de trabalho do IETF (*Internet Enginnering Task Force* – Força Tarefa de Engenharia da Internet). O último rascunho do padrão publicado pelo IETF (*Internet Enginnering Task Force* – Força Tarefa de Engenharia da Internet) data de 16 de abril de 2010 [Dutta ET AL, 2010]. Ele é considerado um mecanismo de otimização de *handover* pois procura minimizar o tempo de latência gerada pelos diversos processos de autenticação que podem estar envolvidos na passagem de uma rede para outra. O framework focaliza não apenas *handover* inter-tecnologias, mas também, e principalmente, *handover* inter-domínios. Ele propõe-se a trabalhar sobre qualquer camada de enlace e com qualquer protocolo de gerenciamento de mobilidade, incluindo Mobile IP [RFC3344], Mobile IPv6 [RFC3775], MOBIKE [RFC4555], HIP [RFC5201]. Possui operações de pré-autenticação e pré-configuração que permitem que muitas das operações inerentes a um *handover* ocorram antes da passagem do equipamento móvel para a outra rede.

Foram consideradas, dentre outras, as seguintes idéias na elaboração da proposta do *framework*:

- Otimização de mobilidade unificada que funcione com qualquer protocolo de gerenciamento de mobilidade;
- Possibilidade de trabalhar em domínios administrativos diferentes de forma segura, na qual exista um relacionamento de confiança entre o terminal móvel e cada domínio administrativo, sendo que não necessariamente esses domínios devam ter associações de segurança firmadas previamente entre eles;

- Suporte a terminais de diversos tipos, dos que têm múltiplas interfaces e que suportam conexões simultâneas com mais de uma rede aos que possuem apenas uma interface.

Essas idéias deram origem a requisitos que são atendidos por meio de quatro procedimentos básicos do MPA [Dutta ET AL, 2008]:

1. Pré-autenticação, na qual há o estabelecimento de uma SA (*Security Association* - Associação de Segurança) entre o equipamento móvel e a CTN (*Candidate Target Network* – Rede Candidata Alvo) para viabilizar a segurança do protocolo de sinalização que será executado em seguida;
2. Pré-configuração, onde é executado um protocolo de configuração que permite a obtenção, por parte do equipamento móvel, de um endereço IP e de outros parâmetros obtidos a partir da CTN;
3. Execução de um protocolo de tunelamento que estabelece o que é conhecido como PHT (*Proactive Handover Tunnel* – Túnel Pró-ativo de *Handover*) entre o equipamento móvel e um roteador de acesso à CTN (salienta-se que os endereços internos do túnel que permitem o tráfego dos pacotes contêm o endereço IP obtido na fase anterior);
4. Exclusão do PHT, o que é feito imediatamente antes do equipamento móvel conectar-se diretamente à CTN, e re-atribuição do endereço interno do túnel excluído à interface física, o que é feito imediatamente após a conexão à nova rede.

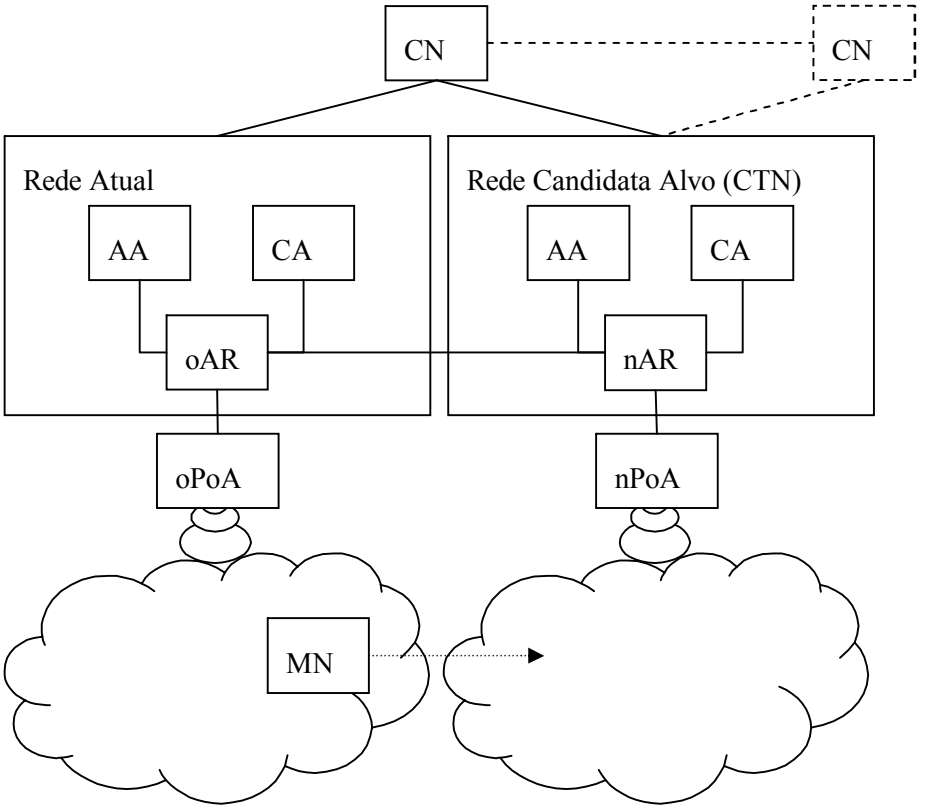

Figura 4.14 – Elementos funcionais do MPA (adaptado de [Dutta ET AL, 2010])

Os elementos funcionais que fazem parte do MPA (ou estão vinculados a este) podem ser entendidos por meio da Figura 4.14:

- CN (Core Network – *Núcleo da Rede*) – não faz parte do *framework* mas está ligado diretamente a este (o MPA prevê a interligação entre núcleos de redes diferentes, o que pode viabilizar, por exemplo, que a CTN esteja vinculada ao núcleo de outra rede).
- AA (*Authentication Agent* – Agente de Autenticação) – responsável pela pré-autenticação do MN por meio da execução de um protocolo de autenticação que permite a geração de uma Associação de Segurança MPA (MPA-SA). Note que esse protocolo de autenticação deve prever a possibilidade de contato do AA com servidores de AAA (do núcleo da própria rede ou de outra vinculada) com o objetivo de se obter material a ser utilizado na criação de chaves que viabilizarão a criação da MPA-SA considerada. Salienta-se que o EAP e seus métodos podem ser considerados protocolos adequados à pré-autenticação tratada no âmbito do MPA.
- CA (*Configuration Agent* – Agente de Configuração) – é o agente responsável pela execução segura do protocolo de pré-configuração que viabilizará a entrega de um endereço IP (e outros parâmetros de configuração) para o MN. Essas informações de configuração devem ser protegidas a partir de material derivado da MPA-SA.
- AR (*Access Router* – Roteador de Acesso) – é o roteador responsável pelo restante do processo de configuração do MN baseado na execução de um protocolo de gerenciamento de túnel para estabelecer o PHT (tratado anteriormente). Dessa forma, todos os pacotes IP transmitidos por meio do PHT deverão ser protegidos por meio de chaves criadas a partir da MPA-SA.
- PoA (*Point of Attachment* – Ponto de Conexão) – elemento de se comunica diretamente com o MN e o AR, podendo estar embutido fisicamente neste último.
- MN (*Mobile Node* – Nó Móvel) – entidade que se encontra em movimento e gera a necessidade de realização do *handover*.

O fluxo básico de mensagens do protocolo MPA pode ser entendido por meio da Figura 4.15:

1. Inicialmente é feita a descoberta das CTN (as comunicações utilizam o endereço de transporte antigo – oCoA – *old Carry-of-Address*). Após isso, tendo sido percebida a queda de sinal (abaixo do limiar 1 configurado) vinculado à conexão com a rede atual, é iniciada a fase de pré-autenticação na qual é criada a MPA-SA que viabiliza a distribuição de chaves criptográficas entre as partes interessadas. Nesse passo são distribuídas as chaves MN-CA *Key* (para a comunicação segura entre o MN e o CA), MN-AR *Key* (análoga à anterior) e

PSK (*Preshared Key* – Chave pré-compartilhada, que viabilizará a troca segura de mensagens entre o nPoA e o MN).

2. Tendo sido constatada uma alta probabilidade de migração para uma dada CTN, executa-se a pré-configuração que tem como principais objetivo a obtenção de um novo endereço de transporte (nCoA – *new CoA*) e o estabelecimento do PHT.
3. Um vez que seja tomada a decisão de se realizar o *handover* para a CTN escolhida, é iniciado o *handover* pró-ativo seguro com a execução da alteração de atribuição (*binding*) por meio do protocolo de gerenciamento de mobilidade e com a utilização do PHT criado anteriormente (já será utilizado o nCoA em substituição ao oCoA).
4. Concluída a alteração de atribuição e estando pronto para realizar a troca de rede, o MN pode executar o protocolo de gerenciamento de túnel para excluir o PHT em uso.
5. Após a exclusão do PHT é iniciado, no nAR (AR da CTN escolhida), o armazenamento temporário dos pacotes em trânsito e, uma vez conectado ao nPoA (*new* PoA), o MN envia um sinal para o nAR a fim de que os pacotes armazenados temporariamente sejam encaminhados para o nPoA.
6. A troca de rede será procedida em seguida e dependerá da política de *handover* utilizada (na figura, a análise está centralizada na razão sinal-ruído – SNR). A partir daí deverá ocorrer um *handover* na camada inferior (camada 2) que poderá embutir um a criação de uma nova SA. Nesse caso, os parâmetros obtidos do processo de pré-autenticação do MPA poderão ser utilizados para minimizar a execução de protocolos completos de autenticação (EAP, por exemplo) resumindo-os a um processo de aperto de mãos de quatro fases (*four way handshake*).
7. Tendo sido concluída a troca de rede, passa-se à fase pós-troca na qual o MN atribui o nCoA à sua interface física que está ligada ao nPoA, o que permite que o nAR pare de armazenar os pacotes.
8. Está concluído o *handover* e os dado podem trafegar normalmente na nova infra-estrutura.

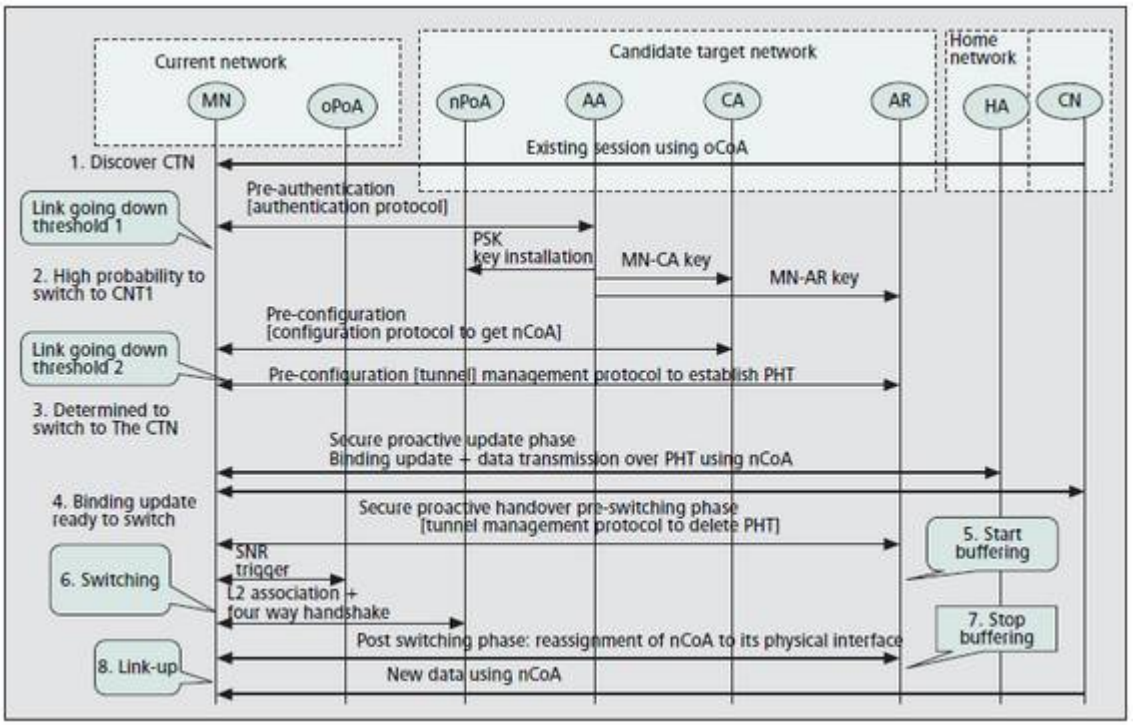

Figura 4.15 – Fluxo de mensagens do MPA (retirado de [Dutta ET AL, 2010])

Fazendo uma ligação com o assunto explorado na seção 4.7.4 (MIH), cabe salientar que o protocolo MPA visto nesta seção pode ter seu papel complementado pelo MIH pois, enquanto o um (MPA) dedica-se mormente à consecução de um *handover* inter-tecnologias pró-ativo seguro, os serviços do outro (MIH) podem fornecer informações valiosas para auxiliar das fases de inicialização e preparação de um *handover*. A integração entre os dois protocolos tem sido alvo de estudos promissores particularmente realizados pelo mesmo grupo de trabalho criado para analisar soluções de segurança para o padrão MIH [Tauil ET AL, 2010].

4.8.2. HOKEY

O HOKEY é uma arquitetura que tem como principal objetivo minimizar o atraso causado pelos processos de autenticação baseados em EAP existentes em *handovers*. Para que isso seja possível, são utilizadas duas abordagens, a autenticação prévia (*early authentication*), a qual permite que a autenticação seja realizada antes do próprio processo de *handover*, e a re-autenticação rápida (*fast reauthentication*), que pode ser aplicada a quaisquer métodos baseados em EAP e na qual é reutilizado material criptográfico gerado na autenticação inicial [Hoper ET AL, 2010].

Em relação à re-autenticação rápida (assunto já tratado na seção 4.5), cabe salientar que em um processo de re-autenticação completa baseada em EAP, ocorrerá a troca de várias mensagens entre o suplicante e o servidor EAP envolvido (fato este que já ocorreu durante a autenticação). O atraso gerado por essa nova troca de mensagens provavelmente interferirá nas transações de dados das quais um equipamento móvel

esteja participando, degradando o serviço e, na pior das hipóteses, forçando a sua interrupção. Esse problema é exatamente o que a re-autenticação rápida busca minimizar.

O problema da autenticação prévia vinculada ao EAP é explorado em detalhes na [RFC 5836] e consiste basicamente na execução de um processo parcial de autenticação (basicamente o estabelecimento de material para criação de chaves a serem utilizadas em uma autenticação EAP) entre um nó móvel (MN) e um ponto de conexão (PoA) potencialmente interessante para a consecução de uma futura conexão, o que viabiliza uma diminuição relevante no processo de autenticação EAP propriamente dito que poderá ocorrer no futuro. Tal qual no caso da re-autenticação rápida, a autenticação prévia também permite uma diminuição no tempo gasto para os processos de autenticação necessários a um *handover*.

A Figura 4.16 ilustra de forma geral processos de autenticação e re-autenticação (na rede de origem – *Home Network* - e em uma rede visitada – *Local Network*) viabilizados por meio do HOKEY. A setas existentes na figura têm o seguinte significado:

1. Execução do método EAP completo entre um MN e a rede de origem da qual faz parte um servidor de AAA (*Home AAA Server*);
2. Distribuição das chaves do servidor AAA para o primeiro AP (*Access Point* - ponto de acesso);
3. Processo de aperto-de-mãos em quatro fases (*four-way handshake*) executado entre o MN e o primeiro AP utilizando o material de chaves recebido em 2;
4. O MN deseja conectar-se ao segundo AP que se encontra em outro domínio administrativo (*handover* INTER-domínios), para tal executa o método ERP (Re-autenticação EAP vista na seção 4.4.2) apoiado pelo servidor de AAA da rede visitada (*Local AAA Server*);
5. O servidor de AAA da rede visitada faz requisições ao servidor de AAA da rede de origem e recebe uma DSRK (*Domain-Specific Root Key* – Chave Raiz de Domínio Específico) a ser utilizada na re-autenticação;
6. O servidor de AAA da rede visitada gera chaves de sessão a partir da DSRK recebida e as entrega para o segundo AP;
7. Processo de *four-way handshake* executado entre o MN e o segundo AP utilizando o material de chaves recebido em 6;
8. O MN deseja conectar-se ao terceiro AP que também se encontra no outro domínio administrativo (*handover* INTRA-domínios), para tal executa o método ERP apoiado pelo servidor de AAA da rede visitada;
9. O servidor de AAA da rede visitada gera novas chaves de sessão a partir da DSRK que já se encontra com ele e as entrega para o terceiro AP;
10. Processo de *four-way handshake* executado entre o MN e o terceiro AP utilizando o material de chaves recebido em 9.

Cabe salientar que, no exemplo explorado, as redes de origem e visitada poderão, além de fazer parte de domínios diferentes, estar baseadas em tecnologias de acesso distintas, o que caracterizará também o *handover* INTER-tecnologias que interessa a esse trabalho

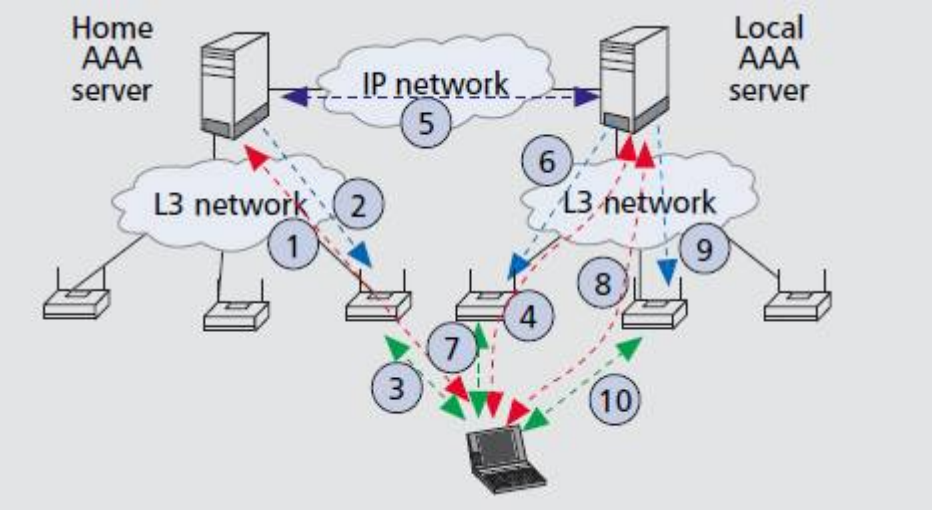

Figura 4.16 – Autenticação e Re-autenticação com HOKEY (retirado de [Clancy, 2008])

4.8.3. PANA

O PANA é descrito pela [RFC 5191] e, diferente de mecanismos constantes de sub-sessões anteriores, é um protocolo executado na camada de rede (baseado em UDP) e viabiliza, nas camadas superiores, a utilização de métodos de autenticação para controle de acesso à rede. De forma mais específica, pode-se dizer que o PANA representa uma camada inferior (modelo conceitual em camadas do EAP – item 4.4.1) sobre a qual funciona o *framework* EAP. Em outras palavras, o PANA "carrega" o EAP (que por sua vez pode "carregar" diversos métodos EAP).

A partir do momento em que viabiliza o emprego do *framework* EAP sobre a camada IP, o PANA permite a execução de quaisquer métodos EAP sobre quaisquer tecnologias de enlace. Dessa forma o PANA é uma ferramenta valiosa no auxilio à execução de *handovers* entre tecnologias de acesso distintas (WLAN e 3G, por exemplo).

A troca de mensagens PANA consiste de uma série de requisições e respostas, tal qual os outros protocolos vistos neste curso, que podem conter vários AVPs (o conceito de AVP já foi explorado na seção **Erro! Fonte de referência não encontrada.** deste trabalho) transportando várias cargas (*payload*). As principais cargas transportadas por esses AVPs são os pacotes EAP.

Por ser baseado em UDP, o PANA precisa implementar seu próprio mecanismo de retransmissão a fim de que haja confiabilidade no processo de troca de mensagens do protocolo.

Uma sessão PANA deve ser executada entre um cliente (PaC – *PANA Client* – Cliente PANA) e um servidor (PAA – *PANA Authentication Agent* – Agente de Autenticação PANA) para que seja possível o fornecimento de autenticação e autorização (AA) aos serviços de controle de acesso a redes. Essa sessão consiste das seguintes fases:

1. Fase de autenticação e autorização - etapa na qual é iniciada uma sessão PANA e onde é executado o método EAP entre o PaC e o PAA (qualquer um dos dois pode iniciar a sessão). Ao término desta fase, o PAA fornece o resultado do processo (AA) para o PaC.
2. Fase de acesso – tendo havido sucesso na fase anterior (AA), o PaC passa a ter acesso à rede almejada e pode realizar seus envios e recebimentos de pacotes de dados normalmente.
3. Fase de re-autenticação – ainda durante a fase anterior (por esse motivo esta fase é considerada uma sub-fase da anterior), o PaC ou o PAA (qualquer um dos dois) podem iniciar a re-autenticação para que seja possível alterar o tempo de vida (*lifetime*) da sessão PANA antes que este expire.
4. Fase de finalização – pode ser, tal qual as anteriores, disparada por qualquer uma das duas entidades envolvidas (PaC ou PAA) a qualquer momento (decisões alheias ao PANA). A entidade interessada no término da sessão remete um sinal (mensagem de desconexão) para a sua parceira e o processo é finalizado.

A Figura 4.17 retrata um cenário típico e utilização do PANA. Neste cenário, inicialmente um MN (PaC) realiza procedimentos de autenticação (utilizando o PANA) a uma rede localizada no domínio A (fase 1 explicada anteriormente). Tendo havido êxito no processo, o MN realiza o acesso à rede e inicia suas transações. Esse MN, desejando movimentar-se para outro domínio (o C no exemplo) mas ainda conectado ao domínio A (acessando a rede original), dispara um processo de re-autenticação (fase 3) direcionado para o domínio C. O contato a ser feito pelo MN com o PAA da rede alvo (localizada no domínio C), aqui representado pelo servidor de AAA do domínio C, deve ser feito de forma direta (PANA não permite, por exemplo, que o PAA atual sirva de *proxy* para o contato com o próximo PAA), entretanto isso não se mostra algo problemático tendo em vista o fato do PANA estar baseado no UDP. Tendo havido sucesso na re-autenticação, o MN pode migrar (*handover*) sem problemas de uma rede para a outra.

Como é possível perceber, o PANA preocupa-se tão somente com os processos de autenticação e re-autenticação, deixando a atividade de *handover* em si para outros protocolos necessariamente suportados pelo MN e demais participantes do processo (salienta-se que o PANA é um dos protocolos propostos para utilização vinculada ao padrão MIH pelo grupo de trabalho que se preocupa com os aspectos de segurança do padrão IEEE 802.21).

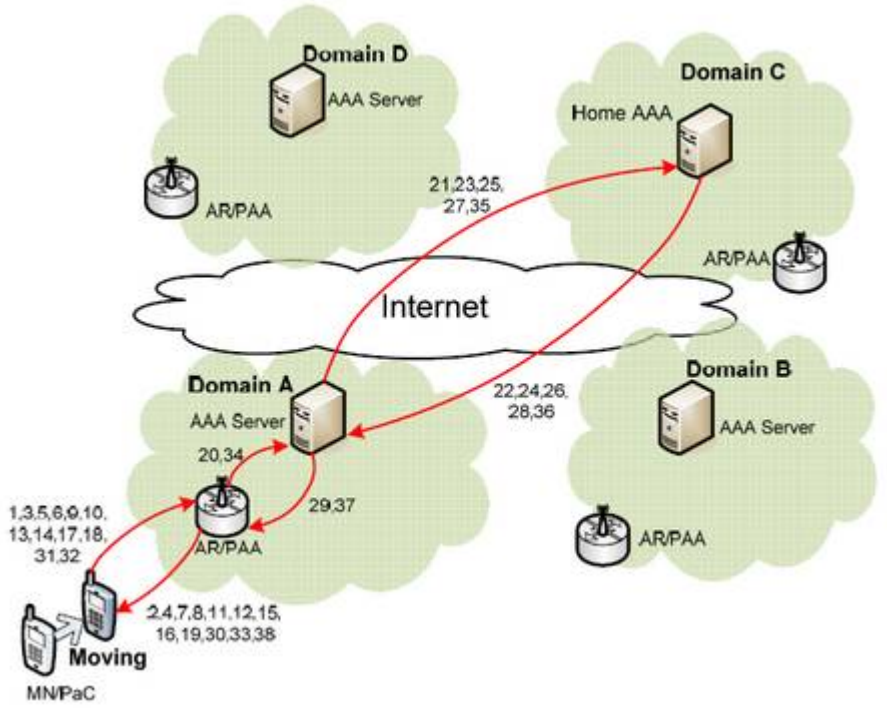

Figura 4.17 – Cenário típico de aplicação do PANA (retirado de [Chamuczynski, 2008])

4.9. Conclusões

O curso encerrado nessa seção teve os objetivos de apresentar aspectos teóricos e práticos ligados à segurança necessária ao *handover* entre redes celulares e WLANs, além de focalizar formas de minimizar a latência gerada pelos processos que viabilizam essa segurança.

Os autores procuraram explorar uma gama bastante grande de assuntos, em muitos casos com pequeno nível de detalhamento, vinculando padrões clássicos a soluções inovadoras e extremamente atuais, com o objetivo de criar no leitor uma idéia ampla e irrestrita das várias áreas ligadas ao assunto-núcleo explorado.

Ao concluir a leitura deste texto, certamente o leitor não poderá se considerar conhecedor profundo de um determinado assunto tratado no curso, entretanto ele poderá ter a certeza de que passou a ter um conhecimento razoável sobre um conjunto bastante grande de assuntos que estão interligados ao tema central. Além disso, ele também perceberá que a palavra chave a ser utilizada na busca da resolução dos problemas abordados é INTEGRAÇÃO. Integrar de forma sinérgica diversas soluções específicas para gerar o resultado esperado.

Especificamente sobre os problemas que foram explorados, pôde-se perceber que, não só no contexto da segurança, mas também em vários outros aspectos (garantia de QoS, modelos de mobilidade, análise de desempenho, etc.), a questão da interconexão entre redes heterogêneas é algo bastante atual e que certamente ainda demandará muito esforço dos diversos elementos envolvidos (comunidade acadêmica, entidades e órgãos

de padronização, fabricantes de equipamentos e provedores de serviços, dentre outros) para que possa ser considerado um problema bem resolvido.

Verificou-se também que, caso as partes integrantes de um processo de comunicação envolvendo usuários, operadoras de redes celulares e provedores de acesso a WLANs não mantenham um relacionamento de confiança entre si, a existência de uma passagem segura entre rede celular e WLAN para ser utilizada por usuários será algo praticamente inviável e, caso venha a ser possível, poderá ser inócua.

Também pode ser extraída do texto a idéia de que a integração entre WLANs e redes celulares, aproveitando aquilo que de melhor elas podem fornecer para o bom provimento de um serviço, é algo muito útil e tem-se mostrado uma tendência de mercado explorada por muitas entidades interessadas.

Em relação ao padrão MIH, foi possível verificar que ele não tem a capacidade de resolver sozinho o problema central tratado neste curso, entretanto pode ser considerado uma ferramenta preciosa que, caso integrada a outros protocolos com atribuições específicas (MPA, por exemplo), gerará resultados importantíssimos no rumo das soluções procuradas.

Sobre os processos de autenticação, pôde-se perceber que, apesar de sua importância incontestável, eles ainda são grandes "vilões" quando o assunto tratado é a diminuição de tempo de um *handover*. Apesar disso, também foi possível observar que muitos esforços têm sido feitos para a proposição de melhorias e adaptações a serem aplicadas nesses métodos de autenticação. Algumas melhorias utilizam idéias de reuso de material criptográfico criado em momentos passados, outras propõem a realização de processos em momentos anteriores ao *handover*, outras buscam aproveitar ambos os benefícios. De qualquer forma, independente do maior ou menor êxito, fica patente o entendimento e a busca de soluções por parte dos inúmeros pesquisadores do assunto.

Também foi possível constatar que os problemas de segurança realmente interferem na qualidade dos serviços prestados por meio de redes sem fio. Sob qualquer prisma, analisando o enfoque da latência gerada pelos processos que se propõem a resolver esses problemas, ou observando as próprias conseqüências danosas advindas da não aplicação dos mecanismos de segurança, fica claro que a não existência, bem como o excesso de segurança, certamente gerarão conseqüências indesejáveis para os serviços fornecidos pelas redes sem fio.

Nesse contexto, percebe-se que ainda há muito trabalho a ser feito e que o grande desafio não é criar uma solução extremamente segura, nem uma tremendamente rápida, mas sim uma que viabilize a segurança necessária e suficiente sem ser "agressiva", ou seja, que não prejudique o próprio "bem" que está sendo protegido: a informação oportuna.

Referências

Aboba, B., Blunk, L., Vollbrecht, J., Carlson, J., and H. Levkowetz, Ed. (2004). "Extensible Authentication Protocol (EAP)". RFC 3748, June 2004.

Aboba, B., Calhoun, P.(2003). "RADIUS (Remote Authentication Dial In User Service) Support For Extensible Authentication Protocol (EAP)". RFC 3579, September 2003.

Aboba, B., Simon, D., Eronen, P.(2008). "Extensible Authentication Protocol (EAP) Key Management Framework". RFC 5247, August 2008.

Arkko, J., Haverinen, H.(2006). "Extensible Authentication Protocol Method for 3rd Generation Authentication and Key Agreement (EAP-AKA)". RFC 4187, January 2006.

Calhoun, P., Loughney, J., Guttman, E., Zorn, G., Arkko, J.(2003). "Diameter Base Protocol", RFC 3588, September 2003.

Chamuczynski, P., Alfandi, O., Werner, C., Brosene, H., Hogrefe, D.(2008). "Performance Study of PANA Pre-authentication for Interdomain *Handover*". In: Fourth International Conference on Networking and Services, ICNS 2008, March 2008.

Clancy, T.(2008). "Secure *Handover* in Enterprise WLANs: CAPWAP, HOKEY, and IEEE 802.11r". In: IEEE Wireless Communications, vol. 15, nr. 5, October 2008.

Clancy, T., Nakhjiri, M., Narayanan, V., Dondeti, L.(2008). "*Handover* Key Management and Re-Authentication Problem Statement". RFC 5169, March 2008.

Congdon, P., Aboba, B., Smith, A., Zorn, G., Roese, J.(2003). "IEEE 802.1X Remote Authentication Dial In User Service (RADIUS) Usage Guidelines". RFC 3580, September 2003.

Dierks, T., Rescorla, E.(2008). "The Transport Layer Security (TLS) Protocol Version 1.2". RFC 5246, August 2008.

Dutta, A., Fajardo, V., Ohba, Y., Taniuchi, K., Schulzrinne, H.(2010). "A Framework of Media-Independent Pre-Authentication (MPA) for Inter-domain *Handover* Optimization". draft-irtf-mobopts-mpa-framework-07, April 2010, trabalho em andamento.

Dutta, A., Famolari, D., Das, S., Ohba, Y., Fajardo, V., Taniuchi, K., Lopez, R., Schulzrinne, H.(2008). "Media-Independent Pre-Authentication Supporting Secure Interdomain *Handover* Optimization". In: IEEE Wireless Communications, vol. 15, nr. 2, April 2008.

Eronen, P.(2006). "IKEv2 Mobility and Multihoming Protocol (MOBIKE)". RFC 4555, June 2006.

ETSI TS 133 234 (2010). "Universal Mobile Telecommunications System (UMTS); LTE; 3G security; Wireless Local Area Network (WLAN) interworking security". V. 9.2.0, Release 9, July 2010

Forsberg, D., Ohba, Y., Patil, B., Tschofenig, H., Yegin, A.(2008). “Protocol for Carrying Authentication for Network Access (PANA)”. RFC 5191, May 2008.

Funk, P., Blake-Wilson, S.(2008). "Extensible Authentication Protocol Tunneled Transport Layer Security Authenticated Protocol Version 0 (EAP-TTLSv0)". RFC 5281, August 2008.

Haverinen, H., Ed., and J. Salowey, Ed.(2006). “Extensible Authentication Protocol Method for Global System for Mobile Communications (GSM) Subscriber Identity Modules (EAP-SIM)”. RFC 4186, January 2006.

Hoeper, K., Decugis, S., Zorn, G., Wu, Q., Taylor, T.(2010). “*Handover* Keying (HOKEY) Architecture Design”. draft-hoeper-hokey-arch-design-03, July 2010, trabalho em andamento.

Hoeper, K., Nakhjiri, M., Ohba, Y.(2010). “Distribution of EAP-Based Keys for *Handover* and Re-Authentication”. RFC 5749, March 2010.

IEEE Standard 802.11 (2007). “Telecommunications and information exchange between systems - Local and metropolitan area networks - Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications”.

IEEE Standard 802.1X (2004). “IEEE Standard for Local and metropolitan area networks Port-Based Network Access Control”.

IEEE Standard 802.21 (2008). “IEEE Standard for Local and metropolitan area networks - Part 21: Media Independent *Handover* Services”.

Johnson, D., Perkins, C., Arkko, J.(2004). “Mobility Support in IPv6”. RFC 3775, June 2004.

Josang, A., Ismail, R., Boyd, C.(2007). “A survey of trust and reputation systems for online service provision”. In: Decision Support Systems, vol. 43, nr. 2. Elsevier Science Publishers B. V., March 2007.

Kaufman, C.(2005). “Internet Key Exchange (IKEv2) Protocol”. RFC 4306, December 2005.

Koien, G. M., Haslestad, T.(2003). “Security Aspects of 3G-WLAN Interworking”. In: IEEE Communications Magazine, vol. 41, nr. 11, November de 2003.

Liu, J., Jiang, S., Lin, H.(2006). “Introduction to Diameter – Get the next generation AAA protocol”. Available in http://www.ibm.com/developerworks/wireless/library/wi-diameter, April 2006, Access: July 2010.

Machan, P., Serwin, S., and Wozniak, J.(2008). “Performance of mobility support mechanisms in a heterogeneous UMTS and IEEE 802.11 network offered under the IEEE 802.21 standard”. In: 1st International Conference on Information Technology, pages 1–4. IEEE. 2008

Moskowitz, R., Nikander, P., Jokela, P., Henderson, T.(2008). “Host Identity Protocol”, RFC 5201, April 2008.

Munasinghe, K. and Jamalipour, A.(2007). “A 3GPP-IMS based approach for converging next generation mobile data networks”. In: International Conference on Communications, pages 5264–5269. IEEE. 2007.

Munasinghe, K. and Jamalipour, A.(2008). “Interworking of WLAN-UMTS networks: an IMS-based platform for session mobility”. In: IEEE Communications Magazine, Vol. 46, nr. 9, IEEE, 2008.

Narayanan, V., Dondeti, L.(2008). "EAP Extensions for EAP Re-authentication Protocol (ERP)". RFC 5296, August 2008.

Ohba, Y., Wu, Q., Zorn, G.(2010). “Extensible Authentication Protocol (EAP) Early Authentication Problem Statement”. RFC 5836, April 2010.

Perkins, C.(2002). “IP Mobility Support for IPv4”. RFC 3344, August 2002.

Postel, J.(1980). “User Datagram Protocol”. RFC 768, August 1980.

Postel, J.(1981). “Transmission Control Protocol”. RFC 793, September 1981.

Rigney, C., Willens, S., Rubens, A., Simpson, A.(2000). “Remote Authentication Dial In User Service (RADIUS)”. RFC 2865, June 2000.

Romkey, J.(1988). “A Nonstandard for Transmission of IP Datagrams Over Serial Lines: SLIP”. RFC 1055, June 1988

Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., Schooler, E.(2002). “SIP: Session Initiation Protocol”. RFC 3261, June 2002.

Salowey, J., Dondeti, L., Narayanan, V., Nakhjiri, M.(2008). “Specification for the Derivation of Root Keys from an Extended Master Session Key (EMSK)”. RFC 5295, August 2008.

Schneier, B. (1996). "Applied Cryptography: Protocols, Algorithms, and Source Code in C". Second Edition. John Wiley & Sons.

Silva, A., Endler, M. Colcher, S.(2008). “Otimização do *Handover* na Camada de Rede (L3) utilizando o Media Independent *Handover* (MIH)”. Tese não publicada. Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática, Rio de Janeiro. 2008.

Simon, D., Aboba, B., Hurst, R.(2008). “The EAP-TLS Authentication Protocol”. RFC 5216, March 2008.

Simpson, W.(1994). “The Point-to-Point Protocol (PPP)”. RFC 1661, July 1994

Song, W., Jiang, H., and Zhuang, W.(2007). “Performance analysis of the wlan-first scheme in cellular/wlan interworking”. In: IEEE Transactions on Wireless Communications, Vol. 6, nr. 5, IEEE, 2007.

Stanley, D., Walker, J., Aboba, B.(2005). “Extensible Authentication Protocol (EAP) Method Requirements for Wireless LANs”. RFC 4017, March 2005.

Tauil, M., Dutta, A., Cheng, Y., Das, S., Baker, D., Yajnik, M., Famolari, D., Ohba, Y., Taniuchi, K., Fajardo, V., Schulzrinne, H.(2010). “Integration of IEEE 802.21

services and pre-authentication framework". In: International Journal of Communication Networks and Distributed Systems, vol. 5, nr.1/2, February 2010.

Yang, P., Deng, H.(2007). "Seamless integration of 3G and 802.11 wireless network". In: 5th ACM international workshop on Mobility management and wireless access, pages 60–65. ACM. 2007.

Yusof, A. L., Ismail, M., and Misran, N.(2007). "Architecture and mobility management protocols for next-generation wireless systems (NGWS)". In: IEEE International Conference on Telecommunications and Malaysia International Conference on Communications, pages 747–752. IEEE, 2007

Zorn, G.(1999). "Microsoft Vendor-specific RADIUS Attributes". RFC 2548, March 1999.

Zorn, G., Cobb, S.(1998). "Microsoft PPP CHAP Extensions". RFC 2433, October 1998.