

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
INSTITUTO DE MATEMÁTICA
NÚCLEO DE COMPUTAÇÃO ELETRÔNICA
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

JAIME CESAR DE CARVALHO JUNIOR

Sutil e Arprap
Duas Propostas de Melhorias no Serviço de Gerência de Handover

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Informática, IM/NCE, Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Informática.

Orientadores: Luci Pirmez, D.Sc.
Flavia Delicato, D.Sc.

Rio de Janeiro

2009

C331 Carvalho Jr, Jaime Cesar de.

Sutil e Arprap: duas propostas de melhorias no serviço de gerência de Handover. / Jaime Cesar de Carvalho Junior. – 2009.
115 f. il.

Dissertação (Mestrado em Informática) – Universidade Federal do Rio de Janeiro, Instituto de Matemática, Núcleo de Computação Eletrônica, 2009.

Orientadores: Luci Pimez ; Flavia Delicato

1. Sutil – Teses. 2. Arprap – Teses. 3. Seleção de Redes – Teses.
4. Programação Linear Inteira – Teses. I. Luci Pimez (Orient.).
II. Flavia Delicato (Orient.). III. Título

CDD

Jaime Cesar de Carvalho Junior

Sutil e Arprap
Duas Propostas de Melhorias no Serviço de Gerência de Handover

Dissertação submetida ao corpo docente do Programa de Pós-Graduação em Informática do Instituto de Matemática e do Núcleo de Computação Eletrônica, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do grau de Mestre em Informática.

Aprovada em: Rio de Janeiro, 14 de dezembro de 2009.

Prof.^a. Luci Pirmez, D.Sc., NCE/UFRJ (Orientadora)

Prof.^a. Flávia Coimbra Delicato, D.Sc., UFRN(Orientadora)

Prof. Luiz Fernando Rust da Costa Carmo, Docteur, INMETRO

Prof. Célio Vinicius Neves de Albuquerque, Ph.D., IC/UFF

Ao prazer do conhecimento.

Agradecimentos

À minha querida esposa, Lina, pelo apoio irrestrito dado nessa jornada.

Aos meus pais, Jaime Cesar (in memoriam) e Sueli, pelo exemplo e dedicação em me proporcionar uma boa formação, mesmo que para isso necessitasse estar longe de casa.

Aos meus professores por compartilharem comigo sua experiência e conhecimento, e às minhas orientadoras, Luci Pirmez e Flávia Delicato, pela dedicação a esse trabalho.

Resumo

CARVALHO JR, Jaime Cesar de. **Sutil e Arprap**: duas propostas de melhorias no serviço de gerência de Handover. 2009. 115 f. Dissertação (Mestrado em Informática) - Instituto de Matemática, Núcleo de Computação Eletrônica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2009.

O avanço tecnológico das redes sem fio, a recente proliferação de dispositivos portáteis, tais como laptops, PDAs e palmtops, aliados ao aumento de popularidade da computação móvel, levaram ao surgimento de aplicações e ambientes de computação ubíquos. Em tais ambientes, é necessário que as estações (móveis ou não), situadas em qualquer localidade e rodando as aplicações de usuários, estejam permanentemente conectadas a redes de diferentes tecnologias, o que impõe uma série de desafios. Um dos desafios envolvidos é estar sempre conectado à melhor rede, que requer o desenvolvimento de soluções que considerem vários critérios para a tomada de decisão de qual seria essa “melhor rede” quando da necessidade de efetuar a transição de usuários móveis entre diferentes células e operadoras de redes. A fim de solucionar este desafio o presente trabalho propõe um mecanismo para seleção de rede em ambiente de redes de próxima geração (NGN) que prioriza redes com maior relevância para aplicação e menor gasto de energia, contribuindo para que os dispositivos móveis e aplicações dos usuários estejam permanentemente conectados e, por consequência, para a viabilidade de sistemas ubíquos. Outro grande desafio ligado a mobilidade de estações é conseguir uma utilização eficiente de redes homogêneas, provendo serviços com qualidade para seus usuários, principalmente em presença de *handover*. A fim de solucionar esse desafio o presente trabalho propõe um melhoramento do processo de *handover* de forma a torná-lo mais eficiente, ou seja, mais rápido, do que o padrão existente para a rede WIMAX e ao mesmo tempo garantir que a qualidade de serviço para os usuários móveis não se deteriore em virtude da transição (*handover*). Para este segundo desafio é proposto um algoritmo que explora o perfil do comportamento do usuário em movimento para maximizar a eficiência da rede e prover melhor QoS para diferentes classes de usuários em um cenário que permite a integração de redes.

Abstract

CARVALHO JR, Jaime Cesar de. **Sutil e Arprap**: duas propostas de melhorias no serviço de gerência de Handover. 2009. 115 f. Dissertação (Mestrado em Informática) - Instituto de Matemática, Núcleo de Computação Eletrônica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2009.

The technological advancement of wireless networks, the recent proliferation of portable devices such as laptops, palmtops and PDAs, added to the growing popularity of mobile computing leads to the emergence of ubiquitous applications and computing environments. In such environments, it is necessary that the stations (mobile or not) situated in any location and running user applications should be permanently connected to networks of different technologies, which places a number of challenges. One of the challenges involved is to be always connected to the best network, which requires the development of solutions that consider multiple criteria for decision making of what that "best network" when the need to handover mobile users between different cells and operators networks. In order to address this challenge, this paper proposes a mechanism for network selection in the context of Next Generation Networks (NGN), which prioritizes networks with higher relevance to the application and lower energy consumption and it enables full and seamless connectivity to mobile user devices and applications. Consequently, it contributes to realize the vision of Ubiquitous Computing. Another major challenge on the mobility of stations is to achieve efficient use of homogeneous networks, providing quality services to its users, especially in the presence of handover. In order to address this challenge this work proposes an improvement of the handover in order to make it more efficient, ie faster than the existing standard for WiMAX network while ensuring that the quality of service for mobile users not deteriorate because of the handover . For this second challenge it is proposed an algorithm that exploits the behavior profile of the mobile user to maximize the network efficiency while meeting QoS requirements for different classes of users in a scenario that supports the integration of wireless broadband networks.

Lista de Figuras

Figura 1. Padronização IEEE para redes sem fio	1
Figura 2. Divisão em camadas, suas funções e pontos de acesso (SAPs)	13
Figura 3. Suporte a QoS em Redes Móveis WiMAX.....	15
Figura 4. Pilha de protocolos do IEEE 802.16e para BS's.....	22
Figura 5. Cenário de uma rede IEEE 802.16e no momento do Handover	1
Figura 6. Etapas do procedimento de <i>Handover</i> da camada MAC	24
Figura 7. Troca de mensagens na descoberta da topologia da rede.....	1
Figura 8. Etapas detalhadas do processo de Handover.....	1
Figura 9. Processo de reentrada na rede	1
Figura 10. Processo MDHO	1
Figura 11. Processo FBSS	1
Figura 12. Convergência global de serviços de telecomunicações	34
Figura 13. Modelo de convergência NGN.....	34
Figura 14. Diagrama de componentes do SUTIL.....	49
Figura 15. Diagrama de componentes do ARPRAP.....	61
Figura 16. Diagrama de Sequência do ARPRAP	62
Figura 17. Diagrama de atividades de implementação.....	69
Figura 18. Componentes do modelo Simulink proposto	71
Figura 19. Modelo para o problema de maximizar utilidades.....	73
Figura 20. Cálculo do EAR	81
Figura 21. TTS Global.....	81
Figura 22. Escalabilidade e a Restrição R3	86
Figura 23. Diagrama de atividades de simulação	89
Figura 24. PFH em 4 áreas	94
Figura 25. PFH em 8 áreas	95
Figura 26. Melhoria na Transferência de Dados	96
Figura 27. Eficiência na Alocação de Recursos	97

Lista de Tabelas

Tabela 1. Características dos padrões 802.16e	11
Tabela 2. Aplicações de Redes Móveis WiMAX e Qualidade de Serviço.....	16
Tabela 3. Classes de aplicações 802.16e	17
Tabela 4. Exemplos de Requisitos e seus RR [KANG 2006]	53
Tabela 5. Matriz Utilidade (ÁREA da FRENTE)	76
Tabela 6. Matriz Utilidade (ÁREA de TRÁS)	77
Tabela 7. Matriz Utilidade.....	78
Tabela 8. Comparação entre os atrasos e perdas de pacote	82

Sumário

1	Introdução.....	1
1.1	Motivação	3
1.2	Objetivos.....	5
1.2.1	Primeiro objetivo específico.....	5
1.2.2	Segundo objetivo específico.....	7
1.3	Organização	8
2	Conceitos Básicos	9
2.1	Padrões IEEE 802 para redes sem fio.....	9
2.2	O Padrão IEEE 802.16 – WiMAX	12
2.2.1	Arquitetura.....	12
2.2.2	Topologias de Rede	13
2.3	QoS.....	15
2.3.1	Classe de Serviço.....	17
2.3.2	Fluxo de Serviço.....	20
2.4	Padrão IEEE 802.16e – Mobilidade no WiMAX.....	21
2.4.1	Conceitos Básicos.....	22
2.4.2	Tipos de <i>Handover</i>	23
2.4.3	MDHO (Macro Diversity HandOver)	29
2.4.4	FBSS (Fast Base Station Switching)	30
2.5	<i>Next Generation Networks</i> (NGN).....	33
2.6	Trabalhos relacionados.....	35
2.6.1	Trabalhos relacionados ao SUTIL.....	35
2.6.2	Trabalhos relacionados ao ARPRAP.....	42
2.7	Considerações finais do Capítulo	44
3	Seleção de Redes baseada em Função de Utilidade	45
3.1	Visão geral e formulação do problema.....	45
3.2	Arquitetura do SUTIL	48
3.3	Operação do SUTIL	50
3.3.1	Eliminar Redes e Construir Serviço Representativo	50
3.3.2	Calcular os valores da Função Utilidade	54
3.3.3	Selecionar Redes	56
3.4	Considerações finais do capítulo	57
4	Reserva Pró-ativa de Recursos Assistida por Perfil	58
4.1	Visão geral e formulação do problema.....	58
4.1.1	Arquitetura do ARPRAP	59
4.2	Descrição do Algoritmo ARPRAP	61
4.3	Mensagens Trocadas no ARPRAP.....	64
4.4	Considerações finais do capítulo	65
5	Implementação e Análise de Resultados do SUTIL.....	66
5.1	O Ambiente para Implementação e Simulação do SUTIL	66
5.2	Implementação.....	68
5.2.1	Construção dos Serviços Representativos (Script M)	70
5.2.2	Cálculo de utilidades (Simulink).....	71
5.2.3	Maximização PLI (AMPL/Cplex).....	72
5.3	Estudo de Caso 1: Aplicando SUTIL em Cenário próximo do real	73
5.3.1	Descrição	73

5.3.2	Avaliação	76
5.4	Estudo de Caso 2: Aplicando SUTIL em Cenário Reduzido.....	78
5.4.1	Descrição	78
5.4.2	Avaliação Numérica dos Resultados para o Cenário Reduzido	79
5.5	Comparando SUTIL com trabalho em [CHEN 2005]	83
5.6	Perfis de QoS.....	84
5.7	Estudo de Caso 3: Aplicando SUTIL em Cenários Maiores	85
5.8	Considerações finais	87
6	Implementação e Análise de Resultados do uso do ARPRAP	88
6.1	O ambiente de Implementação e Simulação do ARPRAP.....	88
6.2	Esquema das simulações	88
6.2.1	Implementação	89
6.3	Simulações e Análises dos resultados do uso do ARPRAP.....	90
6.3.1	Cenários das Simulações	90
6.3.2	Métricas de Desempenho	92
6.3.3	Probabilidade de Falha de <i>Handover</i>	93
6.3.4	Melhoria na Transferência de Dados.....	95
6.3.5	Eficiência na Alocação de Recursos.....	96
6.4	Considerações finais	97
7	Conclusão e Trabalhos Futuros	98
7.1	Principais Contribuições do SUTIL	98
7.2	Principais Contribuições do ARPRAP	100
7.3	Trabalhos Futuros	101
8	Referências Bibliográficas	102

1 Introdução

O avanço tecnológico das redes sem fio, a recente proliferação de dispositivos portáteis, tais como laptops, PDAs e palmtops, aliados ao aumento de popularidade da computação móvel, tornaram realidade o cenário de sistemas ubíquos vislumbrado por Mark Weiser [WEISER 1991]. Sistemas ubíquos caracterizam-se como ambientes nos quais dispositivos (estações), agentes de software e serviços são integrados de forma transparente e não invasiva e cooperam para dar suporte a atividades dos seres humanos, a qualquer hora e em qualquer lugar. Em geral, tais sistemas incluem ambientes instrumentados com sensores e interfaces homem-máquina dotadas de mecanismos de personalização.

Para viabilizar a concretização de sistemas ubíquos, é necessário que as estações (móveis ou não) situadas em qualquer localidade e rodando as aplicações dos usuários estejam permanentemente conectadas a redes sem fio de diferentes tecnologias. A integração de forma transparente das diversas tecnologias de redes sem fio, em particular das redes WiFi, Bluetooth, GSM 3G, WiMAX e satélite, com as redes cabeadas, bem como a criação de um ambiente unificado de redes e serviços, caracterizam as Redes de Próxima Geração (NGN – *Next Generation Networks*). Portanto, as Redes de Próxima Geração consistem num importante componente para viabilizar o estabelecimento dos sistemas ubíquos [ITU-T].

As Redes da Próxima Geração – NGN (Next Generation Networks) representam uma nova estrutura na qual os dados, a voz e as novas aplicações multimídia convergem. Uma Rede de Próxima Geração (NGN) é baseada em pacotes e capaz de prover serviços, inclusive de telecomunicações, fazendo uso de múltiplas tecnologias de transporte de banda larga com provisão de QoS, com acesso irrestrito pelos usuários a diferentes provedores de serviços e com mobilidade generalizada que permitirá uma provisão consistente e ubíqua de serviços para os usuários. Além disso, em contraste com as redes de telecomunicações atualmente existentes, que suportam serviços de voz e dados utilizando para isso plataformas independentes, as redes NGNs tornam possível a redução de custos de operação e manutenção de toda a rede de telecomunicações.

Em suma, as redes de próxima geração substituirão várias redes isoladas (como rede telefônica fixa, rede telefônica móvel, rede de pacotes X.25, rede de dados ATM, etc.) por uma única rede de núcleo (*core network*) que permite interligar redes de diferentes tecnologias.

Nesse contexto, observa-se portanto que as redes sem fio vêm evoluindo e se adaptando aos vários desafios impostos pelas novas gerações de redes. Com relação às chamadas “gerações” das tecnologias sem fio, consideram-se como sistemas de primeira geração (1G) os serviços celulares inicialmente oferecidos com tecnologia de rádio analógica. Quando esta tecnologia foi substituída pela digital na década de 90, surgiram então as redes de segunda geração (2G). Com o passar do tempo esta classificação em gerações deixou de ser exclusiva para redes celulares e a *International Telecommunication Union* (ITU) passou a classificar as demais tecnologias sem fio. No ano 2000, a ITU finalizou o padrão de terceira geração (3G) sob o nome de IMT-2000. Atualmente esta terceira geração está difundida pelo mundo e oferece tecnologias para acesso à Internet de banda larga tais como serviços de hot spot ou ad hoc, redes de acesso multi-hop, com nível de serviço compatível com os cabeados para usuários de Internet móvel. As redes WiFi (IEEE 802.11n), por exemplo, estão classificadas como 3G e as redes WiMAX (IEEE 802.16e), apesar de já serem baseadas em OFDMA e em arquitetura inteiramente IP (requisitos para 4G), em 2008 foram classificadas como 3.9G pelo ITU. O ITU, sob o nome de IMT-Advanced, está trabalhando nas futuras capacidades das redes móveis de banda larga de quarta geração (4G).

Uma das tecnologias de rede sem fio que primeiro se destacou e popularizou foi a tecnologia de redes locais sem fio, aderentes ao padrão 802.11 (WiFi). Porém, apesar do sucesso das redes sem fio baseadas nas tecnologias da família IEEE 802.11, estas sofrem de duas principais limitações: cobertura limitada e escalabilidade restrita [CHANG 2005]. Por outro lado, as redes celulares de terceira geração (3G) fornecem grande cobertura, porém com limitações na largura de banda e com altos custos de infra-estrutura [CHANG 2005]. A introdução da tecnologia WiMAX, baseada no padrão IEEE 802.16, veio aliar flexibilidade, comum nas redes sem fio, com a capacidade de realizar transmissões que atendam os requisitos das atuais aplicações multimídia, tais como maior largura de banda, restrições de atraso e atendimento de um grande número de usuários. Adicionalmente, a especificação 802.16e [IEEE Std 802.16e - 2005] estende o padrão 802.16, incluindo a capacidade de suportar usuários móveis, tornando o WiMAX uma poderosa solução para conectividade sem fio em regiões metropolitanas. E desde 2007 já podem ser vistas as primeiras implementações práticas do padrão IEEE 802.16e surgindo para fornecer novos serviços de acesso banda larga.

Entretanto, apesar das inúmeras vantagens de se utilizar redes sem fio, como o baixo custo de instalação, a facilidade de configuração e a possibilidade de mobilidade dos

dispositivos, existem desafios que ainda precisam ser suplantados para alavancar o amplo uso dessas redes.

1.1 Motivação

Uma rede sem fio inclui uma estação base (BS), que serve uma área geográfica conhecida como célula, e uma ou mais estações clientes. Cada estação base é o nó central de uma célula, e coordena as comunicações entre todas as estações clientes localizadas nesta célula. Em geral, uma rede sem fio é composta de diferentes células, e tem várias estações base conectadas umas as outras por um *backbone*, provendo conectividade para até milhares de usuários, fixos ou móveis. Quando um usuário móvel se afasta de sua estação base, e se aproxima de uma outra, ele finaliza sua conexão com a BS mais distante e estabelece uma nova conexão com a BS mais próxima. A transição de uma estação móvel (MS) de uma BS para outra é chamada de processo de *handoff* ou *handover*. Idealmente, o processo de *handover* deve: (i) ser completamente transparente para os usuários móveis; e (ii) garantir que todos os serviços multimídia não serão interrompidos durante o processo.

Adicionalmente, o processo de transição da estação pode envolver somente redes de mesma tecnologia (chamado de *handover* horizontal) ou redes de tecnologias diferentes, por exemplo, de uma rede IEEE 802.16 para uma rede IEEE 802.11 (chamado de *handover* vertical).

No contexto das redes NGN, um dos desafios envolvidos, “estar sempre conectado à melhor rede” (ABC – *Always Best Connected*) [GUSTAFSSON 2003] requer o desenvolvimento de soluções que considerem vários critérios para a tomada de decisão de qual seria essa “melhor rede” quando da necessidade de efetuar a transição de usuários móveis entre diferentes células e operadoras de redes. Exemplos desses critérios são: prioridade de serviço; preferência e perfil do usuário; contexto da aplicação; força do sinal; requisitos da rede; requisitos do dispositivo; e políticas corporativas ou da operadora.

Além de considerar esses vários critérios na tomada de decisão quanto à “melhor rede”, é importante considerar o gasto de energia no processo de transição, por parte das estações móveis sem fio, uma vez que os mesmos são tipicamente movidos a bateria. Portanto, um requisito básico em qualquer processo de decisão para escolher a melhor rede é considerar redes que demandem um menor gasto de energia para as estações a ela conectadas. Após a seleção da rede, tem início o processo de transição das estações de uma rede sem fio (célula) para outra. No que diz respeito ao mecanismo de seleção da(s) “melhor(es)” rede(s) ,

este torna-se complexo devido a dois fatores: (i) possibilidade de coexistência de diversas tecnologias de acesso heterogêneas; e (ii) necessidade de balancear os requisitos dos usuários com a carga na rede, resultante do processo de handover de um conjunto de usuários móveis.

Fazem-se necessárias, portanto, propostas de mecanismos capazes de selecionar a(s) melhor(es) rede(s) em presença de handover como forma de solucionar o problema apontado.

Outro grande desafio ligado a mobilidade de estações é conseguir uma utilização eficiente de redes homogêneas, provendo serviços com qualidade para seus usuários, principalmente em presença de *handover*. Tal desafio é gerado a partir das próprias limitações do meio sem fio, resultando em flutuações na disponibilidade de recursos. Adicionalmente, com o surgimento de novas aplicações, cada vez mais exigentes em termos de largura de banda, é importante averiguar a influência da mobilidade dos usuários sobre a QoS em uma rede sem fio, de forma a garantir níveis de QoS compatíveis com cada aplicação.

Uma forma de solucionar o problema apontado por esse desafio é propor novos algoritmos de *handover* ou melhorar os processos de *handover* de forma a torná-los mais eficientes, ou seja, rápidos, nos padrões existentes para cada tecnologia e ao mesmo tempo garantir que a qualidade de serviço para os usuários móveis não se deteriore em virtude da transição (*handover*).

Uma das tecnologias mais promissoras em termos de redes de computadores sem fio, alinhada com o conceito de Redes de próxima Geração (NGN) e de quarta geração de redes sem fio (4G) é a rede WIMAX baseada no padrão IEEE 802.16. As principais características as quais tornam as redes WiMAX uma tecnologia de destaque frente a outras tecnologias de redes sem fio são: (i) facilidade e rapidez de instalação da rede; (ii) possibilidade de atingir regiões nas quais não existe infra-estrutura de banda larga com fio; (iii) garantia de QoS para as atuais aplicações multimídias, tais como maior largura de banda, restrições de atraso e atendimento de um grande número de usuários; (iv) área de cobertura abrangente; (v) baixo custo de manutenção; (vi) baixo custo de infra-estrutura (equipamentos de rede são de custo mais baixos se comparados com os da redes celulares); e (vii) suporte a usuários móveis [DELICATO 2008]. Outras melhorias que merecem destaque são: portabilidade, qualidade de serviço, protocolos de autenticação e criptografia de dados e aspectos de economia de energia para os dispositivos da rede. Por estes motivos a rede WIMAX foi escolhida para ambientar o desafio de redes homogêneas neste trabalho.

1.2 Objetivos

O objetivo geral deste trabalho é melhorar os processos relacionados ao serviço de gerência de handover, para maximizar a eficiência da rede e prover melhor qualidade de serviço em presença de handover, assim contribuindo para a viabilização de sistemas ubíquos. Os objetivos específicos estão relacionados aos dois desafios supracitados, ou seja: estar sempre conectado à melhor rede e conseguir uma utilização eficiente de redes homogêneas, provendo serviços com qualidade para seus usuários, principalmente em presença de *handover*, o presente trabalho apresenta duas propostas. O mecanismo SUTIL (Seleção de Redes baseada em Função Utilidade e Programação Linear Inteira) e o algoritmo ARPRAP (Reserva Próativa de Recursos Assistida por Perfil). O **SUTIL** é um mecanismo capaz de escolher a(s) melhor(es) rede(s) em presença de handover e o **ARPRAP** é um algoritmo de Reserva Próativa de Recursos Assistida por Perfil capaz de solucionar o problema de manter a mesma qualidade de serviço para os usuários móveis quando eles se deslocam entre redes homogêneas, no caso redes WiMAX.

1.2.1 Primeiro objetivo específico

O primeiro objetivo deste trabalho é apresentar um Mecanismo de Seleção de Redes em ambientes NGN baseado em Função Utilidade e Programação Linear Inteira, denominado SUTIL. O SUTIL busca escolher a(s) melhor(es) rede(s) em presença de *handover*. A solução provida pelo SUTIL é global no sentido em que, embora considere os requisitos e interesses individuais de cada usuário, considera também a carga imposta na rede resultante de um processo de *handover* de um conjunto de usuários móveis conectados a uma mesma BS, e candidatos a, simultaneamente, participar desse *handover*.

Assim, o SUTIL é um mecanismo de seleção de redes, cujas principais metas são: (i) selecionar redes em um contexto global, ou seja, buscar a melhor solução para o conjunto de MSs simultaneamente candidatas a handover numa mesma BS e o conjunto de redes candidatas; (ii) maximizar a satisfação dos usuários localizados em uma mesma BS e candidatos ao *handover*, atendendo os requisitos dos seus serviços em uso; (iii) garantir a eficiência de uso dos recursos das redes; e (iv) minimizar o gasto de energia do dispositivo ao efetuar um handover.

Assim, tal mecanismo busca garantir que haja uma distribuição dos recursos (balanceamento de carga) solicitados pelos serviços entre as redes selecionadas e, ao mesmo tempo, garantir que os requisitos desses serviços sejam atendidos por tais redes.

O SUTIL considera vários critérios para a tomada de decisão. Esses critérios dizem respeito ao usuário, a rede e ao dispositivo, tendo como exemplo, **prioridade de serviço, preferência de energia do usuário e energia máxima**. Usando-se destes critérios o SUTIL mapeia as preferências (utilidades) dos usuários em relação as redes candidatas e procura uma solução global que maximize estas utilidades com auxílio de Programação Linear Inteira.

O mecanismo SUTIL foi criado para ser executado em cada BS em um sistema NGN, significando que existe uma instância do SUTIL por BS. O mecanismo responsável pela interação entre diferentes instâncias do SUTIL está fora do escopo deste trabalho. Adicionalmente, é importante enfatizar que o SUTIL desempenha especificamente a tarefa de seleção de redes. Depois da melhor rede ser selecionada pelo SUTIL, o próximo desafio é determinar o momento certo para disparar o processo de handover. Este passo está fora do escopo deste trabalho, mas existem vários trabalhos com este propósito, como [MOHANTY 2006].

Para o SUTIL operar corretamente, assume-se que as BSs que são candidatas ao processo de *handover* podem cooperar entre si de forma que possam ser trocadas informações de alto nível, como, por exemplo, as capacidades das BSs. Esta hipótese também é adotada em outros trabalhos, como no trabalho apresentado em [CHALMERS 2006]. De acordo com essa hipótese, o SUTIL não funcionará quando um dos provedores de rede envolvidos em um processo de handover não estiver disposto a colaborar com essa troca informações. Este tipo de cooperação é uma tendência inevitável no desenvolvimento das redes NGN. Já existem iniciativas em andamento para propiciar esse tipo de cooperação. Por exemplo, em [AKYILDIZ 2005], os autores propõem uma arquitetura para a comunicação móvel e ubíqua (*Architecture for Ubiquitous Mobile Communications - AMC*). O AMC é responsável por integrar diferentes tecnologias de rede sem fio (GPRS, cdma2000, UMTS, WLAN, etc.) usando duas entidades: Network Inter-operating Agent (NIA) e um Interworking Gateway (IG). O NIA reside na Internet e atua como um “terceiro” elemento eliminando a necessidade de SLAs diretos entre diferentes operadoras e, deste modo, facilitando a cooperação. Argumenta-se que é provável que a existência de mecanismos como o SUTIL, que realizam a seleção de redes no contexto NGN e consideram os interesses tanto das redes como dos usuários/aplicações, irá motivar ainda mais os operadores a estabelecer acordos e atualizar seus sistemas para incorporá-los.

1.2.2 Segundo objetivo específico

O segundo objetivo deste trabalho é apresentar um algoritmo de Reserva Próativa de Recursos Assistida por Perfil, denominado ARPRAP. O ARPRAP busca solucionar o problema de manter a mesma qualidade de serviço para os usuários móveis quando eles se deslocam entre redes homogêneas, no caso redes WiMAX. Nessa linha de investigação do presente trabalho explora-se o perfil do comportamento do usuário em movimento para maximizar a eficiência da rede e prover melhor QoS para diferentes classes de usuários em um cenário que integra redes WiMAX.

Assim, o ARPRAP é um algoritmo que faz uso do perfil do comportamento do usuário em tempo real para: (i) refinar o conjunto de células previstas para o próximo *handover*, levando em conta não apenas o tipo de usuário envolvido (Premium ou Normal) como também o contexto do serviço e o da rede; e (ii) fazer uma reserva pró-ativa de recursos para a célula escolhida, pelo processo de refinamento, para o próximo *handover*.

O algoritmo ARPRAP é responsável por efetuar uma reserva pró-ativa de canais para a célula de destino escolhida, fato que só é possível quando o comportamento do usuário móvel em relação à utilização dos serviços durante um determinado trajeto for conhecido. Tal comportamento é definido em um perfil do usuário. Assim, graças ao conhecimento do perfil de comportamento do usuário móvel, é possível acionar de forma pró-ativa o processo de geração de requisições de alocação de canal de uma célula destino, uma vez que um futuro procedimento de *handover* "conhecido" necessitará que tal processo seja efetuado.

Se não houver canal disponível na célula destino para um serviço com um dado conjunto de requisitos, a ligação será terminada, o que caracteriza uma falha no *handover*. A probabilidade de ocorrência dessa falha de *handover* é um dos parâmetros que medem a QoS da rede sem fio. Para minimizar essa probabilidade de ocorrência dessa falha de *handover* é aprimorado o algoritmo *Fast handover* da camada 2 do IEEE 802.16 [IEEE Std 802.16e - 2005], refinando-se o processo de escolha das BS vizinhas candidatas ao *handover* com o auxílio do perfil de comportamento do usuário móvel, isto é, considerando aspectos da trajetória, requisitos da aplicação e de rede. O ARPRAP usa três novas estruturas de dados referentes ao usuário: Tabela de Perfil, Tabela de Trajetos e Tabela de Recursos. A **Tabela de Perfil** do usuário contém informações de perfil específicas de usuários em movimento e que estão registrados em sua rede/célula de origem (*home*) tais como: identidade do usuário, tipo do usuário (**Premium**, **Normal**), serviços acordados, ponteiro para a tabela de trajetos, ponteiro para a tabela de recursos. A Tabela de Perfil é uma estrutura armazenada na *Internet*,

ficando acessível de qualquer ponto da rede. A *Tabela de Trajetos* contém uma lista de trajetos usados por um período do dia (início da manhã, manhã, almoço, início da tarde, tarde, noite) por um usuário móvel e a sua respectiva frequência de uso. Um trajeto individual, por exemplo, $\langle a_1, a_2, \dots, a_n \rangle$, é composto por uma lista de identificadores “ a_n ”, onde cada identificador “ a_n ”, denota uma área quadrada de lado igual a um inteiro “ L ”. Cada área é composta por várias células sobrepostas. Alguns trajetos são definidos através de acordo (SLA) entre o usuário e o provedor do serviço informando que o usuário é cliente *Premium*. Outros trajetos são inseridos na lista conforme o mesmo passa a ser utilizado pelo usuário.

Já a *Tabela de Recursos* contém uma lista de recursos, tais como a banda, usados por trajeto e por período por um usuário móvel e a respectiva frequência quanto ao uso do recurso. Alguns recursos são definidos através de um acordo (SLA) entre o usuário e o provedor do serviço informando que o usuário é cliente *Premium*. Outros recursos são inseridos na lista conforme o mesmo passa a ser utilizado pelo usuário.

Uma vez efetuada a escolha da BS candidata ao *handover*, são divulgados previamente os fluxos de serviços, para, a partir de uma dada distância da BS candidata ao *handover*, tais fluxos serem de fato alocados, concretizando, portanto, a operação do *Fast handover*.

1.3 Organização

O restante deste trabalho está estruturado em seis capítulos. No segundo capítulo, são apresentados os conceitos básicos para a compreensão deste trabalho. Serão descritas as particularidades (i) do tipo de rede sobre a qual a arquitetura do ARPRAP proposta pode ser aplicada, (ii) das redes NGN onde está contextualizado o mecanismo SUTIL e os trabalhos relacionados, com os diversos mecanismos e algoritmos propostos para estas redes existentes na literatura. O terceiro e o quarto capítulos apresentam respectivamente a descrição detalhada do mecanismo proposto SUTIL e do algoritmo proposto ARPRAP, descrevendo a arquitetura e metodologia desenvolvida, com suas etapas de atividades. Nos capítulos seguintes, quinto e sexto respectivamente, são detalhadas as implementações, simulações executadas, os resultados e as análises dos resultados do uso do SUTIL e do ARPRAP. Por fim, as conclusões e os trabalhos futuros são apresentados no sétimo capítulo.

2 Conceitos Básicos

Este capítulo apresenta os conceitos básicos necessários para a fundamentação do presente trabalho. A organização deste capítulo segue uma lógica que parte do tipo de rede sobre a qual a arquitetura do ARPRAP proposta pode ser aplicada, do contexto de redes NGN para o qual o SUTIL foi concebido, passando pelos princípios teóricos em que ambos se baseiam. Assim, a Seção 2.1, introduz os conceitos gerais sobre Redes sem fio Metropolitanas e redes sem fio do padrão IEEE 802, a Seção 2.2 descreve os objetivos e as principais características do padrão IEEE 802.16, exemplificando, dessa forma, um dos cenários de aplicação da presente proposta. A Seção 2.3 traz os principais conceitos de QoS utilizados em redes metropolitanas sem fio. A Seção 2.4 resume os principais conceitos sobre o padrão IEEE 802.16e e os tipos de *handover*, pois o ARPRAP e SUTIL são mecanismos que atuam na melhoria do *handover* e da seleção de redes candidatas a *handover* respectivamente. A Seção 2.5 traz os principais conceitos das Redes de Nova Geração (NGN) onde se aplica o SUTIL. A Seção 2.6 traz os trabalhos relacionados do ARPRAP e do SUTIL. Por fim, a Seção 2.7 finaliza este capítulo, tecendo algumas considerações finais.

2.1 Padrões IEEE 802 para redes sem fio

O avanço recente das tecnologias de redes sem fio fez surgir um cenário no qual usuários acessam a Internet a partir de dispositivos portáteis, a qualquer hora e em qualquer lugar, tornando a computação ubíqua uma realidade. Tais usuários fazem uso da rede para executar diversos tipos de serviços e aplicações, incluindo aplicações multimídia de tempo real, que demandam altos requisitos de largura banda e de qualidade (QoS). Para atender a esses usuários, as redes sem fio devem não só prover acesso banda larga a Internet, mas também fornecer suporte a várias classes de serviços. Os serviços demandados podem depender de informações de contexto como localização e preferências do usuário, características do dispositivo e políticas de segurança.

Entretanto, apesar das vantagens de se utilizar redes sem fio, como o baixo custo de instalação, a facilidade de configuração e a possibilidade de mobilidade dos dispositivos, existem desafios que ainda precisam ser suplantados para alavancar o amplo uso dessas redes. Para prover soluções sistemáticas para esses desafios, bem como para viabilizar a interoperabilidade entre equipamentos de diversos fabricantes, surgiram padrões como os do [IEEE Std 802.11 - 1999] e do [IEEE Std 802.16 - 2004].

O órgão de padronização IEEE (*Institute of Electrical and Electronics Engineers*) foi

responsável pelo desenvolvimento de um conjunto de padrões para atender os diferentes tipos de redes sem fio. A Figura 1 ilustra os diferentes padrões IEEE de redes sem fio e a relação de áreas de coberturas das redes.

As *redes sem fio do tipo pessoal* (*Personal Area Network – PAN*) e as redes sem fio geograficamente distribuídas (*Wireless Wide Area Network – WWAN*) são tratadas respectivamente pelos padrões IEEE 802.15 e IEEE 802.20. O padrão IEEE 802.11, que normatiza as *redes locais sem fio* (*Wireless Local Area Network – WLAN*), foi desenvolvido para atender as necessidades de se dispor de uma rede local sem fio para vários usuários. O uso crescente desse tipo de rede seja em locais públicos, residências ou em empresas deve-se principalmente ao fato das mesmas não necessitarem de um cabeamento estruturado e, portanto, terem seus custos de implantação e manutenção reduzidos quando comparados à estrutura cabeada. Entretanto, o padrão 802.11 possui algumas limitações e vulnerabilidades, como o raio de alcance de seus pontos de acesso, a qualidade de serviço oferecida na transmissão de dados e as vulnerabilidades relativas a segurança. Novos adendos ao padrão IEEE 802.11 surgiram com o intuito de solucionar algumas dessas vulnerabilidades. O termo Wi-Fi refere-se a um padrão de referência adotado para certificar equipamentos considerados conformes com o padrão IEEE 802.11 . O Wi-Fi Forum, entidade sem fins lucrativos responsável pela certificação, seleciona um conjunto de características técnicas previstas pelo padrão IEEE 802.11 e confere o título de certificação aos equipamentos que, submetidos a testes específicos, comprovam atender a todo esse conjunto de características. Desse modo, o Wi-Fi Forum busca garantir a interoperabilidade entre os dispositivos IEEE 802.11 produzidos pelos diversos fabricantes existentes.

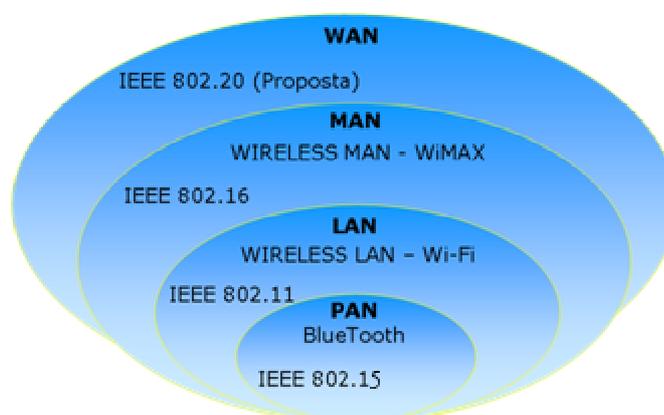


Figura 1. Padronização IEEE para redes sem fio

O padrão IEEE 802.16 trata das *redes metropolitanas sem fio* (Wireless Metropolitan Area Network - WMAN). Na primeira versão do padrão IEEE 802.16, publicada em 2001, foi especificada uma rede ponto-multiponto (PMP), operando necessariamente com enlaces de rádio com visada direta (line-of-sight - LOS) e empregando, entre outros, multiplexação por OFDM (Orthogonal frequency-division multiplexing) no espectro de 10 a 66 GHz. A taxa máxima teórica chegava a 134 Mbps e a distancia máxima a 50 km. Nesse padrão só eram permitidos nós fixos.

O padrão 802.16a, publicado em 2003, introduziu os principais procedimentos para a operação de redes Mesh. Nesse adendo, as redes passavam a poder operar sem visada direta (non-line-of-sight - NLOS) em frequências de 2 a 11 GHz. A taxa máxima teórica chegava aos 75 Mbps e o alcance máximo a 10 Km [ZHANG 2007].

No adendo 802.16-2004, homologado em 2004, encontra-se uma revisão dos adendos 802.16a, 802.16b e 802.16c. Nesse adendo são padronizados, entre outros pontos, a provisão de acesso fixo e nomádico, em ambientes com linha de visada direta (LOS) ou sem linha de visada direta (NLOS), como também a operação em frequências licenciadas e não licenciadas.

No adendo 802.16e, aprovado em 2005, é abordado tanto acesso fixo como acesso móvel, onde se encontram especificados o suporte ao processo de *handover*, modos de economia de energia e novas versões de protocolos de segurança. Um aperfeiçoamento do modo de operação Mesh também foi realizado nesse adendo. A taxa máxima teórica pode chegar a 15 Mbps por canal de 5 MHz de largura de banda [ZHANG 2007] e [AHSON 2007]. A Tabela 1 resume a evolução do padrão 802.16.

Tabela 1. Características dos padrões 802.16e

Aspecto	802.16e
Acesso fixo	Sim
Portabilidade	Sim
<i>Handover</i>	Sim
Taxa de transmissão	Até 15 Mbps
Largura do canal	Entre 1,25 e 20 MHz
Área de cobertura	2 – 5 Km

Espectro	2-6 GHz
Canal de Propagação	LOS e NLOS
Raio de célula	Típico 1,5 a 5 Km

Da mesma forma que foi mencionado para o termo WiFi, o termo WiMAX refere-se a um padrão de referência adotado para certificar equipamentos considerados conformes com o padrão IEEE 802.16. O WiMAX Forum, entidade sem fins lucrativos responsável pela certificação, seleciona um conjunto de características técnicas previstas pelo padrão IEEE 802.16 e confere o título de certificação aos equipamentos que, submetidos a testes específicos, comprovam atender a todo esse conjunto de características. Desse modo, o WiMAX Forum busca garantir a interoperabilidade entre os dispositivos IEEE 802.16) produzidos pelos diversos fabricantes existentes.

2.2 O Padrão IEEE 802.16 – WiMAX

Nas redes WiMAX, a forma de transmissão do sinal é similar à usada nas redes de telefonia celular. Uma torre, a estação base (BS - Base Station), fornece sinal diretamente para as estações assinantes (SS - Subscriber Station), possibilitando a cobertura da rede nas áreas programadas. As torres secundárias funcionam como replicadores para aumentar o alcance do sinal.

Nos próximos subitens são apresentados a arquitetura de comunicação, os elementos que compõem a rede (estações base e estações subscritoras) e as topologias de comunicação suportadas, bem como as bandas de frequências especificadas para uso pelas redes WiMAX.

2.2.1 Arquitetura

No padrão IEEE 802.16 são especificadas a camada física (PHY) e a camada de acesso ao meio (MAC) para as redes metropolitanas de banda larga sem fio (BWA). A subcamada MAC encontra-se dividida em três subcamadas: a subcamada de Convergência de Serviços (Convergence Sub-layer - CS), a subcamada de Parte Comum (Common Part Sub-layer - CPS) e a subcamada de Privacidade (Privacy Sub-layer - PS). A Figura 2 ilustra essa divisão de camadas, seus pontos de acesso e ainda apresenta as principais funções das subcamadas.

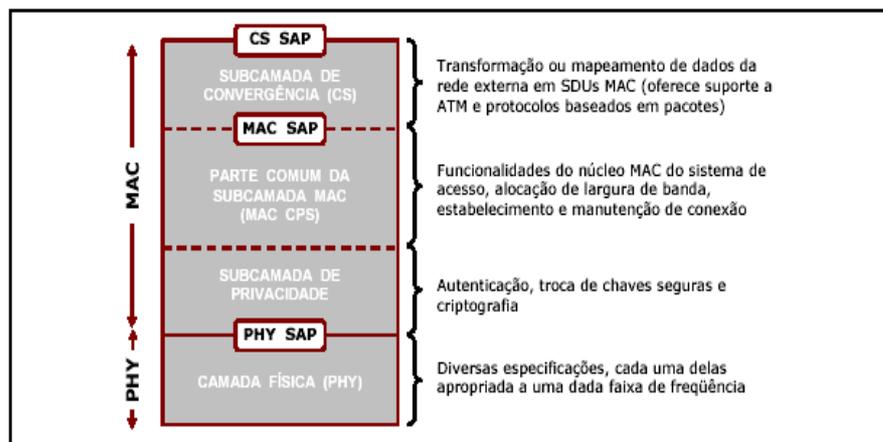


Figura 2. Divisão em camadas, suas funções e pontos de acesso (SAPs)

Na camada física são especificados o espectro de frequência, o esquema de modulação, as técnicas de correção de erros, a sincronização entre transmissor e receptor, a taxa de dados e a estrutura de multiplexação. Por sua vez, na subcamada de acesso ao meio (MAC) são especificadas as funções associadas aos serviços oferecidos aos usuários, que incluem a transmissão de dados em quadros e o controle do acesso ao meio sem fio compartilhado. Na camada MAC é definido como e quando a estação base ou os assinantes podem iniciar a transmissão no canal.

2.2.2 Topologias de Rede

O padrão IEEE 802.16 define dois tipos de topologias: ponto-multi-ponto (PMP) e Mesh. No modo PMP, a estação base tem a função de coordenar e retransmitir todo o tráfego entre as estações assinantes da rede. Toda e qualquer estação assinante tem a obrigatoriedade de comunicar-se com a BS para transmitir qualquer tipo de dado para outra SS. No modo Mesh, cada SS pode comunicar-se diretamente com as SS vizinhas sem qualquer coordenação com a BS. Vale ressaltar que em instalações Mesh típicas, têm-se ainda alguns nós que farão o papel de BS, conectando a rede Mesh ao link de backhaul (Internet, telecomunicações, etc.).

Para o suporte a mobilidade, o padrão IEEE 802.16e especificou novas funcionalidades aos elementos de rede já existentes no padrão IEEE 802.16 e adicionou novos elementos. Esses novos elementos e funcionalidades são sintetizados a seguir.

- **Estação móvel (MS)** – Uma estação móvel é uma estação assinante (SS) com intenção de ser utilizada durante o movimento ou não, em pontos não específicos, escolhidos aleatoriamente.

- **Estação Base (BS) vizinha** – Para qualquer estação móvel (MS), uma BS vizinha é aquela cujo canal de transmissão de download pode ser recebido pela MS.
- **Serving BS** – Para qualquer estação móvel, a Serving BS é aquela com que mais recentemente a MS realizou seu registro durante a entrada na rede ou durante o processo de *handover*.
- **BS Alvo** – A BS Alvo é aquela com que uma MS pretende registrar-se ao final do processo de *handover*.
- **BS Ativa** – Uma BS ativa é informada das capacidades das MS, parâmetros de segurança, fluxos de serviço e informações de contexto MAC completas.
- **Conjunto ativo** – O Conjunto ativo contém a lista das BS ativas para uma MS. O Conjunto ativo é gerenciado pela MS e pela BS, e é utilizado para o *Fast BS Switching* (FBSS) e *Macro Diversity Handover* (MDHO).
- **BS Âncora** – A BS Âncora é a BS onde as MSs estão registradas, sincronizadas e monitorando o canal de downlink para obter informações de controle, permitindo suporte ao processo de *handover* dos tipos FBSS e MDHO para as MSs.
- **Atribuição de frequência (FA)** – FA indica a atribuição lógica da frequência central do canal downlink (DL) e a largura de banda programada para a BS.
- **Handover (HO)** – Processo no qual uma MS migra de uma área coberta por uma BS para uma área coberta por outra BS. São definidas duas variantes:
 - *break-before-make HO*: HO no qual o serviço com a BS Alvo inicia após a desconexão com a Serving BS anterior.
 - *make-before-break HO*: HO no qual o serviço com a BS Alvo inicia antes da desconexão com a Serving BS anterior.
- **Intervalo de varredura** – Período de tempo para a MS monitorar as BSs vizinhas para determinar as prováveis BS Alvo para *handover*.
- **Backbone de rede** – Infra-estrutura onde duas ou mais BSs comunicam-se entre si e pode incluir comunicação com outras redes.
- **Procedimento de power down** - Procedimento que uma MS executa quando está sendo desligada.

2.3 QoS

O mecanismo mais importante de fornecimento de QoS nas redes WiMAX está associado à transferência de quadros da subcamada MAC, usando o conceito de fluxo de serviço. Um fluxo de serviço, identificado unicamente por um identificador de fluxo de serviço (SFID - Service Flow Identifier), é um fluxo unidirecional de quadros, caracterizado por um conjunto de parâmetros de QoS tais como: latência, retardo e garantias de vazão. Graças a esses parâmetros de QoS, é possível escalonar os fluxos de serviço na rede WiMAX. Um fluxo de serviço, por sua vez, está associado a uma única conexão. Uma conexão é identificada de forma única por um identificador de conexão (CID - Connection Identifier). Na camada MAC do WiMAX móvel, o QoS é fornecido através dos fluxos do serviço como ilustrado na Figura 3.

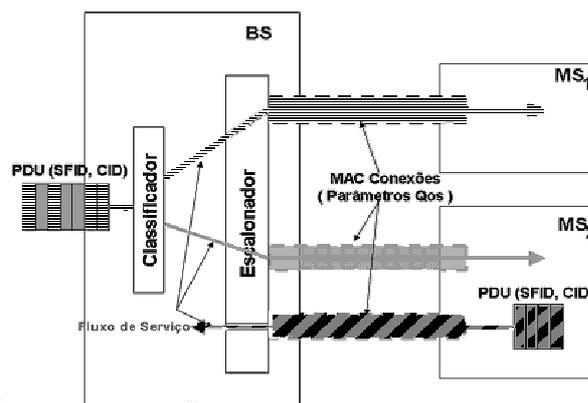


Figura 3. Suporte a QoS em Redes Móveis WiMAX

Em outras palavras, os fluxos de serviço provêm um mecanismo para gerenciamento de QoS no sentido uplink e downlink, além de serem parte integrante do processo de alocação de banda. Uma SS requisita banda no uplink baseada na conexão, implicitamente identificando o fluxo de serviço, já que para cada CID só existe um SFID correspondente. A banda então é alocada pela BS para a SS como um agregado das solicitações por conexão feitas pela SS. Os parâmetros de QoS associados com o fluxo de serviço definem os requisitos da transmissão que será requisitada e assim a banda para atender esses requisitos. Na Figura 3 observa-se um fluxo de serviço (cinza) destinado a uma conexão com a MS2 e o fluxo (listas horizontais) para a MS1 ambos transmitidos no canal de downlink (BS→MS). A BS separa as PDUs destinadas a MS diferentes usando um classificador. Para tal, é consultado o ID do

fluxo de serviço (SFID) e o ID da conexão (CID) de cada PDU. Após o procedimento de separação ser realizado, cada PDU passa por um escalonador, o qual definirá a sua ordem de atendimento de forma a garantir a QoS requerida. Já o fluxo de serviço (listas em diagonal), acontece no sentido uplink (MS→BS). Portanto, com a utilização de QoS orientado a conexão, torna-se possível fornecer o controle exato sobre o tráfego do link aéreo, principal gargalo de saída da rede.

Tabela 2. Aplicações de Redes Móveis WiMAX e Qualidade de Serviço

Categoria QoS	Aplicações	Especificações
UGS (Unsolicited Grant Service)	VoIP	Taxa máxima, Latência máxima e Tolerância a Jitter
rtPS (real-time Polling Service)	Streaming Audio ou Video	Prioridade de Tráfego, Taxa mínima e Latência máxima
ErtPS (Extended real-time Polling Service)	Voz com detecção de uso (VoIP com DTX)	Taxa máxima, Latência máxima, Tolerância a Jitter e Prioridade de Tráfego
nrtPS (non real-time Polling Service)	Protocolo de Transferência de Arquivo (FTP)	Taxa mínima garantida e Prioridade de Tráfego
BE (Best Effort)	Transferência de dados, Navegação Web	Prioridade de Tráfego e Taxa máxima

Os parâmetros do fluxo de serviço podem ser dinamicamente gerenciados através de mensagens MAC para acomodar serviços sob demanda e prover QoS para conexões nos sentidos downlink e uplink. A WiMAX móvel suporta uma larga escala de serviços de dados e aplicações com exigências variadas de QoS (Tabela 2). Na Tabela 3 são apresentadas sucintamente as 5 classes de serviço, detalhadas na seção 2.4.1, que mapeiam os tipos possíveis de aplicações e suas necessidades de QoS específicas.

Tabela 3. Classes de aplicações 802.16e

Classe	Aplicação	Requisito estimado de banda		Requisito estimado de latência		Requisito estimado de jitter	
1	Jogos interativos multiplayer	Baixo	50 kbps	Baixo	< 25 ms	N/D	
2	VoIP&videoconferência	Baixo	32 a 64 kbps	Baixo	< 160 ms	Baixo	< 50 ms
3	Streaming	Baixo a alto	5 kbps a 2 Mbps	N/D		Baixo	< 100 ms
4	Web&IM (Instant Messaging)	Moderado	10 kbps a 2 Mbps	N/D		N/D	
5	Downloads de mídia	Alto	> 2 Mbps	N/D		N/D	

2.3.1 Classe de Serviço

O padrão IEEE 802.16 define quatro classes de serviço as quais devem ser tratadas de forma diferenciada pelo mecanismo de escalonamento da camada MAC. Essas classes são: Unsolicited Grant Service (UGS), Real-Time Polling Service (rtPS), Non Real-Time Polling Service (nrtPS) e Best Effort (BE). Além das classes existentes no padrão [IEEE Std 802.16 - 2004] foi especificado no padrão [IEEE Std 802.16e - 2005] uma nova classe, chamada Extended rtPS (ertPS), fundamentada na eficiência do UGS e rtPS. Segundo esse mecanismo, a estação base (BS) pode conceder banda não solicitada como no UGS, porém o tamanho da banda alocada é variável. As classes de serviços são descritas com mais detalhes em seguida.

- **Unsolicited Grant Service (UGS)**

Projetada para oferecer suporte aos fluxos de serviço em tempo real que geram pacotes de dados de tamanho fixo em intervalos periódicos. Essa categoria de serviço pode ser representada pelo tráfego gerado por emulação T1/E1 e por aplicações de voz sobre IP sem supressão de silêncio. O serviço oferece periodicamente concessões não solicitadas para a transmissão dos dados, eliminando a sobrecarga e a latência ocasionada pelas requisições de banda unicast das estações para enviar pedidos de transmissão. No UGS, a SS é proibida de usar qualquer requisição de contenção e a BS não oferece qualquer oportunidade de requisição unidirecional para a SS. Requisições de piggyback (pedidos de banda) não são permitidas no UGS. Os principais parâmetros para

essa categoria são: Taxa máxima (Maximum Sustained Rate), Latência máxima (Maximum Latency Tolerance) e Tolerância a Jitter (Jitter Tolerance).

- **Real-Time Polling Service (rtPS)**

Essa categoria oferece suporte aos fluxos de serviço de tempo real que geram pacotes de dados de tamanho variável em intervalos periódicos, como por exemplo, transmissões de vídeo sob demanda ou transmissões de pacotes VOIP (voz sobre IP) com supressão de silêncio. Periodicamente são oferecidas diversas oportunidades de requisição de banda unidirecional, que vão ao encontro das necessidades da conexão em tempo real (largura de banda) e que permitem que a SS especifique qual será o tamanho da concessão desejada. A SS não tem permissão de utilizar qualquer método de requisição de contenção ou de piggyback. Os principais parâmetros para essa categoria são: Prioridade de Tráfego (Traffic Priority), Taxa mínima (Minimum Reserved Rate), Taxa máxima (Maximum Sustained Rate) e Latência máxima (Maximum Latency Tolerance).

- **Non Real-Time Polling Service (nrtPS)**

A categoria nrtPS oferece suporte aos fluxos de serviço que não são em tempo real que geram pacotes de tamanho variável em intervalos periódicos, como, por exemplo, o tráfego gerado por aplicações FTP, e-mail, SMS, multicast/broadcast, MMS, entre outros. O serviço oferece periodicamente oportunidades de requisição unicast (polls). Uma oportunidade de requisição unicast é um intervalo do canal de upstream no qual somente uma SS específica pode enviar uma requisição de largura de banda para a BS. Esse processo difere de requisição de contenção no qual muitas SSs disputam entre si para transmitir suas requisições de banda. As alocações de banda para atender os fluxos de serviços nessa categoria serão fornecidas em intervalos maiores que os na categoria rtPS. Nessa categoria é permitido que sejam fornecidas às SS oportunidades de requisição mesmo durante períodos em que a rede esteja congestionada. Além disso, cada SS pode utilizar oportunidades de requisição de contenção e de piggyback. Os principais parâmetros para essa categoria são: Intervalo entre polls (Nominal Polling Interval), Taxa mínima garantida (Minimum Reserved Traffic Rate), Taxa máxima (Maximum Sustained Rate), e Prioridade de Tráfego (Traffic Priority).

- **Best Effort (BE)**

A intenção no agendamento de concessão do tipo BE é proporcionar um serviço eficiente para o tráfego de melhor esforço no sentido uplink. Para este serviço funcionar corretamente, a política de requisição/transmissão deve ser ajustada para que SS utilize requisições de contenção, o que resulta no uso das oportunidades de contenção além das oportunidades de transmissão unicast e as oportunidades de transmissão de dados concedidas pela BS. O serviço de melhor esforço é tipicamente oferecido para tráfegos de serviços Web, navegação pela Internet. É importante mencionar que para as categorias nrtPS e BE, o padrão especifica que a BS deve usar o parâmetro prioridade de tráfego para determinar a precedência na requisição do serviço e na geração da concessão para a transmissão de dados dessas duas categorias. Os principais parâmetros para essa categoria de serviço são: Taxa máxima (Maximum Sustained Rate) e Prioridade de Tráfego (Traffic Priority).

- **Extended Real-Time Polling Service (ErtPS)**

O objetivo desta classe é suportar serviços em tempo real que utilizem pacotes de tamanho variável, tais como serviços de voz sobre IP com supressão de silêncio. A BS pode prover alocações periódicas do canal de uplink que poderão ser utilizadas para solicitação de banda ou transmissão de dados. Por padrão, o tamanho das alocações corresponde ao valor do parâmetro Maximum Sustained Traffic Rate utilizado na conexão. A BS não pode alterar o espaço do canal uplink alocado até receber um pedido de mudança do tamanho de banda realizada por uma estação móvel. Esse pedido pode ser feito através de uma requisição piggyback ou utilizando um campo de sinalização do cabeçalho MAC. Se não houver banda unicast alocada para a SS, ela pode utilizar contenção para informar a BS que existem dados para serem enviados. Os principais parâmetros para essa categoria de serviço são: taxa Máxima (Maximum Sustained Rate), Taxa mínima (Minimum Reserved Rate), Latência máxima (Maximum Latency Tolerance), Tolerância ao Jitter (Jitter Tolerance) e Prioridade de Tráfego (Traffic Priority).

2.3.2 Fluxo de Serviço

Um fluxo de serviço, conforme já mencionado, é caracterizado por um conjunto de parâmetros de QoS tais como latência, retardo e garantias de vazão. Assim, após a SS ter efetuado seu registro, conexões são associadas a fluxos de serviço (uma conexão por fluxo de serviço) provendo referência para a requisição de banda. Adicionalmente, serviços do cliente podem necessitar modificar conexões já estabelecidas. O Fluxo de serviço define os parâmetros de QoS para as PDUs que são transmitidas na conexão.

O conceito de fluxo de serviço permite que haja um mecanismo para gerenciamento de QoS nos canais de downlink e uplink e é fundamental para o processo de alocação de banda. Uma SS solicita banda uplink por conexão (identificando implicitamente o fluxo de serviço), já a banda concedida pela BS para uma SS é um conjunto de concessões em resposta às solicitações por conexão feitas pela SS. Os fluxos de serviço são caracterizados pelos atributos descritos em seguida.

- **Identificador do Fluxo de Serviço (SFID)** – Um SFID identifica um único fluxo de serviço. Tanto na SS como na BS, o SFID serve como o principal identificador para o fluxo. Um fluxo de serviço tem no mínimo um SFID e uma direção associada.
- **Identificador da Conexão (CID)** – Uma conexão recebe um identificador único (CID) que é associado a um fluxo de serviço (SFID) quando este é admitido.
- *AdmittedQoSParamSet, ActiveQoSParamSet e AdmittedQoSParamSet* – Esses atributos são preenchidos de acordo com o tipo que o fluxo de serviço irá assumir, os quais podem ser: Provisionado, Admitido e Ativo. A descrição desses atributos, para cada um dos tipos é dada a seguir.
- **Provisionado** – Caracterizado por um conjunto de parâmetros de QoS fornecido por mecanismos externos aos definidos no padrão 802.16, como, por exemplo, pelo sistema de gerenciamento de rede. Os parâmetros *AdmittedQoSParamSet* e *ActiveQoSParamSet* para esse tipo de fluxo assumem valores nulos.

- **Admitido** – Define um conjunto de parâmetros de QoS para a BS utilizar, garantindo reserva da largura de banda e memória para viabilizar a ativação do fluxo. Esse tipo de fluxo de serviço possui recursos reservados pela BS para os parâmetros de `AdmittedQoSParamSet`, mas esses parâmetros não estão ativos (ou seja, os parâmetros de `ActiveQoSParamSet` assumem valores nulos). Os fluxos de serviço admitidos podem ter sido provisionados ou sinalizados por algum outro mecanismo.
- **Ativo** – Especifica um conjunto de parâmetros de QoS que define o serviço sendo atualmente provido para o fluxo de serviço. Somente um fluxo de serviço ativo pode encaminhar pacotes. Nesse tipo de fluxo de serviço, os recursos estão comprometidos pela BS para o seu conjunto de parâmetros `ActiveQoSParamSet`. Por exemplo, a estação base pode estar enviando mapas contendo concessões não solicitadas para a transmissão de um fluxo de serviço UGS (Unsolicited Grant Service). O conjunto de parâmetros `ActiveQoSParamSet` desse fluxo não é nulo.

2.4 Padrão IEEE 802.16e – Mobilidade no WiMAX

O padrão IEEE 802.16e - 2005 (Mobile WiMAX) [CHANG 2005] adiciona mobilidade ao padrão IEEE 802.16 - 2004, permitindo que as Estações Assinantes, agora nomeadas de Mobile Station (MS), movimentem-se livremente na área de cobertura de sua Estação Base, realizando os procedimentos de *handover* para as células cobertas por outras Estações Base sempre que necessário. O processo de *handover* ocorre quando há a transição de uma estação móvel de uma determinada BS para outra. Em outras palavras, um usuário móvel associado a uma dada BS, ao se distanciar de sua BS atual e se aproximar de outra BS, termina sua conexão com a BS mais distante e estabelece outra conexão com a BS mais próxima. Assim, a mobilidade provoca a necessidade de uma reserva adicional de recursos para suportar a transferência de conexões ativas entre células (*handover*).

No contexto de redes sem fio, a mobilidade tem grande importância, pois há uma tendência à conectividade contínua e onipresente nas vidas dos usuários, que cada vez mais necessitam manter conectividade e fazer uso de serviços de comunicação mesmo quando em movimento.

O padrão IEEE 802.16e - 2005 permite o modo de uso fixo e o modo de uso móvel sem linha de visada (Non Line of Sight - NLOS). Isto se torna possível devido a utilização de um novo método de multiplexação denominado Scalable OFDMA (SOFDMA) [AHSON 2007] e [ZHANG 2007]. Além disso, esse padrão inclui melhorias como:

- aumento do ganho do sinal, melhorando a penetração em ambientes internos (indoor);
- aumento da área de cobertura, com a introdução das tecnologias Adaptive Antenna Systems (AAS) e Multiple Input Multiple Output (MIMO);
- aumento do desempenho e da segurança, com a introdução das técnicas de codificação Turbo Coding e Low-Density Parity Check (LDPC);
- possibilidade de obter um desempenho semelhante com qualquer espaçamento de canais RF entre 1.25 e 14 MHz, pois elimina as dependências em relação ao comprimento do canal.

2.4.1 Conceitos Básicos

O padrão IEEE 802.16e tem por finalidade incrementar o mercado de acesso banda larga, aproveitando-se da mobilidade inerente dos dispositivos sem fio. Além disso, o intuito é preencher uma lacuna entre as redes locais sem fio, que suportam altas taxas de transmissão, e os sistemas celulares, que são altamente móveis, suportando serviços fixos e móveis para mercados corporativos e domésticos.

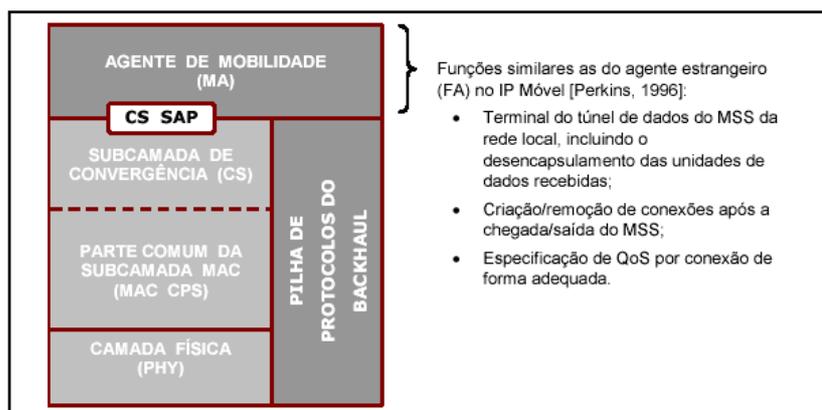


Figura 4. Pilha de protocolos do IEEE 802.16e para BS's

No padrão IEEE 802.16e encontram-se especificações para clientes móveis que trafeguem em velocidade veicular. O padrão inclui especificações para combinar o acesso por

estações fixas e móveis. Também são especificadas funções para suportar o processo de *handover*. Na Figura 4, são apresentados detalhes desse padrão, suas camadas e seus principais agentes.

2.4.2 Tipos de *Handover*

Handover, conforme já mencionado, é o processo que suporta a transição de uma estação móvel (MS) de uma célula para outra em uma rede sem fio. Quando esse processo suporta a transição entre redes de tecnologias distintas (por exemplo, de uma rede padrão IEEE 802.16 para outra rede padrão 802.11) tem-se o *handover* vertical. Por outro lado, quando envolvem redes de mesma tecnologia o processo é denominado de *handover* horizontal (Figura 5). Espera-se que a migração das estações móveis seja feita de forma transparente, isto é, rápida e sem perdas de dados, de modo a não interromper a continuidade dos serviços e aplicações. O processo de *handover* ocorre quando: (i) uma estação móvel, ao se deslocar, percebe que o sinal proveniente da sua estação base (BS) está abaixo de um limiar aceitável devido a fatores como atenuação e interferências que prejudicam a qualidade do sinal; (ii) uma estação móvel pode ser atendida por uma BS que oferece melhor nível de qualidade de serviço (QoS).

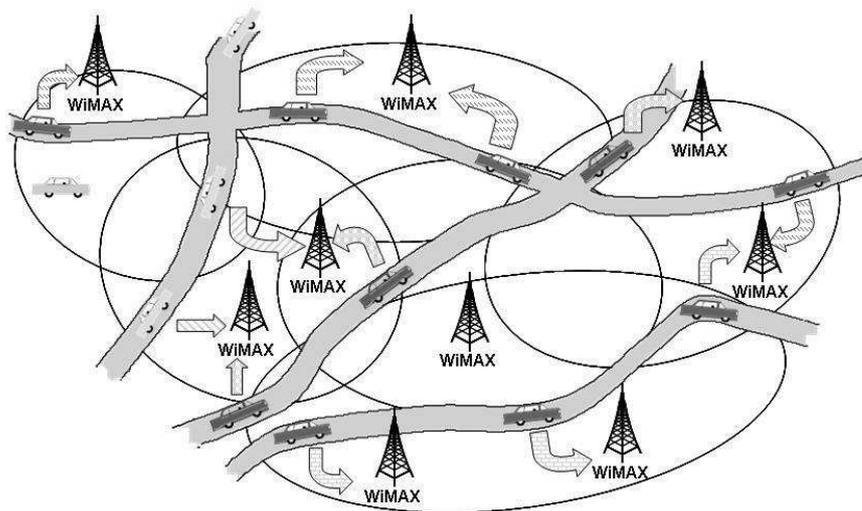


Figura 5. Cenário de uma rede IEEE 802.16e no momento do Handover

O padrão IEEE 802.16e adota três tipos de *handover*: (i) *Hard Handover* (HHO), denominado neste trabalho como *handover* tradicional; (ii) *Macro Diversity Handover*

(MDHO); e (iii) *Fast Base Station Switching* (FBSS). Desses, o HHO é obrigatório, enquanto o FBSS e o MDHO são métodos opcionais. A seguir, serão detalhados os três tipos de *handover*.

2.4.2.1 HHO (*Hard HandOver*)

O procedimento de *handover* do tipo HHO do padrão IEEE 802.16e [IEEE Std 802.16 - 2004] consiste de dois processos secundários: (i) a descoberta da topologia da rede antes do pedido de *handover*; e (ii) o processo de *handover* propriamente dito, incluindo a decisão de *handover*, a iniciação, *ranging* e reentrada na rede.

A Figura 6 mostra as etapas do procedimento de *handover* de cada um dos dois processos componentes do *handover* do tipo HHO, que serão detalhados a seguir.

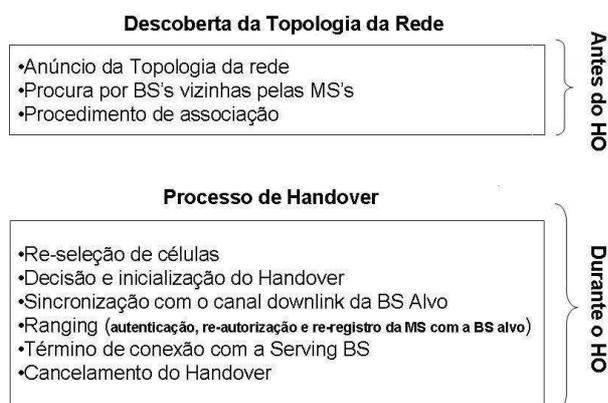


Figura 6. Etapas do procedimento de *Handover* da camada MAC

- ***Descoberta da Topologia da Rede***

O processo de descoberta da topologia da rede é realizado antes da iniciação do *handover* e envolve tanto a MS como a BS a qual ela está conectada. Esse processo possui três etapas: (i) anúncio de Topologia; (ii) procura por BSs vizinhas, feita pelas MSs; e (iii) associação.

A seqüência das trocas de mensagens para a descoberta da topologia da rede na camada MAC é ilustrada na Figura7. Na primeira etapa, anúncio da topologia da rede, a Serving BS transmite periodicamente, via backbone da rede, a informação de topologia, através da mensagem MOB_NBR-ADV, que inclui a informação de canal das BSs vizinhas. A Serving BS é aquela com que mais recentemente a MS realizou seu registro durante a entrada na rede ou durante o processo de *handover*.

De posse das informações provenientes dessas mensagens, a MS pode sincronizar-se com as BSs vizinhas sem ser necessário transmitir mensagens pelos canais downlink e uplink da Serving BS. Na etapa seguinte, procura por BSs vizinhas, a MS faz a varredura de suas BSs vizinhas para medir a qualidade do sinal das mesmas e seleciona alguma BS como candidata para o *handover* futuro.

Para tal, cada BS pode alocar intervalos de tempo para que suas estações móveis procurem e monitorem as BS vizinhas como possíveis alvos para o *handover*. O tempo dedicado para as MS realizarem a varredura por BS disponíveis é denominado “intervalo de varredura”. Quando uma MS necessita realizar com frequências buscas de possíveis BS Alvos, ela pode solicitar a alocação de grupos de intervalos de varredura e, com isso, criar múltiplas oportunidades de varredura. A BS Alvo é aquela com que uma MS pretende registrar-se ao final do processo de *handover*. Na mensagem de solicitação de varredura (MOB_SCN-REQ), a MS informa o tempo estimado necessário para realizar a varredura.

Uma MS, ao se afastar da Serving BS, constata que a qualidade dos sinais recebidos por algumas das BSs vizinhas da Serving BS são melhores do que a da própria Serving BS. Essas BSs vizinhas passam então a ser candidatas a BS Alvo para um futuro processo de *handover*. O processo de *handover* é iniciado pela MS quando essa transmite uma mensagem de pedido de *handover* a Serving BS. Em seguida, a Serving BS e as BSs vizinhas candidatas a BS Alvo trocam informações sobre a potência de sinal da MS com cada uma dessas BSs, através de mensagens que são normalmente enviadas no *backbone* da rede. Com base nas informações recebidas, a Serving BS seleciona a BS Alvo mais indicada. A MS, ao receber a mensagem de resposta ao pedido de *handover* da Serving BS, libera a conexão com a Serving BS, estabelece uma nova conexão com a BS Alvo indicada pela Serving BS e conduz o processo de reentrada na rede controlada pela BS Alvo. Após o processo de reentrada nessa rede, a MS pode transmitir ou receber mensagens da "nova" BS.

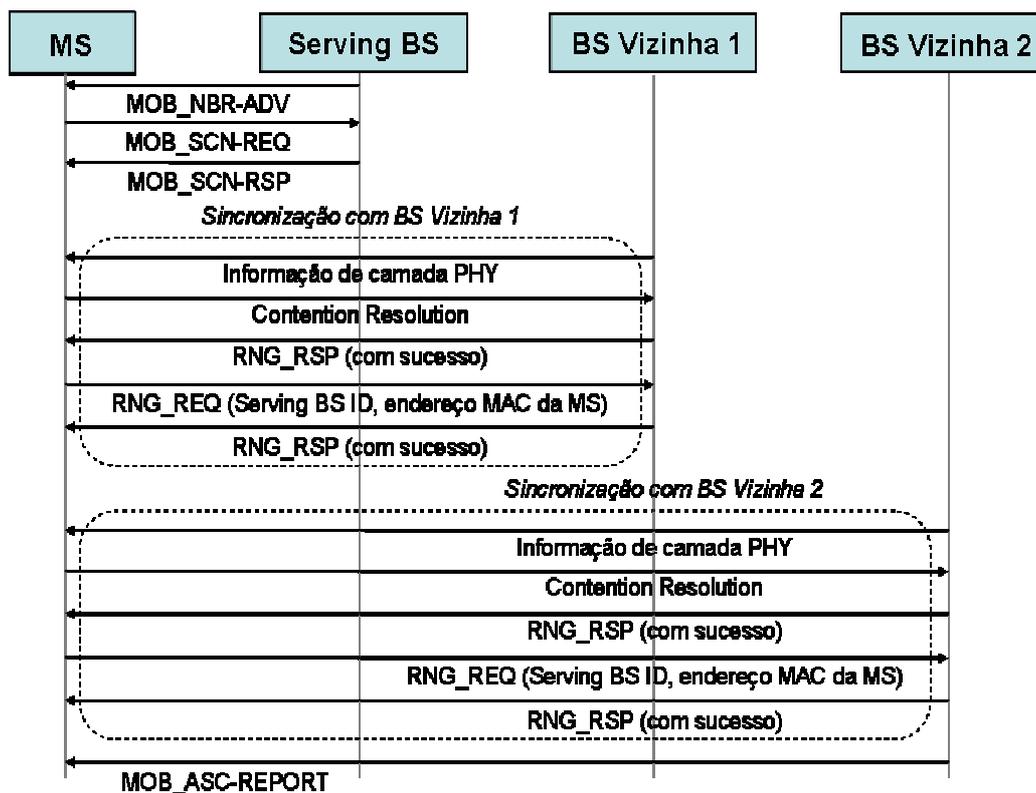


Figura 7. Troca de mensagens na descoberta da topologia da rede

- As mensagens do tipo MOB_SCN-REQ / RSP permitem que a MS possa realizar a sincronização com o canal downlink de cada BS vizinha. Os dados recebidos pela MS, durante o processo de varredura, são guardados na Serving BS. A terceira e última etapa da descoberta da topologia da rede é a associação entre a MS e a candidata a BS Alvo. Opcionalmente, o processo inicial de ranging pode ser executado durante o procedimento de varredura. A MS mantém valores iniciais de ranging das BSs para uso em um *handover* futuro.

- **Processo de HandOver**

Após o processo de descoberta de topologia, tem lugar o processo de *handover* do tipo HHO propriamente dito, o qual consiste de seis etapas: Re-seleção de células (detecção de *handover*), Decisão e iniciação, Sincronização com o canal *downlink* da BS Alvo, *Ranging*, Término da conexão com *Serving BS* e Cancelamento do *handover*. Cada uma dessas etapas é detalhada em seguida.

- **Re-seleção de células** - MS realiza uma varredura e/ou associação com uma ou mais BSs com o objetivo de escolher uma BS Alvo, segundo alguns requisitos como a qualidade do sinal e/ou a do serviço (QoS). Esta etapa pode ser abreviada, uma vez que inclui as mesmas operações que o processo de descoberta da topologia da rede.
- **Decisão e iniciação do *handover*** - MS deve migrar de uma Serving BS (BS na qual a MS está atualmente conectada) para uma BS Alvo, sendo que tal decisão pode ser originada na estação móvel ou na Serving BS.
- **Sincronização com o canal *downlink* da BS Alvo** - etapa onde são obtidos os parâmetros de uplink e downlink.
- ***Ranging*** - etapa onde são realizadas negociações de capacidades e re-entrada na rede (autenticação, re-autorização, re-registro da estação MS com a nova BS).
- **Término do Contexto da MS** – Serving BS é responsável por liberar as conexões da MS e por terminar o contexto associado a MS (informações nas filas, contadores, temporizadores, informações de supressão de cabeçalho).
- **Cancelamento do *handover*** – A MS pode solicitar o cancelamento do *handover* a qualquer momento durante o processo.

O processo de *handover*, ilustrado na Figura 8, do tipo HHO é executado quando a MS move-se da Serving BS para a BS Alvo. No início, a MS é responsável pela re-seleção da célula usando a informação da topologia da rede obtida do processo de descoberta da topologia da rede anterior. Em seguida, a decisão do *handover* e o processo de iniciação são efetuados. Tanto a MS como a Serving BS podem executar a etapa de decisão e iniciação do *handover*, usando respectivamente, a mensagem MOB_MSHO-REQ e MOB_BSHO-REQ. Neste minicurso será abordado apenas o procedimento do *handover* iniciado na MS.

A Serving BS notifica o pedido do *handover* as BSs vizinhas emitindo a mensagem de notificação de *Handover* (HO notification) e notifica a MS emitindo a mensagem MOB_MSHO-RSP. Após ter recebido a resposta da Serving BS, a MS escolhe a BS Alvo para o *handover* e transmite a mensagem MOB_MSHO-IND a Serving BS para que o *handover* de fato possa ser iniciado.

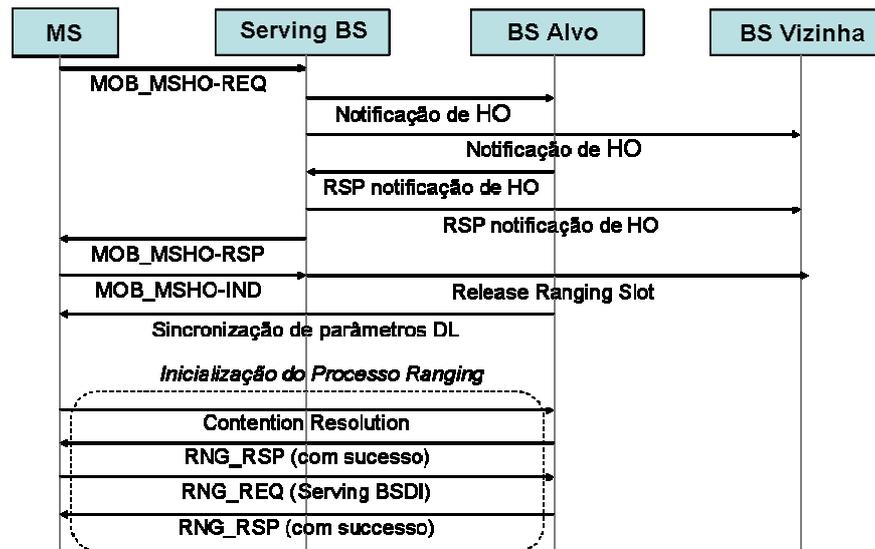


Figura 8. Etapas detalhadas do processo de Handover

Após a iniciação do *handover*, a etapa de sincronização com o canal de downlink da BS é efetuada. Em seguida, é iniciado o processo de ranging, que pode ser realizado em período de contenção ou não para melhor obtenção das características físicas do canal. Vale ressaltar que a MS deve estar sincronizada no canal downlink com a BS Alvo e obter parâmetros de downlink e uplink via mensagens DCD/UCD, quando então as mensagens RNG_REQ e RNG_RSP são trocadas para terminar o processo de ranging. Se a mensagem RNG_REQ contiver o BSID, identificação da Serving BS, a BS Alvo pode obter a informação da MS da Serving BS através do backbone da rede. Caso a MS já esteja associada a uma BS Alvo no estágio de sincronização com o canal downlink da BS Alvo, esse processo pode ser abreviado.

Após a etapa inicial de ranging ser realizada com sucesso, isto é, todos os parâmetros físicos serem negociados, a etapa de reentrada na rede é iniciada conforme mostrado na Figura 9. Essa etapa inclui a autorização e o registro da MS com a nova BS. A BS Alvo pede informações de autorização da MS ao servidor de autorização através do backbone da rede. Após a autorização ser concedida, o registro da MS na nova BS é executado através da troca de mensagens REG-REQ e REG-RSP. As etapas de negociação da capacidade, autorização da MS e mudança de chave e registro podem ser abreviadas, dependendo do nível da associação usado após o registro bem sucedido da MS na BS Alvo. A MS emite uma mensagem MOB_HO-IND à Serving BS para notificar que o *handover* foi terminado.

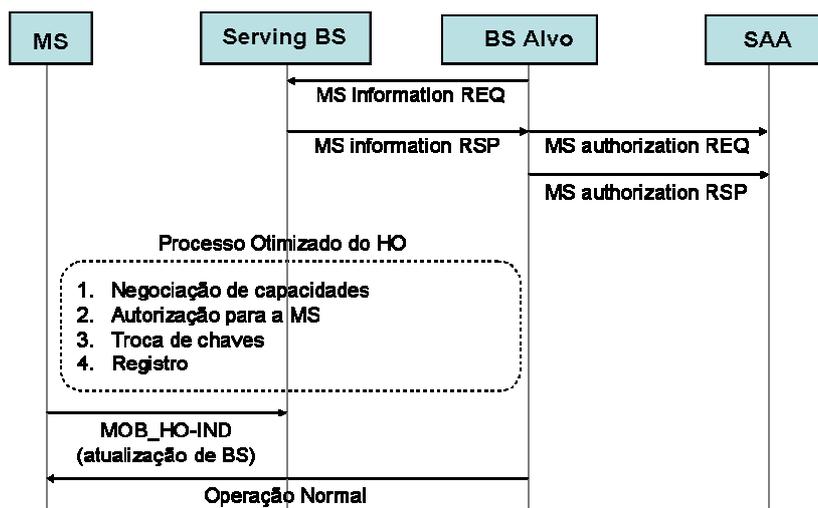


Figura 9. Processo de reentrada na rede

Durante o processo do *handover* do tipo HHO, a estação móvel não é capaz de transmitir ou receber dados e, portanto, estes dados podem chegar ao seu destino atrasados. Em consequência, pacotes pertencentes a aplicações sensíveis ao atraso, tais como serviços de voz sobre IP (VoIP) ou serviço de fluxo de vídeo (video streaming service), podem ser entregues normalmente com atrasos de 20 a 25ms [HEYAIME-DUVERGE 2002]. Se o atraso for grande, pacotes serão rejeitados e a probabilidade de perda do pacote será aumentada. Em suma, no processo de *handover*, dependendo do tipo de aplicação, os pacotes recebidos com atraso são descartados e, conseqüentemente, é aumentada a probabilidade de perda de pacotes. Logo, um grande desafio é restabelecer a conexão com a próxima BS em presença de *handover* no menor tempo possível. Assim, é necessário que a detecção do início de *handover* seja feita antes que haja a perda da conexão. Portanto, é crucial achar o ponto certo para iniciar o processo de *handover*.

O tempo gasto para efetuar o processo de *handover* pode ser visto como um dos principais fatores que afetam a qualidade de serviço oferecida aos usuários. Diversas técnicas foram desenvolvidas para aperfeiçoar o HHO do padrão 802.16e. As melhorias no processo de *handover* podem ser vistas nos *handovers* do tipo MDHO e FBBS.

2.4.3 MDHO (Macro Diversity HandOver)

O processo de *handover* do tipo MDHO, conforme ilustrado na Figura 10, inicia-se quando a MS decide transmitir e receber mensagens para múltiplas BSs ao mesmo tempo.

O conjunto de BSs envolvidas em um *handover* do tipo MDHO com uma estação móvel é denominado de conjunto de vizinhas (conjunto Diversity). Este conjunto é

monitorado pela MS que verifica continuamente a força do sinal das BSs integrantes e pode sugerir atualizações varrendo as BSs vizinhas e escolhendo novas BSs para o conjunto. Dentre as BSs do conjunto de vizinhas, uma é selecionada para ser a BS Âncora. A BS Âncora é a BS na qual a MS pretende se registrar ao final de um processo de *handover*. Na verdade, ela é a BS que deixa de ser “Alvo” para se tornar Âncora. Um caso particular é quando esse conjunto é formado por uma única BS e, portanto, essa mesma BS também será a BS Alvo e Âncora.

No canal downlink, quando se utiliza o método MDHO, duas ou mais BSs transmitem dados os quais são combinados pela MS. Para o canal uplink MDHO, a transmissão de uma MS é detectada por múltiplas BSs Vizinhas, sendo que algumas receberão a mensagem mas não serão incluídas no conjunto Diversity devido ao fato do nível do sinal não ser suficiente.

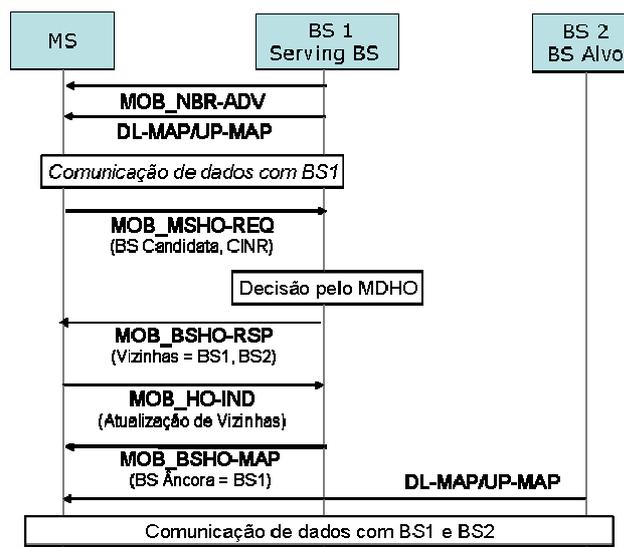


Figura 10. Processo MDHO

2.4.4 FBSS (Fast Base Station Switching)

O processo de tomada de decisão no *handover* do tipo FBSS é realizado na BS. Nesse tipo de *handover*, conforme é ilustrado na Figura 11, a MS monitora continuamente as BSs de seu conjunto de vizinhas e uma delas é selecionada para ser a BS Alvo. Após ser efetuada a seleção da BS Alvo, a MS comunica-se somente com a BS Alvo escolhida tornando-a BS, Âncora, responsável pelo envio/recepção de mensagens de uplink e downlink, incluindo

conexões de tráfego e de gerência. A transição de uma BS Âncora para outra (isto é, migração de BSs) ocorre sem a invocação de mensagens explícitas de *handover*, isto é, são eliminadas algumas etapas do processo de *handover* do tipo HHO. Nesse *handover*, existem dois mecanismos para que a MS e a BS executem a troca de BSs Âncora. O primeiro mecanismo usa mensagens de *handover* (HO) e no segundo a MS e a BS usam o canal de retorno rápido (fast-feedback) para atualizar a lista do conjunto de vizinhas. Nos dois mecanismos, a BS Âncora escolhida deve pertencer à lista do conjunto de vizinhas (conjunto Diversity).

No primeiro mecanismo, que usa mensagens de *handover*, é possível escolher uma das duas alternativas descritas a seguir. Na primeira alternativa, a MS reporta suas informações à nova BS Âncora através da mensagem MOB_MSHO-REQ. Na segunda, a antiga BS Âncora é quem informa à MS da mudança de BSs âncora através da mensagem MOB_BSHO-REQ ou da mensagem MOB_BSHO-RSP, passando o tempo estimado para a troca. A MS pode aceitar ou rejeitar a troca de BSs âncora através da mensagem MOB_HO-IND (campo MDHOFBSS_IND_type).

No segundo mecanismo, seleção de BS Âncora com retorno rápido (Fast Anchor BS selection feedback), a MS e a BS usam o canal de retorno rápido (fast-feedback) para atualizar a lista do conjunto Diversity. Caso a MS necessite enviar uma informação de seleção de BS Âncora, também é encaminhado o identificador temporário (TEMP_BSID). Todas as BS recebem ao entrar na lista do conjunto Diversity um identificador temporário.

Para que o *handover* do tipo FBSS possa ser realizado entre a MS e um grupo de BSs, é necessário que ocorram as condições descritas em seguida.

- BSs envolvidas nesse tipo de *handover* estejam sincronizadas.
- Quadros enviados pelas BSs cheguem na MS dentro de um intervalo de tempo pré-definido.
- Quadros das BSs envolvidas nesse tipo de *handover* estejam sincronizados.
- BSs envolvidas operem na mesma frequência.
- BSs envolvidas sejam obrigadas a compartilhar ou transferir o contexto MAC. Tal contexto inclui toda informação que uma BS e uma MS trocam durante o processo de entrada na rede (Network Entry), particularmente o estado de autenticação, de forma a aproveitar esta informação em outras BSs do mesmo conjunto de vizinhas. O contexto inclui também Fluxos de Serviços, o

mapeamento correspondente de conexões associadas à MS e chaves de autenticação e criptografia atuais associadas com as conexões.

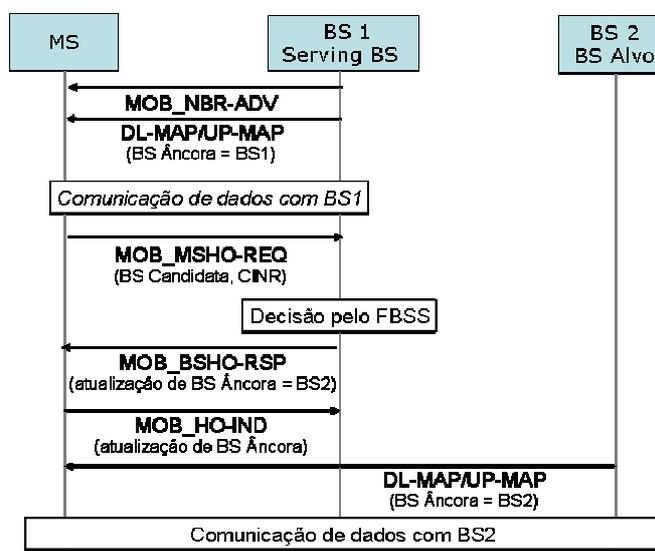


Figura 11. Processo FBSS

Analisando as formas de *handover* do IEEE 802.16e, pode-se observar que o processo de HHO não é capaz de atender os requisitos de QoS das conexões dos clientes móveis, já que durante o processo do *handover* do tipo HHO, a estação móvel não é capaz de transmitir ou receber dados e, portanto, estes dados podem chegar ao seu destino atrasados. O processo de tomada de decisão no *handover* MDHO é realizado pela MS, enquanto no FBSS é realizado pela BS. Assim, o MDHO é iniciado quando uma MS decide transmitir/receber mensagens para/de todas as BSs do conjunto de vizinhas no mesmo intervalo de tempo. Já no *handover* FBSS, a MS comunica-se somente com a BS Âncora para o envio/recepção de mensagens de uplink e downlink. A transição de uma BS Âncora para outra ocorre sem a invocação de mensagens explícitas de *handover*, isto é, são eliminadas as etapas do processo de *handover* do tipo tradicional. Os fatores descritos em seguida são os responsáveis por tornar o FBSS/MDHO rápido (fast).

- Possuir uma lista de BSs candidatas (Conjunto de Vizinhas) ao *handover*.
- Negociar somente com uma BS (BS Âncora).
- Coordenação e influência nas decisões de *handover* serem mais centralizadas na BS.
- Evitar vários passos do *handover* tradicional.

- Compartilhar informações de fluxo, de autenticação, de entrada de rede previamente com as BSs envolvidas.
- As BSs listadas no conjunto de vizinhas recebem as necessidades da MS antes de iniciar o processo de *handover*. Um exemplo de melhora do *handover* do tipo fast é reduzir a lista das BSs vizinhas [CARVALHO JR 2005]. No FBSS, uma melhoria é que o contexto MAC é compartilhado com as BSs do conjunto de vizinhas.

2.5 *Next Generation Networks* (NGN)

A International Telecommunication Union (ITU-T) [ITU-T], mantém um grupo de estudo desde 2003 sobre o tema NGN, e agora se concentra no desenvolvimento de normas detalhadas necessárias para a implementação dessas redes, buscando harmonizar mundialmente, em conjunto com outros organismos, as diferentes abordagens de arquitetura para as NGN. Um rede de nova geração (NGN) é uma rede baseada em pacotes capaz de: (i) prover serviços inclusive de telecomunicações; (ii) fazer uso de múltiplas tecnologias de transporte de banda larga com QoS; (iii) ter serviços (Vídeo, Dados e Telefonia) independentes das tecnologias subjacentes relacionadas a transporte (tecnologias de rede acesso e rede *core*) (figura 13); (iv) oferecer acesso irrestrito pelos usuários a diferentes provedores de serviços e (v) suportar mobilidade generalizada que permitirá uma provisão consistente e ubíqua de serviços para os usuários. Para evitar os custos da troca abrupta dos equipamentos usados atualmente, a NGN pode evoluir a partir da geração atual de redes (CGN), o que implicaria em uma fase intermediária (Near GN) antes da convergência global de serviços de telecomunicações esperado nos sistemas NGN (Figura 12)[ITU-R]. Para alcançar a mobilidade generalizada (item v da definição NGN ITU) é necessária a convergência entre as redes fixas e móveis, contexto onde estão inseridas as propostas do presente trabalho. Já a NGN final deve ser puramente baseada em pacotes, incluir suporte completo a multimídia, banda larga e uni/multi/broad-casting. Inclusive com capacidade de personalização como nos serviços de IPTV, onde cada usuário poderá escolher sua programação de forma independente.

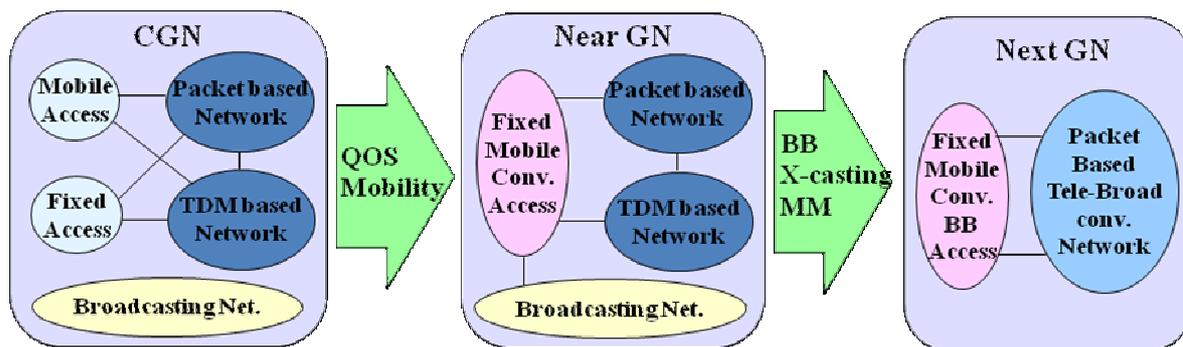


Figura 12. Convergência global de serviços de telecomunicações

O desafio é grande pois os grupos de estudo ITU-T terão que integrar as redes orientadas a pacote e as redes orientadas a circuito nas Redes de Próxima Geração (NGN) suportando vários serviços multimídia o que demandará uma extensão dos princípios arquitetônicos em perspectivas tanto verticais quanto horizontais. No aspecto vertical, mais estudos são necessários nas áreas de controle de serviços (serviços controlados pela rede de núcleo multi-serviços) e suporte de operações para atender a vários serviços multimídia sobre múltiplas tecnologias. Nos aspectos horizontais, suportar situações complexas que envolvem os ambientes de usuários (incluindo redes domésticas) e várias tecnologias de acesso. Essas capacidades são necessárias para suportar as futuras redes NGN.

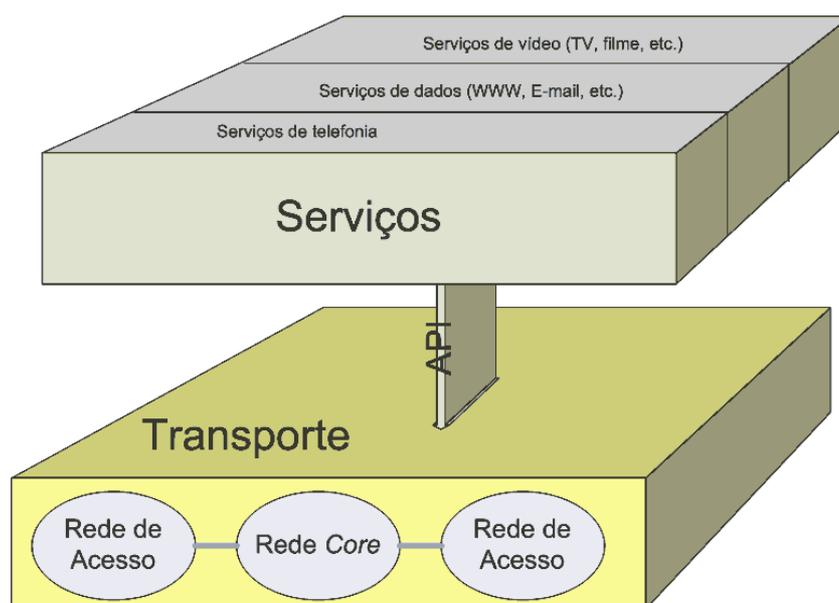


Figura 13. Modelo de convergência NGN

Além disso existe uma necessidade da indústria em fornecer *roaming* global e total mobilidade para os usuários de diferentes tecnologias de acesso e/ou diferentes provedores de acesso, para que possam desfrutar de serviços transparentes enquanto se movem dentro ou através de redes. Ainda segundo a ITU-T, mais de 50% dos assinantes no mundo inteiro são móveis em oposição aos assinantes da rede fixa, e a proporção continua a mudar a favor de assinantes móveis. Em muitos mercados, principalmente os países em desenvolvimento, essa proporção é muito maior, muitas vezes superior a 90%. Portanto, a mobilidade tem grande importância e deve ser estudada e definida em um contexto global, uma vez que as redes móveis são muito mais do que as tecnologias de acesso via rádio. Estas exigem uma grande infra-estrutura de rede para se conectar.

2.6 Trabalhos relacionados

As propostas apresentadas no presente trabalho situam-se em dois contextos diferentes: (i) o mecanismo SUTIL é voltado para *handovers* verticais em ambiente NGN e (ii) o algoritmo ARPRAP tem foco em *handovers* horizontais em redes WIMAX. A Seção 2.6.1 descreve os trabalhos relacionados ao mecanismo SUTIL, enquanto que na Seção 2.6.2 estão apresentados os trabalhos relacionados ao algoritmo ARPRAP.

2.6.1 Trabalhos relacionados ao SUTIL

Nos últimos anos, diversos trabalhos ([CHANG 2005], [KIM 2005],[ZAHARAN 2006], [KANG 2006] e [HORRICH 2007]), apresentaram propostas para melhorias no processo de *handover* (HO), a fim de fornecer garantias de QoS em aplicações multimídia na presença de mobilidade, por meio da redução do atraso e perda de pacotes gerados pelo processo de *handover*. No entanto, nenhum desses trabalhos propôs um mecanismo de seleção de redes que atendesse ao mesmo tempo diversos usuários candidatos a *handover* e, para cada usuário, todo o conjunto de serviços em uso. Além disso, a maioria dos trabalhos atuais não adotam uma abordagem holística, que considera tanto os pontos de vista do usuário quanto da rede, assim, alcançando uma solução global e evitando a sobrecarga das redes de destino. Um algoritmo adaptativo baseado em tempo de vida para *handover* vertical é apresentado em [ZAHARAN 2006], que considera como parâmetros para o processo de seleção: a força do sinal da rede sem fio, a latência do *handover*, e também os requisitos de QoS da aplicação e a

tolerância a atraso. Este algoritmo procura minimizar a carga de sinalização imposta pelo processo de *handover* e satisfazer os diferentes requisitos do aplicativo usando um limiar associado à força do sinal (um limiar diferente para cada aplicativo). O algoritmo foi analisado em um modelo onde redes celulares 3G e WLAN estão em sobreposição e as estações móveis são equipadas com duas interfaces que lhes permitem comunicar-se com ambas as redes. Além disso, um software cliente de mobilidade multi-interface está instalado nas MSs. Este software faz a sinalização *IP Móvel* com agentes locais e externos. Ele verifica periodicamente as interfaces disponíveis e mede o *Received Signal Strength (RSS)*. Em seguida, ele seleciona de forma inteligente a melhor rede de acesso de acordo com o algoritmo pré-definido de *handover* vertical (VHO). Portanto, em [ZAHARAN 2006] a decisão não considera o ponto de vista da rede, enquanto no SUTIL a decisão é baseada em três diferentes pontos de vista: as aplicações, as estações móveis, e as redes. Outra diferença entre o trabalho descrito em [ZAHARAN 2006] e nosso trabalho é que a solução fornecida pelo SUTIL é global, uma vez que considera, para uma dada estação base, todos os dispositivos que são simultaneamente candidatos a *handover*. No SUTIL, além de considerar alguns dos mesmos parâmetros utilizados em [ZAHARAN 2006], a solução é baseada em Função de Utilidade e Programação Linear Inteira, prioriza as redes com menor consumo de energia e de maior relevância para a aplicação, considerando os parâmetros de rede, como largura de banda, perda de pacotes e atrasos, bem como as preferências do usuário e da aplicação.

Os autores em [PANDEY 2004] propõem uma estratégia para dinamicamente alocar recursos para as aplicações com requisitos severos de QoS no conjunto de células-alvo para o próximo *handover* em uma rede celular. Outro trabalho [CARVALHO JR 2008] propõe uma melhoria do processo de *handover* para redes WiMAX com base em um aperfeiçoamento do mecanismo para prever o conjunto de células-alvo para a próxima entrega, levando em consideração o perfil comportamental de um usuário e os contextos de redes e serviços. O mecanismo de seleção de redes SUTIL pode ser usado para melhorar ainda mais o processo de *handover* dos trabalhos propostos em ([PANDEY 2004] e [CARVALHO JR 2008]).

Sobre as soluções *middleware* existentes para o gerenciamento de *handover*, o trabalho em [KANG 2006] apresenta uma arquitetura que engloba um módulo para gerenciamento pró-ativo, que é capaz de decidir quando um *handover* deve ser feito como uma consequência de requisitos pré-definidos de aplicação, as preferências do usuário, bem como informações sobre a rede. Para fornecer conectividade contínua em ambientes de

computação ubíqua composto de interfaces de rede heterogêneas, esse trabalho propõe um esquema inteligente de gerenciamento de *handover* baseado em contexto e uma arquitetura *middleware* baseada em um conjunto de perfis. O conjunto de perfis engloba *perfis de aplicação*, que representam os requisitos das aplicações relevantes; *perfis de usuário*, refletindo quaisquer informações específicas do usuário e requisitos; *perfis de trabalho*, que são informações de contexto inferida a partir de uma *context engine*, e um *perfil abstrato*, que é captada por um gerenciador de aplicação a partir do **perfil de trabalho** de todas as aplicações e cujo objetivo é suportar multi-aplicações. Com base nestes perfis, os autores propõem um gerenciador de decisão de VHO que realiza *handover* vertical eficaz e um agente de aplicativo que gerencia o modo de operação da aplicação para um serviço contínuo. A concepção do SUTIL foi influenciada por este trabalho no sentido de que também utilizam o conceito de perfis de aplicação. No entanto, esta é a única semelhança nossa proposta apresenta para o trabalho descrito em [KANG 2006]. O SUTIL utiliza uma abordagem de solução global para alcançar a melhor solução possível para atender ao mesmo tempo serviços de todos os usuários de uma BS específica, que são candidatos a *handover*.

O artigo [CHALMERS 2006] propõe uma solução chamada dyCARD (*dynamic Candidate Access Router Discovery*), que tem uma abordagem distribuída baseada em aprendizagem para o problema de coletar e distribuir informações de alto nível sobre as capacidades de pontos de acesso (AP) alvo e roteadores de acesso (AR) correspondentes. Estas capacidades descrevem os serviços que são suportados por um roteador de acesso em particular, bem como as propriedades da conexão, como a largura de banda disponível ou custo. No artigo, os autores modelam o processo de propagação de informação como uma instância de um cache distribuído e hierárquico. Roteadores de acesso coletam informações que descrevem o que está disponível na vizinhança imediata a fim de facilitar a escolha pela estação móvel do melhor caminho através da rede. O protocolo proposto adota uma abordagem baseada em aprendizagem, onde os roteadores de acesso vizinhos descobrem uns aos outros através dos padrões de *handover* da estação móvel. Uma vez descoberto, dois roteadores vizinhos de trocam diretamente as capacidades. Enquanto uma estação móvel se prepara para fazer um *handover*, ela consulta o seu roteador de acesso local para obter informações sobre pontos de acesso alcançáveis. Os recursos retornados podem então ser passados como parâmetros para um algoritmo de *handover* apropriado baseado em política. Estes algoritmos de *handover* baseados em política fazem uso de propriedades da camada

física, tais como *Received Signal Strength (RSS)*, *Bit Error Rate (BER)*, ou *Signal to Noise Ratio (SNR)*, juntamente com estas informações de camada superior. No SUTIL abordamos a forma de melhorar a decisão de *handover* e, diferentemente de dyCARD, não estamos preocupados com a forma como as informações necessárias para tal decisão é conseguida ou trocada entre as estações. Então, a solução dyCARD poderia ser usada junto com mecanismo SUTIL, coletando e disseminando todas as informações necessárias para o processo de decisão do SUTIL.

O trabalho [HORRICH 2007] aborda gerência de mobilidade entre Radio Access Networks (RANs) heterogêneos usando uma solução baseada em lógica nebulosa e aprendizagem por rede neural. Um algoritmo multi-critério nebuloso de *handover* vertical que melhora o desempenho do sistema é proposto. O algoritmo usa como entrada: a potência do sinal recebido do sistema atual, a potência do sinal recebido do alvo pré-selecionado, a carga no sistema atual, a carga no sistema alvo e da velocidade da estação móvel. Uma rede neural adapta automaticamente os parâmetros *Fuzzy Logic Controller* às flutuações do ambiente e as variações de tráfego fornecendo uma saída para entradas imprecisas, incertas ou incompletas. Assim, esse método é adequado para o problema de decisão de *handover* vertical, uma vez que supera a flutuação e a incerteza do ambiente de rádio, bem como a heterogeneidade dos parâmetros entre sistemas (por exemplo: sombreamento, as variações do modelo de tráfego). As principais diferenças entre este trabalho e o SUTIL estão nas técnicas adotadas para resolver o problema de decisão de *handover* e nos parâmetros considerados na decisão em ambas as soluções. A solução SUTIL prioriza a escolha das redes com menor consumo de energia e de maior relevância para a aplicação, considerando os parâmetros de rede, como largura de banda, perda de pacotes e atrasos, bem como as preferências do usuário e aplicação. Um importante diferencial da nossa abordagem é que a solução fornecida pelo SUTIL é global, uma vez que considera, por uma dada estação base (BS), todos os dispositivos que são simultaneamente candidatos a *handover*. O trabalho [HORRICH 2007] assume uma arquitetura *internetworking* centralizada entre UMTS e WLAN controlada por uma entidade, chamada *Common RRM*, que é responsável pelo processo de decisão de *handover*. Depois de decidir a melhor rede, a entidade *Common RRM* solicita à estação móvel que se anexe ao novo RAN. A decisão leva em conta o ponto de vista dos terminais móveis e não considera o ponto de vista das redes de destino. Portanto, neste trabalho, a solução

fornecida não pode ser considerada como global, e pode sofrer com o problema de sobrecarga de um RAN que, simultaneamente, receber um grande número de MSs.

Em [AKYILDIZ 2005] os autores apresentam uma nova arquitetura para NGN chamada *Architecture for Ubiquitous Mobile Communications (AMC)*. AMC integra sistemas sem fio heterogêneos (GPRS, cdma2000, UMTS, WLAN, etc) usando duas entidades: um *Network Inter-operating Agent (NIA)* e um *Internetworking Gateway (IG)*. O IG reside em cada sistema sem fio atuando como *gateway*, conectado a uma única entidade, o NIA. O NIA é suportado por um provedor de terceiros e reside na Internet para que os provedores de rede não tenham que criar SLAs separados (Service Level Agreements) com cada outro provedor. Em vez disso, apenas um SLA é necessário, o que torna todo o sistema heterogêneo totalmente escalável e mais atrativo para os provedores de rede, contribuindo assim para permitir um cenário de total interoperabilidade como está previsto para as NGN. A AMC consegue transparência em relação à heterogeneidade de sistemas individuais usando o Internet Protocol (IP) como o protocolo de inter-conexão. A solução inclui o projeto de protocolos de autenticação, autorização e faturamento. Além disso, abrange algoritmos para a seleção das melhores redes e protocolos de gerência da mobilidade para resolver a questão de mobilidade entre os sistemas. O NIA inclui um subsistema chamado *handover management unit*, que decide se um pedido de *handover* vertical pode ser cumprido, verificando com o banco de dados do operador a existência de um SLA com o operador de origem da MS. Sempre que for aplicável, o NIA também atua como mediador entre as diferentes redes (por exemplo, para a transferência de perfis de serviço do usuário). Além disso, a *handover management unit* decide a melhor rede disponível, usando um esquema de seleção de rede híbrida que combina mecanismos seleção baseados nas estações móveis e na rede. Nossa proposta pode ser considerada como um complemento a este trabalho. Por um lado, o algoritmo SUTIL pode substituir a solução baseada na estação móvel, fornecendo uma solução que considera, para uma estação base, todos os dispositivos que estão conectados, ou ela pode agir como o tomador de decisão à *handover management unit* em execução na Internet (NIA) em vez de executar nas BSs como propomos. Por outro lado, NIA poderia ser a entidade de rede central para SUTIL.

O artigo [JIA 2006] propõe um esquema de seleção de rede centralizado, de baixa complexidade, o com o objetivo de distribuir otimamente os usuários finais sobre as redes disponíveis em um sistema heterogêneo sem fio, no sentido de maximizar a eficiência no

espectro global. Dois algoritmos sub-ótimos são propostos para resolver este problema de otimização, que se baseiam em medidas de *absolute bandwidth request (ABR)* e *relative bandwidth request (RBR)*, respectivamente. O artigo também usa uma entidade de rede central, o *Network Interoperating Agent (NIA)*, previamente descrito em [AKYILDIZ 2005]. Essa entidade central é responsável pela difusão de informações do sistema periodicamente a todas as MSs em redes heterogêneas. O esquema proposto é para ser realizado no *handover management unit (NIA)*, evitando assim a sobrecarga de difusão de informações do sistema que ocorrem em esquemas distribuídos e híbridos. Portanto, a proposta é muito atraente para os operadores de rede que estão relutantes a divulgar seu sistema de informação para os concorrentes comerciais. O esquema de seleção de redes centralizado proposto no artigo leva em consideração apenas a largura de banda na sua decisão. O SUTIL pode fazer uso de uma entidade de rede central, como NIA e seu mecanismo de decisão é executado de forma distribuída em cada BS do sistema sem fio. Além da banda, SUTIL leva em conta informações de perda de pacotes, atraso e energia para maximizar a eficiência global entre todas as MSs que são candidatas simultaneamente ao processo de *handover*.

Em [CHEN 2005] os autores propõem um esquema de *handover* vertical, chamado *Active Application Oriented*, no qual a MS pode ativamente pedir e iniciar o *handover*, em contraste com outros sistemas, onde as MSs participam passivamente do processo de *handover*. A proposta visa reduzir o consumo de energia causada pela ativação desnecessária de interfaces de rede durante o *handover*. Os autores argumentam que permitir à MS ativamente decidir quando fazer *handover* e a qual rede conectar de acordo com os requisitos das aplicações em execução [LATVAKOSKI 2002] permite a aplicação ser atendida pela rede mais adequada. Uma função para quantificar o nível de QoS fornecido por uma determinada rede é calculada, função esta que considera fatores como largura de banda, latência, custo, consumo de energia, etc. O artigo também apresenta um método de descoberta baseado em um *Location Service Server (LSS)*, bem como algoritmos de decisão de *handover* onde a decisão é centralizada na MS. Portanto, podemos notar que em [CHEN 2005], a decisão baseia-se apenas o ponto de vista da estação móvel, enquanto no SUTIL a decisão é baseada em três diferentes pontos de vista: as aplicações, as estações móveis, e as redes. A solução fornecida pelo SUTIL é global, uma vez que considera, por uma dada estação base, todos os dispositivos que estão conectados a ela. No SUTIL, além de considerar alguns dos mesmos parâmetros utilizados em [CHEN 2005], a solução é baseada em Função de Utilidade e Programação Linear Inteira, e prioriza as redes com menor consumo de energia e de maior

relevância para a aplicação, considerando os parâmetros de rede como largura de banda, perda de pacotes e atraso, bem como as preferências do usuário e aplicação.

O trabalho [MCNAIR 2004] apresenta um tutorial sobre o projeto e desempenho para a gerência de *handover* vertical em um vislumbrado ambiente 4G de múltiplas redes e descreve várias arquiteturas de rede e tecnologias atualmente em desenvolvimento além do 3G, incluindo WLANs, celular e IP Móvel. O problema de *design* do *handover* vertical no contexto do ambiente imaginado é explorada. Os problemas de pesquisa abertos para alcançar uma decisão *handover* transparente e interoperável e algoritmo de detecção são descritos. Os autores em [MCNAIR 2004] argumentam que o processo de *handover* tradicional é insuficiente face aos desafios dos ambientes de 4G, porque só a força do sinal e disponibilidade de canais são considerados como métricas de *handover* (qualidades que são medidas para dar uma indicação se um *handover* é necessário ou não). No sistema 4G vislumbrado, as seguintes novas métricas foram propostas para uso em conjunto com medidas de força do sinal: **tipo de serviço** - os diferentes tipos de serviços exigem várias combinações de confiabilidade, latência e taxa de dados; **custo monetário** - o custo é sempre uma questão importante para os usuários, já que redes diferentes podem empregar estratégias de faturamento diferentes que podem afetar a escolha do *handover* pelo usuário; **condições da rede** - parâmetros relacionados, tais como tráfego, largura de banda disponível, a latência de rede e congestionamento (perda de pacotes) podem precisar ser considerados para o uso eficaz da rede; o uso de informações da rede na decisão de *handover* também pode ser útil para balanceamento de carga através de diferentes redes, possivelmente aliviando o congestionamento em determinados sistemas; **desempenho do sistema** - para garantir o desempenho do sistema, uma variedade de parâmetros podem ser utilizados na decisão de *handover*, tais como as características de propagação do canal, *path loss*, a interferência entre canais, *signal-to-noise ratio (SNR)*, e *bit error rate (BER)*; A **energia da bateria** pode ser outro fator crucial para determinados usuários; quando o nível da bateria está fraca, o usuário pode escolher mudar para uma rede com baixo consumo de energia, como uma rede *ad hoc Bluetooth*; **condições da MS** – As condições da estação móvel (MS) incluem fatores dinâmicos, como velocidade, padrão de movimentação, mudanças de histórico e informações sobre a localização e as **preferências do usuário** - as preferências do usuário podem ser utilizadas para atender a pedidos especiais de um tipo de sistema sobre outro. A solução provida pelo SUTIL está em conformidade com as principais conclusões do trabalho em [MCNAIR 2004], já que leva fortemente em conta métricas como o **tipo de serviço**,

condições de rede, energia da bateria, condições da MS e preferências do usuário mencionados no artigo. Portanto, podemos dizer que o SUTIL está pronto para enfrentar os principais desafios dos sistemas 4G. A métrica **custo monetário** não é considerada no SUTIL na versão atual, mas pode facilmente ser incluída na composição das preferências de nossos usuários no futuro.

Finalmente, o trabalho em [ZHU 2006] apresenta um algoritmo de decisão de *handover*, que faz uso de uma função de custo para selecionar, por uma MS específica, a melhor rede candidata, tendo em conta os múltiplos serviços em uso ao mesmo tempo (por um único usuário). O SUTIL apresenta os seguintes diferenciais em comparação com [ZHU 2006]: (i) procura satisfazer simultaneamente a todos serviços de usuários candidatos a *handover* de uma BS específica, (ii) faz uso de técnicas baseadas em Função de Utilidade e Programação Linear Inteira durante a fase de seleção, (iii) busca a eficiência no uso dos recursos da rede para que os limites das redes selecionadas não sejam ultrapassados e (iv) tenta evitar a ocorrência de sincronização *handover*.

2.6.2 Trabalhos relacionados ao ARPRAP

Uma questão importante quando se fala de QoS em presença de *handover* é a realização do *handover* de forma transparente para o usuário e sem interrupção da comunicação. Porém, nas redes WIMAX, conforme especificado no padrão, no HO original (*Hard Handover*) uma estação móvel não é capaz de transmitir ou receber dados antes do processo de *handover* estar concluído. Em [CHOI 2005], um algoritmo de *Fast Handover* é proposto de forma que cada estação móvel pode receber em tempo real o serviço de downlink da BS-alvo durante o processo de *handover*. Os autores modificaram o *Hard Handover* para introduzir uma nova mensagem de gerência para receber dados do *downlink* durante o processo de *handover*. Como resultado, o algoritmo proposto pode ignorar o tempo de processamento de re-entrada na rede durante o *handover* para serviço de downlink. Conseqüentemente, há uma redução, tanto na probabilidade de atraso de transmissão de dados quanto na probabilidade de perda de pacotes. No ARPRAP, utiliza-se o *Fast Handover FBSS* [IEEE Std 802.16e - 2005] que também faz uma transição suave do usuário, sem perda de conexão e onde o usuário não pára de transmitir e receber dados durante o processo de *handover*. O grande diferencial é que o FBSS é aprimorado no ARPRAP com o uso do perfil do usuário para a escolha mais acurada de BSs-Alvo e com a alocação prévia de recursos para os usuários.

Uma arquitetura é proposta em [KIM 2005] para a integração das redes 3G e 802.16, enfocando aspectos de QoS, enquanto proporciona total mobilidade das estações. Para tal, usa-se um modelo fracamente acoplado (*Loosely-coupled*) onde a integração acontece via Internet e é baseada em protocolo IP. Essa arquitetura incorpora procedimentos para: (i) ativar sessões de QoS que mapeiam os requisitos da aplicação em parâmetros de QoS de acordo com a rede específica, (ii) fazer o mapeamento entre diferentes classificações de QoS específicas de rede, (iii) prover suporte a QoS de camada de rede usando *DiffServ*; e (iv) prover suporte a QoS de camada de sessão através do protocolo SIP para atender serviços em tempo real, evitando o atraso induzido pelo roteamento triangular realizado em IP móvel. O objetivo principal em [KIM 2005] é oferecer QoS adequado na rede de destino de acordo com o perfil de QoS do usuário e requisitos de aplicação. O trabalho em [KIM 2005] e o ARPRAP têm como foco o padrão IEEE 802.16 e a QoS, porém em [KIM 2005] o objetivo é a integração entre as redes 802.16 e as redes 3G UMTS. Já o ARPRAP promove a melhoria da QoS em presença de *handover*, sendo complementar ao trabalho apresentado em [KIM 2005].

Em [CHANG 2005], um esquema de gerenciamento hierárquico de mobilidade para redes WiMAX foi proposto. A proposta de IP móvel do IETF foi modificada de forma a minimizar os seus inconvenientes em uma rede metropolitana, oferecendo *Fast Handover*, micro-mobilidade (o usuário móvel se move dentro de um domínio) e *handover vertical*. O trabalho proposto em [CHANG 2005] e o ARPRAP se assemelham no que tange a preocupação com a mobilidade, porém o foco em [CHANG 2005] é otimizar a mobilidade na camada 3, com soluções de IP móvel, enquanto as soluções do ARPRAP estão baseadas em otimizações de *handover* na camada 2 das redes WiMAX, o que as torna soluções complementares.

Em [PANDEY 2004] é apresentada uma proposta em que os recursos são alocados de forma dinâmica e eficiente para os usuários classificados como *Premium* nas células alvo previstas para o próximo *handover* da rede celular, atingindo um alto grau de utilização de recursos. Usuários *Premium* apresentam maiores requisitos de QoS, por exemplo, grande largura de banda, e participam de um Service Level Agreement (SLA) com a provedora de acesso. A principal contribuição da reserva dinâmica adotada do trabalho em [PANDEY 2004] é o algoritmo PARMA, que é baseado na previsão de caminhos e de serviços, considerando o perfil do usuário. A proposta do ARPRAP é baseada no trabalho [PANDEY 2004] com as seguintes diferenciações: (i) um cenário integrando diferentes redes WiMAX é

utilizado, (ii) o perfil do usuário também inclui variáveis temporais e espaciais, (iii) um novo algoritmo para a alocação de canal é proposto (iv) o conjunto de células previsto para a passagem seguinte é ainda mais refinado, tendo em conta o comportamento do usuário, bem como serviços e contextos de aplicação.

2.7 Considerações finais do Capítulo

Neste Capítulo, foram apresentados os conceitos básicos necessários à fundamentação dos mecanismos ARPRAP e SUTIL propostos. Foram descritas de forma resumida as principais tecnologias de rede sem fio, em especial as de alcance metropolitano com ênfase no padrão IEEE 802.16e e o conceito de Redes de Próxima Geração (NGN). Por estarem intimamente ligadas aos mecanismos propostos, os diversos tipos de *handover* e suas características foram introduzidos, além dos desafios impostos pela transição das redes atuais para NGN. Em seguida, os conceitos de QoS em redes foram abordados, destacando-se a QoS em presença de *handover*, foco deste trabalho.

3 Seleção de Redes baseada em Função de Utilidade

Neste capítulo é descrito o SUTIL, o mecanismo para seleção de redes em ambientes NGN proposto neste trabalho. SUTIL é um mecanismo para seleção de redes executado em cada BS de um sistema NGN. O SUTIL pode ser fornecido como parte do serviço de gerenciamento de mobilidade [JIA 2006] de dispositivos móveis de um Sistema de Middleware de Próxima Geração. A Gerência de Mobilidade tipicamente engloba a Gerência de Localização, que está fora do escopo desse trabalho, e a gerência de *Handover* (HO). A Gerência de *Handover* torna possível que uma estação móvel mantenha a conexão com a rede enquanto ocorre o processo de *handover*, isto é, a transição da estação móvel de uma estação base para outra. A fim de permitir uma completa integração entre redes heterogêneas NGN, além do serviço de gerência de mobilidade são necessários módulos e serviços adicionais. O SUTIL deve interagir com esses módulos adicionais e serviços, trocando informações com eles sempre que for necessário. Todos os resultados do mecanismo de decisão SUTIL são encaminhados para o módulo responsável pela execução do processo de *handover*. No trabalho descrito em [AKYILDIZ 2005] é proposta uma solução que fornece uma completa integração entre redes heterogêneas. A interação entre o SUTIL e outros serviços NGN está fora do escopo deste trabalho.

Este capítulo está organizado da seguinte forma: Na Seção 3.1 uma visão geral do mecanismo e a formulação do problema por ele endereçado são apresentadas. Na Seção 3.2 é descrita com detalhes a operação do SUTIL. A Seção 3.3 finaliza o capítulo com as considerações finais.

3.1 Visão geral e formulação do problema

O processo de seleção no SUTIL é subdividido em três etapas. Considerando que existe um conjunto de MSs prontas para fazer um *handover* e um conjuntos de BSs candidatas para receber estas MSs, a *primeira etapa* engloba duas atividades: (i) verificar se todas as redes candidatas são capazes de atender aos requisitos de todos os serviços em uso em cada MS do conjunto; e (ii) construir serviços que representem todos os serviços em uso por um dado usuário. Na *segunda etapa* é atribuída uma utilidade a cada uma das redes pré-selecionadas na etapa anterior em relação a cada um dos usuários candidatos a efetuar o *handover* e localizados em uma dada BS. Uma função utilidade é definida de modo a informar o "quão promissora" é a conexão de um dado usuário com uma dada rede alvo. Na

terceira etapa busca-se por uma solução, fazendo uso de Programação Linear Inteira (PLI), que maximize a função utilidade de forma a estabelecer o máximo de conexões possíveis entre os usuários e as respectivas redes alvo preferenciais, garantindo desta maneira um alto grau de satisfação dos usuários e o uso eficiente dos recursos das redes. Cada uma das três etapas do SUTIL serão detalhadamente descritas na Seção 3.3.

Para o correto funcionamento do SUTIL, algumas premissas se fazem necessárias. Inicialmente, como mencionado no Capítulo 1, é preciso que as BSs que são candidatas ao processo de *handover* possam cooperar entre si de forma a trocar informações de alto nível, como, por exemplo, as capacidades das BSs.

Outro requisito importante é que, em função da abordagem global do SUTIL se faz necessário haver um *pool* de MSs simultaneamente prontas para fazer um *handover* e um conjunto de BSs candidatas para receber estas MSs. A otimização de alocação de recursos pelo SUTIL depende de considerar simultaneamente vários usuários. No entanto, existe um paradoxo a ser investigado em relação ao tempo que o SUTIL deve aguardar a formação do *pool* de MSs em transição. Valores elevados de tempo de espera podem gerar atrasos do início do processo de *handover*. Portanto, no SUTIL se faz necessário um mecanismo para calcular o limite máximo no que diz respeito ao tempo de espera para a formação de *pool*. Esse mecanismo consiste em calcular um valor médio de um conjunto de valores de tempo de início de *handover* vertical, fornecido por um módulo externo ao SUTIL. Como o SUTIL só é responsável pela seleção de rede e não pelo desencadeamento do processo de *handover*, os valores de tempo relacionados ao início do processo de *handover* são obtidos a partir de outros módulos existentes em uma rede sem fio. O SUTIL só calcula a média. Um exemplo desse tipo de módulo de em uma Rede Sem Fio de Próxima Geração (NGN) é o *seamless roaming module* descrito em [MOHANTY 2006]. O *seamless roaming module* usa a detecção de camada física e MAC para determinar o tempo adequado para a iniciação do processo de *handover* vertical. Em um cenário real, o atraso de sinalização do *handover* vertical varia de alguns segundos a várias dezenas de segundos, dependendo de vários fatores, por exemplo, o nível de tráfego na rede backbone, qualidade da ligação da rede sem fios, e a distância entre o usuário e sua rede de origem. Assim, os autores em [MOHANTY 2006] argumentam que os protocolos baseados em um atraso de sinalização de *handover* vertical fixo têm um fraco desempenho. Para solucionar tal deficiência, os autores propõem o uso de um limiar dinâmico de *Received Signal Strength (RSS)* e de um sistema para prever o atraso de sinalização do *handover* vertical com antecedência. Assim, assumimos que SUTIL interage com o *seamless*

roaming module e armazena informações sobre o tempo de *handover* para calcular um valor médio a ser utilizado para a formação do pool.

Com relação ao processo de seleção da rede propriamente dito, o SUTIL considera vários critérios para a tomada de decisão de qual seria a “melhor rede” quando da necessidade de *handover*. Esses critérios dizem respeito ao usuário, a rede e ao dispositivo. Por exemplo, o critério **prioridade de serviço** indica a prioridade de um serviço/aplicativo específico em relação a outros serviços executados no dispositivo de um usuário. Já o critério **preferência de energia do usuário** é o grau de preferência que um usuário tem em relação aos requisitos de consumo de energia, quando uma rede é selecionada. O critério **energia máxima** informa o valor máximo de energia que a estação pode consumir ao usar uma rede.

Os critérios para prioridade de serviço englobam a banda mínima, a taxa máxima de perda de pacotes e o atraso máximo requeridos/tolerados por uma aplicação específica. Dentre os valores possíveis que um requisito de um serviço pode assumir, denomina-se ‘melhor valor’ aquele valor que melhor atende ao dado serviço. Por exemplo, para o requisito de perda, quanto menor for o valor da perda mais próximo esse valor está deste ‘melhor valor’. O mesmo princípio aplica-se ao atraso, onde o ‘melhor valor’ é o menor valor de atraso sugerido para o requisito. Em contrapartida, para o requisito de banda quanto maior for o valor da banda mais próximo esse valor está deste ‘melhor valor’. Adicionalmente, denomina-se ‘pior valor’ do requisito aquele valor que pior atende ao dado serviço. Para o requisito de perda (ou atraso), observa-se que ao aumentar o valor da perda (ou atraso) mais próximo esse valor está deste ‘pior valor’. Em contrapartida, para o requisito de banda quanto menor for o valor da banda mais próximo esse valor está deste ‘pior valor’. Quanto ao contexto de rede, os seguintes parâmetros (ou critérios) são considerados: a banda disponível, o atraso fim a fim e a perda de pacotes. A energia residual, localização geográfica, velocidade e padrão de movimento (a pé ou locomotivo) e SNR (Signal to Noise Ratio) são parâmetros da MS que são também considerados.

O problema de seleção de redes tratado pelo SUTIL pode ser formulado conforme descrito em seguida. Seja uma topologia composta por um conjunto $J = \{1, \dots, M\}$ de estações base (BSs) e um conjunto $I = \{1, \dots, N\}$ de estações móveis (MSs) candidatos ao processo de *handover*, os quais estão conectados a uma mesma estação base $j \in J$. Cada uma das estações móveis $i \in I$ possui s serviços de um conjunto S de serviços sendo usados simultaneamente.

O SUTIL é executado em todas as BSs do conjunto J e é ativado pelo processo de gerência de *handover* quando um novo *handover* é necessário.

3.2 Arquitetura do SUTIL

A arquitetura proposta para o SUTIL possui componentes ilustrados na Figura14 que implementam as fases de operação do SUTIL. Além disso, a arquitetura prevê interações com serviços externos ao SUTIL como o de localização e o módulo de execução de *handover* (ver Seção 3.1).

Os componentes que integram o SUTIL são:

- Avaliação de Serviços;
- Cálculo de Utilidades;
- Seleção de Redes.

O primeiro componente – **Avaliação de Serviços** – é responsável pela primeira etapa de operação do SUTIL, a qual consiste em eliminar redes incapazes de atender aos requisitos de pelo menos uma MS do *pool* e construir o serviço representativo dos serviços em uso em cada MS candidata a *handover*. Esta etapa será detalhada na Seção 3.3.1.

O segundo componente – **Cálculo de Utilidades** – é responsável pela segunda etapa de operação do SUTIL que consiste em calcular a utilidade que cada BS candidata tem para cada uma das MSs candidatas a *handover*. Etapa detalhada na Seção 3.3.2.

O terceiro componente – **Seleção de Redes** – está a cargo da terceira etapa de operação do SUTIL que consiste em selecionar as melhores redes possíveis para MSs candidatas ao processo de *handover* e localizadas em uma mesma estação base, segundo as utilidades das redes obtidas do segundo componente . Etapa detalhada na Seção 3.3.3.

A Gerência de *Handover* é o componente responsável pela execução do processo de *handover* para o qual o mecanismo de seleção SUTIL repassa os resultados. A Gerência de Localização é usada pelo SUTIL para verificar o posicionamento das MSs e BSs e assim selecionar BSs coerentes com o deslocamento do usuário. Ambos estão fora do escopo do SUTIL.

O *Seamless Roaming Module* [MOHANTY 2006] é um exemplo de módulo capaz de determinar o tempo adequado para a iniciação do processo de *handover* vertical usado pelo SUTIL para determinar o tempo usado para a formação do *pool* de MSs simultâneamente candidatas a *handover*.

Estes componentes se interconectam da seguinte forma: (i) a *Gerência de Handover* aciona o SUTIL pela interface *SetPool* passando MSs e BSs candidatas a *handover*; (ii) a *Avaliação de Serviços* usa a média de tempo de início de handover calculado a partir de informações obtidas pela interface *GetHOTime* oferecida pelo *Seamless Roaming Module*, para limitar o tempo de formação do *pool*; utiliza o serviço oferecido pela *Gerência de Localização* para obter o posicionamento das MSs e BSs pela interface *GetCoordenadas*; filtra as redes e monta os serviços representativos (SR), e finalmente passa esses dados para o componente *Cálculo de Utilidades* pela interface *SetSR* que esse oferece; (iii) o *Cálculo de Utilidades* então calcula as utilidades das redes para as MSs do *pool* e repassa a *matriz utilidade* via interface *SetUtilidade* para o componente *Seleção de Redes*; e (iv) o componente *Seleção de Redes* seleciona as redes para as MSs de forma maximizar as utilidades. O resultado desta seleção é devolvido ao componente *Gerência de Handover* através da interface *SetSeleçãoSUTIL* para a execução dos handovers sugeridos pelo SUTIL.

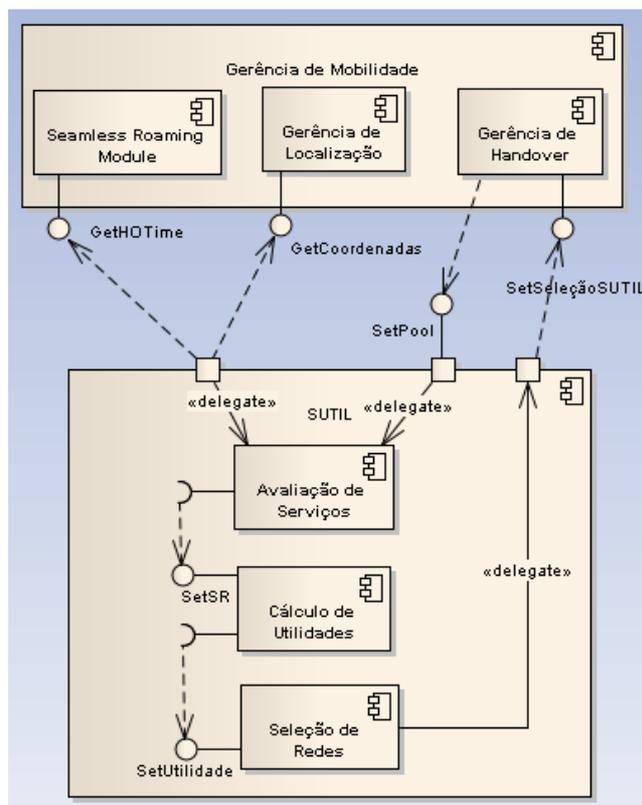


Figura 14. Diagrama de componentes do SUTIL

3.3 Operação do SUTIL

Os procedimentos que descrevem o funcionamento do SUTIL são agrupados em 3 etapas, detalhadas nas próximas seções:

- eliminar redes inadequadas e construir Serviços Representativos;
- calcular Função de Utilidade; e
- selecionar Redes.

Esses procedimentos são realizados apenas para as MSs que estão em movimento, informação essa obtida pela BS através de consultas regulares feitas à Gerência de Localização. O SUTIL ao ser ativado leva em consideração todas as estações móveis simultaneamente candidatas ao processo de *handover*.

3.3.1 Eliminar Redes e Construir Serviço Representativo

A primeira etapa de funcionamento do SUTIL é responsável por eliminar as redes candidatas inadequadas ao processo de *handover* para posterior seleção. Em outras palavras, o SUTIL elimina redes que são candidatas ao processo de *handover*, mas não preencheram os requisitos mínimos de serviços em uso pelo conjunto de MSs em transição. Para tal, o SUTIL faz uso de um filtro simples. Este filtro compara os requisitos de largura de banda, perda de pacotes e atraso dos serviços executados em cada MS em transição com as capacidades de todas as redes candidatas. Sempre que uma rede não é capaz de satisfazer os requisitos dos serviços em execução em uma das MSs em transição, o mecanismo SUTIL atribui esta rede para a primeira MS entre as MSs que são menos exigentes do conjunto de MSs em transição, desde que a rede seja capaz de atender seus requisitos. Em tais casos, tanto a MS menos exigente e a rede que lhe foi atribuída são retiradas das demais etapas do processo de seleção SUTIL já que a escolha MS-BS já foi realizada. A rede candidata é completamente removida do processo de *handover* apenas quando ela não é capaz satisfazer aos requisitos dos serviços em execução de nenhuma MS do conjunto de MSs em transição. Após esta etapa, cada MS pode ser atendida por qualquer uma das redes filtradas. Assume-se que as redes depois de filtradas formam um subconjunto $J' \subseteq J$.

Em seguida, é construído um *Serviço Representativo (SR)* para cada MS candidata a efetuar o processo de *handover*, considerando que há s serviços simultaneamente em uso em cada uma das N MSs. Cabe ressaltar que no presente trabalho considera-se que cada usuário possui apenas um dispositivo móvel em uso.

Define-se um serviço representativo (SR) para uma MS i específica como sendo o serviço hipotético usado para representar todos os serviços em execução nesta MS i . Mais especificamente, um SR é representado por um conjunto de *Requisitos Representativos (RRs)* onde cada RR é calculado de forma a atender ao requisito em questão para todos os s serviços em uso na MS i . Cada RR caracteriza-se pelos valores ‘pior valor’ e ‘melhor valor’ de um intervalo de valores possíveis que é determinado para o RR em questão. Para construção de RRs, duas estratégias são adotadas pelo SUTIL: (i) estratégia de soma de valores, e (ii) a estratégia do valor extremo.

A *estratégia de soma de valores* é empregada para obter, para a MS i , os limites denominados $BRRLowLim_i$ (Equação I) e $BRRUpLim_i$ (Equação II), esta estratégia consiste em adicionar os valores limites correspondentes aos serviços em uso na MS i , o limite inferior do **RR de banda** é obtido através da soma dos valores do requisito $BRRLowLim_i$ correspondentes, relacionados aos limites mínimos de banda, de cada um dos serviços em uso na MS i . Da mesma forma, o limite superior do **RR de banda** é obtido através da soma dos valores do requisito $BRRUpLim_i$ correspondentes, relacionados aos limites da banda máximas, de cada um dos serviços em uso na MS i . Estes cálculos são expressos nas Equações I e II, onde $BRLowLim_{i,k}$ and $BRUpLim_{i,k}$ são, respectivamente, os limites inferiores e superiores de requisito de banda para o serviço k em uso na MS i . (Observando que existem s serviços em uso.)

$$BRRLowLim_i = \sum_{k=1}^s BRLowLim_{i,k} \quad I$$

e

$$BRRUpLim_i = \sum_{k=1}^s BRUpLim_{i,k} \quad II$$

A *estratégia do valor extremo*, por outro lado, é adequada para determinar os limites de **RR de Perda** e **RR de atraso**. Consiste em pegar os valores extremos entre os requisitos correspondentes aos serviços em uso na MS i .

O *limite inferior* (*‘melhor valor’*) do **RR de perda**, é obtido ao adotar-se o maior valor dentre os limites inferiores do requisito de perda, para o serviço k em uso na MS i . Enquanto o *limite superior* (*‘pior valor’*) do **RR de perda** é obtido ao adotar-se o menor valor entre os valores superiores do requisito de perda para o serviço k em uso na MS i . Em outras palavras, o limite inferior (*‘melhor valor’*) e o limite superior (*‘pior valor’*) do RR de perda, para a MS i , denominados $PLRRLowLim_i$ (Equação III) e $PLRRUpLim_i$ (Equação IV) são calculados respectivamente na Equação III e na Equação IV.

$$PLRRLowLim_i = \max_{k=1}^s \{ PLRRLowLim_{i,k} \} \quad III$$

e

$$PLRRUpLim_i = \min_{k=1}^s \{ PLRRUpLim_{i,k} \}, \quad IV$$

onde $PLRRLowLim_i$ e $PLRRUpLim_i$ são, respectivamente, o limite inferior e superior do requisito de perda para o serviço k em uso na MS i .

Analogamente, o *limite inferior* (*‘melhor valor’*) do **RR de atraso**, denominado $DRRLowLim_i$ na Equação V, é obtido ao adotar-se o maior valor entre os limites inferiores do requisito de atraso, para o serviço k em uso na MS i . Já o *limite superior* (*‘pior valor’*) do **RR de atraso**, denominado $DRRUpLim_i$ na Equação VI, é obtido ao adotar-se o menor valor entre os valores de limites superiores do requisito de atraso para o serviço k em uso na MS i .

$$DRRLowLim_i = \max_{k=1}^s \{ DRRLowLim_{i,k} \} \quad V$$

e

$$DRRUpLim_i = \min_{k=1}^s \{ DRRUpLim_{i,k} \}, \quad VI$$

Na Tabela 4 [KANG 2006] são ilustrados os valores típicos de RR. A coluna ‘voz e vídeo’, por exemplo, é um RR determinado com os valores das colunas ‘voz’ e ‘fluxo de vídeo’.

Tabela 4. Exemplos de Requisitos e seus RR [KANG 2006]

Requisitos Representativos		voz	fluxo de vídeo	fluxo de dados	voz e vídeo	voz e dados	vídeo,voz e dados	vídeo e dados
Banda(Kbps)	$BRRUpLim_i$	64	128	500	192	564	692	628
	$BRRLowLim_i$	9	30	128	39	137	167	158
Atraso(ms)	$DRRLowLim_i$	1	5	5	5	5	5	5
	$DRRUpLim_i$	150	150	120	150	120	120	120
Perda(%)	$PLRRLowLim_i$	2	1	2	2	2	2	2
	$PLRRUpLim_i$	8	3	8	3	8	3	3

A fim de levar em conta a mobilidade das MSs, o SUTIL não considera em suas decisões, as redes candidatas localizadas na direção oposta ao deslocamento das MSs. Para realizar essa estratégia, todo o conjunto de BSs candidatas que são vizinhas da BS atual é dividido em dois subgrupos nomeados ÁREA da FRENTE (AF) e ÁREA de TRÁS (AT). A área AF engloba BSs que estão localizados em coordenadas X e/ou Y maiores que as da BS atual. Por outro lado, a área AT abrange BSs com coordenadas X e/ou Y inferiores as da BS atual. Portanto, um conjunto de MSs se deslocando para a frente em relação a sua BS atual (isto é, com as coordenadas X e/ou Y aumentando), denominado MS-F, terá como BSs candidatas àquelas designadas para a área AF, enquanto que um conjunto de MSs se deslocando na direção contrária quanto à sua BS atual, denominado MS-T, terá como candidatas aquelas BSs atribuídas à área AT. Considerando-se esta estratégia, o SUTIL será executado duas vezes, uma para o grupo denominado MS-F e um para o MS-T, a fim de utilizar apenas o subconjunto de BSs candidatas localizadas no sentido supostamente seguido por cada MS pertencente a cada grupo. Apesar da versão atual do SUTIL considerar apenas duas áreas, é importante ressaltar que pode-se dividir o conjunto de BSs candidatas que são vizinhas da BS atual em mais divisões. Inicialmente optou-se por dividir o conjunto de BSs em apenas duas áreas, a fim de analisar conjuntos maiores de MSs (mais usuários). Além disso, ao dividir os conjuntos de BSs em mais de duas áreas, por exemplo, norte, sul, leste, oeste, é possível que uma mesma BS esteja presente em mais de uma área, como nordeste. Portanto, estas BSs podem servir simultaneamente a mais de um *pool*, e estar atribuída a mais de um sub-conjunto de BSs. Em tais casos a solução torna-se ligeiramente mais complexa.

3.3.2 Calcular os valores da Função Utilidade

Na segunda etapa do SUTIL, para cada MS i candidata a efetuar o processo de *handover*, é calculada a utilidade de cada uma das redes $j \in J'$ (do ponto de vista deste usuário). Para o cálculo da Função Utilidade F_{ij} , adotou-se uma abordagem na qual, para uma MS i específica, é dada prioridade às redes que tenham maior **relevância** para os serviços em uso na MS i e que, ao mesmo tempo, proporcionem a maior **energia residual**. Considerando todos os serviços em uso na MS i , a **relevância** R_{ij} de uma rede $j \in J'$ depende não somente dos valores correntes de banda disponível da rede, a perda de pacotes da rede e o atraso fim-a-fim da rede, mas também dos RRs de banda, atraso e perda de pacotes formando o SR da MS i . A relevância R_{ij} (dada pela Equação I) de uma rede j é obtida em quatro passos. Primeiro, relaciona-se a banda de rede disponível, denominada $NetB_j$, com os valores $BRRUpLim_i$ e $BRRLowLim_i$ segundo a Equação VII.

$$B_{ij} = (\min\{NetB_j, BRRUpLim_i\} - BRRLowLim_i) / (BRRUpLim_i - BRRLowLim_i), \quad VII$$

Observe que $0 \leq B_{ij} \leq 1$. Quanto mais próximo B_{ij} é de 1, mais este valor contribui para a determinação da relevância R_{ij} . Por isso, B_{ij} pode ser visto como a “aptidão” da rede j para a MS i no que diz respeito a banda. No Segundo passo, relaciona-se a perda de pacotes da rede, denominada $NetPL_j$ com $PLRRUpLim_i$ e $PLRRLowLim_i$ segundo a equação VIII.

$$P_{ij} = (PLRRUpLim_i - \max\{NetPL_j, PLRRLowLim_i\}) / (PLRRUpLim_i - PLRRLowLim_i). \quad VIII$$

Em seguida, relaciona-se o atraso fim-a-fim da rede, denominado $NetD_j$ (Equação IX), com $DRRUpLim_i$ e $DRRLowLim_i$:

$$D_{ij} = (DRRUpLim_i - \max\{NetD_j, DRRLowLim_i\}) / (DRRUpLim_i - DRRLowLim_i), \quad IX$$

onde P_{ij} , D_{ij} significa a aptidão da rede j para a MS i no que diz respeito a perda de pacotes e atraso, respectivamente. Novamente observe que $0 \leq P_{ij} \leq 1$ e $0 \leq D_{ij} \leq 1$. Finalmente, para determinar a relevância R_{ij} (Equação X), combina-se os valores B_{ij} , P_{ij} e D_{ij} usando três coeficientes de calibração λ, ϕ, δ da seguinte forma:

$$R_{ij} = \lambda B_{ij} + \phi P_{ij} + \delta D_{ij}, \quad X$$

onde $\lambda + \phi + \delta = 1$. A equação acima implica $0 \leq R_{ij} \leq 1$; quanto mais próximo R_{ij} é de 1, mais relevante a rede j é para a MS i . Em um cenário onde a banda, a perda e o atraso são priorizados uniformemente, tem-se $\lambda = \phi = \delta = 1/3$. Contudo, muitos cenários plausíveis

podem ser modelados calibrando convenientemente λ, φ, δ . Por exemplo, para garantir para uma MS específica i que o atraso é mais importante que a banda e a perda de pacotes, pode-se obter fazendo $\delta > \lambda$ e $\delta > \varphi$.

Quanto à **energia residual** U_i de um dado usuário i , busca-se sempre por uma tecnologia de rede que gere o menor consumo de energia para a MS em questão. O tipo de interface de rede utilizada pela MS i é um dos fatores que afetam o custo de energia W_{ij} quando uma transição (*handover*) para a rede j é efetuada. A energia residual U_i é definida como sendo a quantidade de energia atual que um dispositivo MS i possui em um dado instante de tempo. Assim, o termo $U_i - W_{ij}$ denota a energia final da MS i , no caso dela efetuar a transição para rede j (energia inicial U_i menos o consumo de energia W_{ij} da MS i no processo de *handover*). Em suma, busca-se escolher uma tecnologia de rede para a qual o termo $U_i - W_{ij}$ seja maximizado, valor este normalizado por U_i . Logo, para uma MS i específica, a Função utilidade F_{ij} (equação XI) de uma rede j pode ser representada por:

$$F_{ij} = \alpha R_{ij} + \beta (U_i - W_{ij}) / U_i, \quad XI$$

onde α, β são também coeficientes de calibração satisfazendo a relação $\alpha + \beta = 1$. O valor F_{ij} significa a utilidade da rede j para a MS i . Como para R_{ij} , tem-se $0 \leq F_{ij} \leq 1$.

É interessante notar que o SUTIL pode priorizar a relevância da rede frente ao serviço representativo do dado usuário i , ou o tempo de vida do dispositivo obtido ao selecionar uma tecnologia específica de rede. Neste trabalho, o par (α, β) constitui um **perfil de QoS** da rede. Existem 3 possíveis perfis de QoS:

- (i) baseado no desempenho, que prioriza a relevância da rede, buscando atender os requisitos do SR (serviço representativo) do dado usuário candidato ao *handover*. Nesse caso, $\alpha = 1, \beta = 0$.
- (ii) baseado no tempo de vida do dispositivo móvel, que prioriza o consumo de energia do mesmo. Nesse caso, $\alpha = 0, \beta = 1$.
- (iii) baseado na razão desempenho/tempo de vida, que balanceia o tempo de vida do dispositivo e o parâmetro de desempenho, buscando a melhor razão custo/benefício entre os dois parâmetros. Os valores de α e β devem ser ajustados de acordo.

Como resultado final desta etapa, para cada par (i,j) representando uma MS i (candidata ao *handover*) e uma rede filtrada $j \in J'$, a utilidade F_{ij} é determinada.

3.3.3 Selecionar Redes

A principal meta da terceira etapa do mecanismo de seleção é selecionar as melhores redes possíveis para MSs candidatas ao processo de *handover* e localizados em uma mesma estação base, segundo as utilidades das redes obtidas na etapa anterior. Em outras palavras, busca-se uma solução que maximize a função utilidade F_{ij} de forma a estabelecer o máximo de conexões possíveis entre os usuários i e as redes do conjunto J' , garantindo o uso eficiente dos recursos das redes. Com esse propósito, define-se uma função linear inteira como se segue:

$$\text{Maximizar } \sum_{i=1}^N \sum_{j \in J'} (z_{ij} \cdot (F_{ij} - Y_{ij})) \quad XII$$

Sujeito a três restrições:

$$\sum_{j \in J'} z_{ij} = 1, \text{ para cada } i \in \{1, \dots, N\} \quad (1)$$

$$\sum_{i=1}^N z_{ij} \cdot BRRUpLim_i \leq NetB_j, \text{ para cada } j \in J' \quad (2)$$

$$\sum_{i=1}^N z_{ij} \leq P, \text{ para cada } j \in J' \quad (3)$$

onde $z_{ij} \in \{0,1\}$, para cada par i,j tal que $i \in \{1, \dots, N\}$ e $j \in J'$

Na equação XII e nas restrições (1), (2) e (3), i é um índice para as MSs, j é um índice para as redes. A variável binária z_{ij} assume valor 1 se a MS i estabelece conexão com a rede j , caso contrário assume valor 0. O valor F_{ij} é a utilidade da rede j em relação a MS i (calculada na etapa anterior). O valor Y_{ij} é um ajuste “força-mudança”, que tem a intenção de forçar cada MS i a fazer *handover* para uma nova rede diferente da rede que esta está atualmente conectada. Por exemplo, Y_{ij} pode ser ajustado para 0,5 se j é a rede corrente que a MS i está conectada, e 0 caso contrário. Por outro lado, ajustando $Y_{ij} = 0$ para cada j , a rede corrente tem a mesma prioridade das outras. Também é possível ajustar $Y_{ij} = F_{ij}$ a rede corrente j , obtendo como efeito a proibição da seleção da rede corrente.

O objetivo da função é simplesmente maximizar o somatório das “utilidades ajustadas” das redes selecionadas sobre todas as MSs. A restrição (1) assegura que tem-se exatamente uma rede selecionada para cada MS i , e a restrição (2) que o somatório dos requisitos representativos de banda de todas as estações móveis conectadas a uma mesma rede

j não exceda $NetB_j$, para cada $j \in J'$. É importante ressaltar que $NetB_j$ refere-se à banda atualmente disponível na rede j ; portanto, o mecanismo leva em conta os usuários já conectados e aloca os novos usuários em função da banda que ainda não foi usada. A Restrição (3) garante que nenhuma rede receba mais do que P MSs a fim de evitar que uma rede escolhida como preferencial por todas as MSs receba um número de MSs muito alto. Essa restrição foi definida para lidar com o fato de que, para algumas tecnologias sem fio o atraso e a perda de pacotes depende do número de usuários ligados simultaneamente à rede. P é um valor que deve ser ajustado para satisfazer as restrições específicas de tecnologias sem fio que são sensíveis a este problema. Simulações extensivas devem ser feitas em diferentes cenários a fim de definir os valores adequados de P em cada situação. Há vários trabalhos, por exemplo [SHAABAN 2008] relatando os resultados de análise sobre o desempenho de sistemas sem fio em resposta a diferentes condições de carga e número de usuários, incluindo a avaliação do impacto destes fatores em aplicações sensíveis a QoS.

Como poderá ser visto no Capítulo 6, experimentos computacionais mostram que é possível usar métodos exatos para encontrar a solução usando um programa linear inteiro, desde que o número de variáveis seja tratável nos cenários típicos.

3.4 Considerações finais do capítulo

Neste capítulo foi apresentada detalhadamente a descrição do mecanismo de seleção proposto, incluindo os passos de operação seguidos pelo mecanismo: eliminar redes inadequadas e construir Serviços Representativos, calcular Função de Utilidade e selecionar Redes. Adicionalmente, foi apresentada uma visão geral do problema de selecionar redes levando em consideração MSs simultaneamente prontas para fazer um *handover* e um conjunto de BSs candidatas para receber estas MSs.

O próximo capítulo apresenta a proposta de um algoritmo de Reserva Pró-ativa de Recursos Assistida por Perfil (ARPRAP) para solucionar o problema de manter a mesma qualidade de serviço para os usuários móveis quando eles se deslocam entre redes homogêneas, no caso redes WiMAX.

4 Reserva Pró-ativa de Recursos Assistida por Perfil

Nesse capítulo é apresentada uma proposta de um algoritmo de Reserva Pró-ativa de Recursos Assistida por Perfil (ARPRAP) para solucionar o problema de manter a mesma qualidade de serviço para os usuários móveis quando eles se deslocam entre redes homogêneas, no caso redes WiMAX. Para tal, é aprimorado o algoritmo *Fast Handover* da camada 2 do IEEE 802.16e [IEEE Std 802.16 - 2005], refinando-se o processo de escolha das BS vizinhas candidatas ao *handover* com o auxílio do perfil de comportamento do usuário móvel, isto é, considerando aspectos da trajetória, requisitos da aplicação e de rede. Uma vez efetuada a escolha da BS candidata ao *handover*, são divulgados previamente os fluxos de serviços, para, a partir de uma dada distância da BS candidata ao *handover*, tais fluxos serem de fato alocados, concretizando, portanto, a operação do *Fast Handover*.

Assim, o ARPRAP explora o perfil do comportamento do usuário em movimento para maximizar a eficiência da rede e prover melhor QoS para diferentes classes de usuários em um cenário que integra redes WiMAX. Em suma, o ponto-chave dessa proposta é que o algoritmo faz uso do perfil do comportamento do usuário em tempo real para (i) refinar o conjunto de células previstas para o próximo *handover*, levando em conta não apenas o tipo de usuário envolvido (*Premium* ou *Normal*) como também o contexto do serviço e o da rede; e (ii) fazer uma reserva pró-ativa de recursos para a célula escolhida, pelo processo de refinamento, para o próximo *handover*.

Este capítulo está organizado da seguinte forma: Na Seção 4.1 uma visão geral do algoritmo e a formulação do problema por ele endereçado são apresentados. Na Seção 4.2 é descrito detalhadamente o algoritmo ARPRAP. Na Seção 4.3 são descritas as mensagens trocadas no ARPRAP. A Seção 4.4 finaliza o capítulo com as considerações finais.

4.1 Visão geral e formulação do problema

O algoritmo de Reserva Pró-ativa de Recurso Assistida por Perfil (ARPRAP) busca garantir melhor qualidade de serviço para diferentes classes de usuários em um cenário que permite a integração de redes WiMAX. Para tal, esse algoritmo faz uso do perfil do usuário em movimento para dinamicamente reservar recursos de forma pró-ativa apenas naquelas áreas previstas para o próximo *handover* da rede sem fio. O ARPRAP procura selecionar a “melhor” BS (BS alvo) a partir do conjunto de BSs vizinhas, baseado na trajetória do usuário móvel, nos requisitos da aplicação em execução no dispositivo móvel e nos requisitos de redes (recursos disponíveis na rede gerenciada pela BS Alvo). O ARPRAP faz uso: (i) do

handover do tipo *Fast* (FBSS), (ii) do protocolo SNMP (*Simple Network Management Protocol*) v2 [HARRINGTON 2002] para o gerenciamento de recursos dos dispositivos da rede sem fio, inclusive os pré-provisionados; e (iii) da MIB proposta para o IEEE 802.16 para MAC e PHY [CHOU 2004].

Através do mecanismo de predição de áreas alvos usando o perfil de comportamento do usuário móvel, é obtida uma lista mais apurada de BSs a ser usada no *Fast Handover*, resultando, conseqüentemente, em menos transmissões entre BSs e em um atendimento personalizado quanto ao quesito de QoS. Assim, as estações de usuários da categoria **Premium** teriam previamente alocados os fluxos de serviços de acordo com seu perfil de uso, permitindo um processo de *handover* rápido pela assertividade ao escolher a BS Alvo. Nas subseções a seguir, são descritas as estruturas de dados e mensagens usadas pelo algoritmo, bem como o seu procedimento detalhado.

4.1.1 Arquitetura do ARPRAP

A MIB (Management Information Base) proposta em [CHOU 2004] consiste em uma base com informações coletadas tanto de MSs como de BSs, cujas tabelas importantes para o ARPRAP são detalhadas nessa Seção. Além da MIB, o ARPRAP usa três novas estruturas de dados referentes ao usuário: Tabela de Perfil, Tabela de Trajetos e Tabela de Recursos. A **Tabela de Perfil** do usuário contém informações de perfil específicas de usuários em movimento e que estão registrados em sua rede/célula de origem (*home*) tais como: identidade do usuário, tipo do usuário (**Premium**, **Normal**), serviços acordados, ponteiro para a tabela de trajetos, ponteiro para a tabela de recursos. A Tabela de Perfil é uma estrutura armazenada na *Internet*, ficando acessível de qualquer ponto da rede. A **Tabela de Trajetos** contém uma lista de trajetos usados por um período do dia (início da manhã, manhã, almoço, início da tarde, tarde, noite) por um usuário móvel e a sua respectiva frequência de uso. Um trajeto individual, por exemplo, $\langle a_1, a_2, \dots, a_n \rangle$, é composto por uma lista de identificadores “ a_n ”, onde cada identificador “ a_n ”, denota uma área quadrada de lado igual a um inteiro “L”. Cada área é composta por várias células sobrepostas. Alguns trajetos são definidos através de acordo (SLA) entre o usuário e o provedor do serviço informando que o usuário é cliente **Premium**. Outros trajetos são inseridos na lista conforme o mesmo passa a ser utilizado pelo usuário.

Já a **Tabela de Recursos** contém uma lista de recursos, tais como a banda, usados por trajeto e por período por um usuário móvel e a respectiva frequência quanto ao uso do

recurso. Alguns recursos são definidos através de um acordo (SLA) entre o usuário e o provedor do serviço informando que o usuário é cliente *Premium*. Outros recursos são inseridos na lista conforme o mesmo passa a ser utilizado pelo usuário.

Na MIB (*Management Information Base*) *wmanIfMib*, definida por [CHOU 2004], para o padrão IEEE 802.16, são armazenadas informações coletadas pelas MSs e BSs sobre os objetos gerenciados. A *wmanIfMib* contém três tabelas que fornecem suporte à provisão do fluxo de serviço. A tabela de Fluxos pré-provisionados, *wmanIfBsProvisionedSfTable*, tem os campos MAC, *sfindex* e *Qosindex* apontando para tabelas que contêm perfis de fluxo de serviço pré-provisionado para MSs. O ARPRAP adiciona nesta tabela as necessidades de recursos (fluxos de serviços) futuras de um usuário móvel baseadas no seu perfil de utilização para cada faixa de horário e trajeto. Já a tabela *wmanIfBsServiceClassTable* contém parâmetros de QoS (*prioridade do tráfego, taxa de pico, latência máxima, jitter, atraso* e etc) associados aos fluxos de serviço. Por fim, a tabela *wmanBsClassifierRuleTable* é indexada por *sfindex* e possui as regras para que o classificador possa mapear os pacotes de *downlink* e *uplink* para o fluxo de serviço. Além dessas três tabelas, toda vez que uma estação, móvel ou não se registra em uma BS é criada uma entrada em outra tabela, denominada *wmanIfBsregisteredTable*, a qual é indexada por endereço MAC da estação. Assim, uma BS, com base no endereço MAC da estação, pode verificar as outras três tabelas e verificar se a MS tem fluxos pré-provisionados. É importante destacar que a tabela *wmanIfCmnSsChmeasurementTable* contém dados de utilização do canal pela SS/MS e que o gerenciador de perfil, embora não faça parte do escopo do ARPRAP, consulta esses dados para montar a Tabela de Perfil. A arquitetura do ARPRAP pode ser vista no diagrama de componentes ilustrado na Figura 15. Nesse diagrama de componentes podem ser observadas as interconexões entre os componentes utilizados pelo ARPRAP. Uma MS tem sua troca de mensagens de *Handover* com sua BS Âncora representada pela interface *ServHO*, que utiliza o *Fast Handover* do tipo FBSS. O componente FBSS obtém a lista de BSs Vizinhas refinadas e as indicações da BS Alvo escolhida através da interface *GetDiversitySetRefinada* do componente ARPRAP. O componente ARPRAP, por sua vez, pega as informações que necessita da MIB (interface *GetInfo*), da Gerência de Perfil (interface *GetPerfil*) e da Gerência de Localização (interface *GetCoordenadas*). Tanto a BS quanto a MS atualizam suas informações se utilizando do protocolo SNMP na MIB 802.16, serviço disponibilizado que está representado pela interface *SetInfo*.

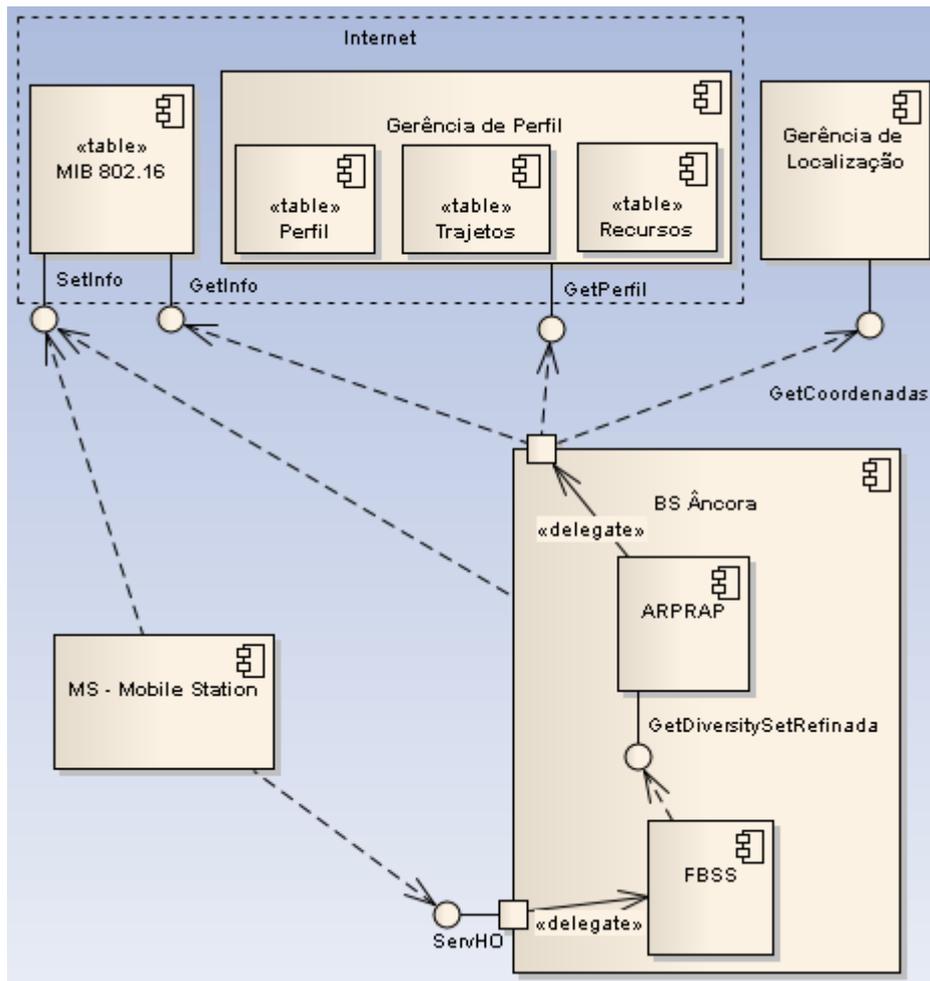


Figura 15. Diagrama de componentes do ARPRAP

4.2 Descrição do Algoritmo ARPRAP

Embora o algoritmo proposto seja genérico quanto ao número de classes de usuários, com o intuito de garantir uma descrição simplificada do mesmo, no ARPRAP serão consideradas em um acordo SLA apenas duas classes de serviços, uma para usuário *Premium* e outra para usuário *Normal*. O algoritmo proposto é executado na BS âncora e os procedimentos descritos a seguir detalham o seu funcionamento. Esses procedimentos são realizados apenas no caso da MS estar em movimento, informação essa obtida pela BS através de consultas regulares feitas ao Serviço de Localização. A Figura 16 apresenta o diagrama de seqüência ilustrando tais procedimentos.

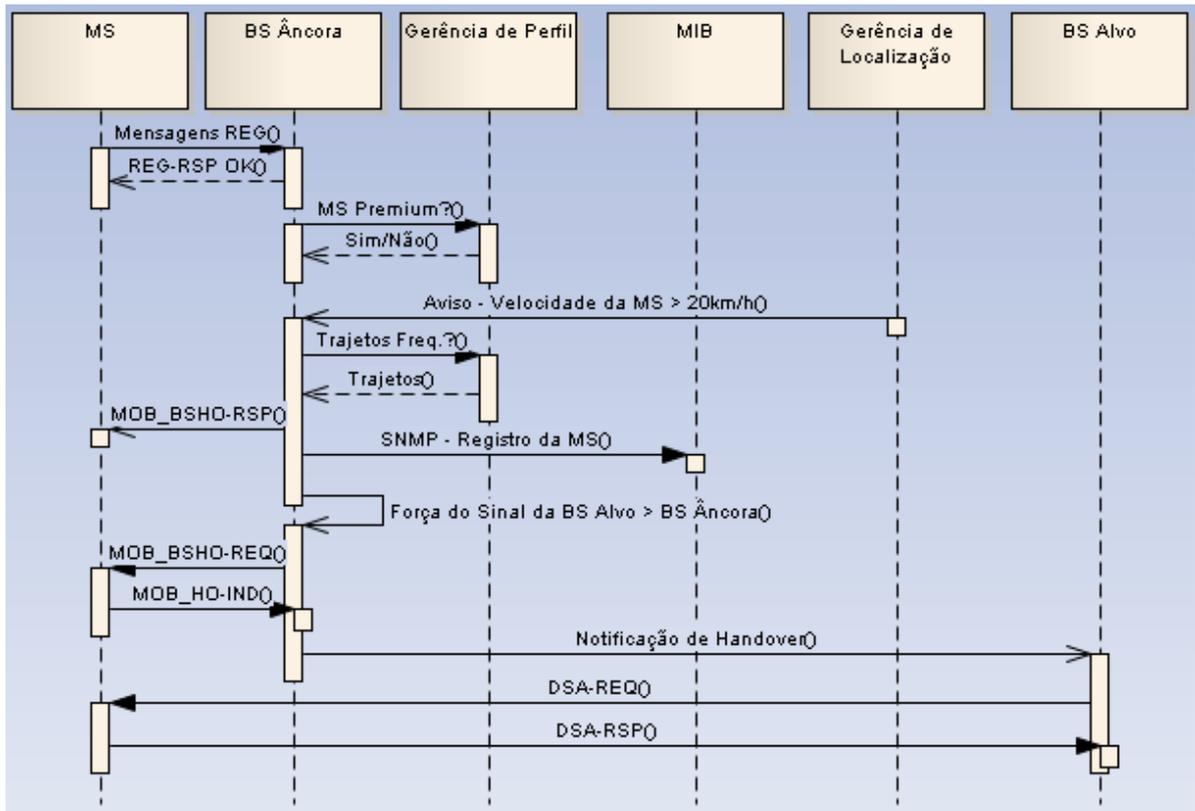


Figura 16. Diagrama de Seqüência do ARPRAP

1. Determinar se a estação móvel (MS) é do tipo cliente *Premium* ou *Normal*. Tal informação é obtida através de consulta a Tabela de Perfil do Usuário.
2. Se a MS é do tipo *Normal*, essa estação seguirá os procedimentos normais de *handover*, isto é, nenhuma reserva será feita de forma pró-ativa.
3. Se a MS é do tipo *Premium*, verificar se o usuário tem trajetos freqüentes para o horário em questão, e se nesses trajetos alguns recursos são usados comumente. Os procedimentos desse passo são descritos a seguir.
 - a. A BS âncora verifica os trajetos usados com freqüência na faixa de horário em questão, consultando a *Tabela de Trajetos* do Usuário. No ARPRAP é escolhido o trajeto mais utilizado nessa faixa de horário.
 - b. A BS âncora verifica, para o trajeto selecionado, os recursos a serem usados com freqüência na faixa de horário em questão, consultando a *Tabela de Recursos do Usuário*. De posse das informações de recursos, é alterada a tabela na MIB (*wmanIfBsServiceClassTable*) que contém os parâmetros de QoS associados ao fluxo a ser provisionado. Assim,

- em um futuro próximo, quando o processo de *handover* for iniciado, a nova BS âncora ativará estes fluxos.
- c. A BS âncora atualiza a Tabela de Fluxos pré-provisionados (*wmanIfBsProvisionedSfTable*) de forma que a mesma contenha os fluxos que atenderão a predição de recursos obtida no passo anterior.
 - d. A BS âncora busca prever para o dado trajeto a área alvo a partir da localização da MS obtida através do Serviço de Localização, e dos trajetos obtidos no passo (a). Vale ressaltar que a determinação da área alvo é obtida através da respectiva função, sendo dados a localização atual do usuário, a área atual do usuário e o trajeto mais frequentemente usado no dado período.
 - e. É criada uma lista das BSs contidas na área alvo prevista. Nesse passo, a BS âncora usa as informações de MOB_NBR-ADV recebidas das BSs vizinhas, para prever quais são as melhores BSs, do ponto de vista dos requisitos de QoS, na área alvo escolhida. De posse da lista das melhores BSs da área prevista, a BS âncora escolhe a BS alvo que melhor atende aos requisitos de QoS sugeridos pelo perfil do usuário, envia uma das mensagens, MOB_BSHO-RSP ou MOB_BSHOREQ, sugerindo reduzir a lista de BSs à BS alvo escolhida. Caso várias BSs atendam ao requisitos do usuário, é escolhida a que tem mais recursos disponíveis. Se houver empate, escolhe-se a primeira BS da lista que atende ao perfil. As informações dos recursos obtidas da Tabela de Recursos são armazenadas na tabela de fluxo-provisionado e em uma cache na BS âncora para consulta rápida posterior.
 - f. Deve ser repassado via *backbone* para todas BSs pertencentes a lista, o contexto da BS âncora, incluindo informações da Tabela de Fluxos pré-provisionados.
4. A seguir, considerando que a força do sinal da BS alvo se torne maior que a da BS Âncora, verifica-se se a predição está correta, isto é, se o usuário móvel permaneceu na trajetória prevista. Em caso positivo, a BS alvo passará a ser a nova BS âncora. No ARPRAP, a BS informa à MS da mudança de BS âncora passando o tempo estimado para a troca enviando a mensagem MOB_BSHO-

REQ ou MOB_BSHO-RSP. A MS pode aceitar/rejeitar a troca de BS através da mensagem MOB_HO-IND (campo MDHOFBSS_IND_type).

5. Na nova BS âncora, os fluxos pré-provisionados passarão a ser fluxos ativos, isto é, a nova BS âncora alocará dinamicamente estes recursos enviando a mensagem DSA (*Dynamic Service Addition*), disponibilizando esses fluxos de serviço a MS.

4.3 Mensagens Trocadas no ARPRAP

Em uma rede sem fio WiMAX, uma MS é capaz de receber de tempos em tempos mensagens do tipo MOB_NBR_ADV, contendo a intensidade do sinal de um conjunto de BSs vizinhas que estejam na sua área de alcance e o tempo de propagação entre a MS e cada uma dessas BSs vizinhas. A MS faz uma avaliação das medidas de intensidade de sinal e de tempo de propagação recebidas e repassa para sua BS âncora, via mensagem MOB_MSHO-REQ, um conjunto de possíveis BSs alvos. O conjunto dessas BSs é denominado de *Diversity Set* (conjunto de BS's vizinhas). Deve-se destacar também que as capacidades de cada BS integrante de um *backbone* são conhecidas pela BS âncora, pois as mesmas são difundidas por cada uma dessas BSs via mensagem MOB_NBR_ADV. Capacidade da BS refere-se a requisitos de redes, tais como a largura de banda e o atraso, os quais dependem dos serviços de agendamento suportados [IEEE Std 802.16e - 2005] – UGS (Unsolicited Grant Service), rtPS (Real-Time Polling Service), nrtPS (Non Real-Time Polling Service) e BE (Best Effort). Como mencionado, o ARPRAP faz uso do *handover* do tipo *Fast* (FBSS). No FBSS, a MS somente se comunica com a BS âncora para mensagens dos canais *Uplink* (UL) e *Downlink* (DL). A troca de BS âncora é efetuada sem invocar o HO tradicional. A MS envia a mensagem MOB_MSHO-REQ para adicionar/retirar BSs da lista (*Diversity*) ou para solicitar a troca de BS âncora, e esta responde com MOB_BSHO-RSP. A mensagem MOB_BSHO-REQ é usada com a mesma finalidade da MOB_MSHO-REQ, só que por iniciativa da BS âncora.

É importante salientar que, ao se usar o *handover* do tipo *Fast*, o contexto da BS âncora é repassado periodicamente, via *backbone*, para as BSs vizinhas. Embora tal procedimento seja considerado fora do escopo da norma, no ARPRAP foi necessário criar a

mensagem **BBONE_CONTEXT** para esse fim. Essa mensagem possui os seguintes campos: Tipo da mensagem de gerenciamento, Identificador da estação base, Identificadores das estações bases vizinhas e Contexto MAC.

No algoritmo proposto, foram acrescentadas novas informações a mensagem de MOB_NBR-ADV tais como: data e hora dos trajetos usados com mais frequência, e o(s) recurso(s) usado(s) nesse(s) trajeto(s).

4.4 Considerações finais do capítulo

Um dos grandes desafios tecnológicos na área de redes sem fio é conseguir uma utilização eficiente da rede, provendo serviços com qualidade para seus usuários, principalmente em presença de *handover*. O ARPRAP busca prover uma possível solução para lidar com tal desafio. Para tal foi proposto um algoritmo de predição de área alvo que leva em conta o perfil de comportamento do usuário, além do contexto do serviço e o da rede, para obter uma lista mais apurada de BSs a ser usada no *Fast Handover*. A utilização do algoritmo resulta, conseqüentemente, em (i) um menor número de transmissões entre BSs; (ii) um atendimento personalizado quanto ao quesito de QoS; (iii) um processo de *handover* rápido pela assertividade ao escolher a BS Alvo; (iv) uma redução do número de interrupções de conexões; e (v) uma redução de perdas na qualidade referente a serviço de tempo real, como o de voz sobre IP. Uma aplicação útil de tal algoritmo seria efetuar a reserva de recursos em trajetos previamente estabelecidos, por exemplo, de serviços considerados de emergência como os efetuados dentro de ambulâncias, viaturas policiais e bombeiros.

Em suma, nesse capítulo é apresentada uma proposta de um algoritmo que dinamicamente aloca recursos para as aplicações de usuários considerados Premium na célula alvo, escolhida pelo mecanismo de predição, obtendo alta utilização de recursos de forma a garantir níveis de QoS compatíveis com cada aplicação. Para tal, é aprimorado o algoritmo *Fast Handover* da camada 2 do IEEE 802.16e, refinando-se o processo de escolha das BS vizinhas candidatas ao *handover* (HO) com o auxílio do perfil de comportamento do usuário móvel, isto é, considerando aspectos da trajetória, requisitos da aplicação e de rede.

O próximo capítulo apresenta a implementação e análise de resultados do mecanismo de seleção de redes SUTIL.

5 Implementação e Análise de Resultados do SUTIL

Nesse capítulo são descritas as simulações realizadas para avaliação e verificação da validade da estratégia que foi adotada na concepção do SUTIL, conforme descrito no Capítulo 3. A fim de ilustrar o uso do SUTIL e avaliar o mecanismo de seleção numa situação próxima da real apresenta-se um estudo de caso e descrevem-se os resultados do conjunto de experimentos realizados no cenário apresentado. Além do cenário do estudo de caso, adicionalmente usou-se um cenário com maior restrição de banda com a intenção de capturar o comportamento do SUTIL nestas condições.

Este capítulo está organizado em cinco Seções. A Seção 5.1 descreve as ferramentas utilizadas para implementação do protótipo do SUTIL e para a realização das simulações. Na Seção 5.2 é apresentada a implementação do protótipo. A Seção 5.3 descreve a simulação e as análises dos resultados obtidos com a utilização do SUTIL. A última seção traz as considerações finais do capítulo em questão.

5.1 O Ambiente para Implementação e Simulação do SUTIL

O ambiente MATLAB, as ferramentas Simulink, a linguagem de modelagem para programação Matemática (AMPL) e o solucionador CPLEX foram escolhidos para a implementação do protótipo do SUTIL. O MATLAB foi escolhido porque é uma linguagem de alto nível que facilita a implementação, tem um ambiente interativo que permite executar tarefas de computação intensiva rapidamente e incorpora ferramentas para geração de gráficos a partir dos dados. Além disso, o MATLAB permite a integração dos seus scripts com as ferramentas Simulink.

O *Simulink* é um pacote de *software* para modelagem, simulação e análises de sistemas dinâmicos. Para a modelagem, o *Simulink* oferece uma interface gráfica com o usuário (GUI) para a criação de modelos em forma de diagrama de blocos. Cada bloco representa uma função ou um conjunto de funções associadas que recebe o nome de *subsistema*. Ou seja, a modelagem é representada graficamente em diagramas de blocos em vários níveis, pois um bloco ou subsistema pode constituir-se de vários outros blocos ou subsistemas. Esses podem ser interligados diretamente ou, quando não adjacentes, através do compartilhamento de uma área de memória específica para o *Simulink*. Além da utilização das ferramentas que compõem o *Simulink* foi também usado o editor de códigos do *MatLab* na confecção dos *scripts* necessários aos testes e às simulações. A linguagem utilizada nos *scripts* foi a do próprio ambiente MATLAB (*scripts *.m*).

A *linguagem de modelagem para programação Matemática* (AMPL) [FOURER 1993] é uma linguagem de modelagem algébrica poderosa para problemas de otimização lineares e não lineares, que permite o uso de variáveis discretas ou contínuas. A linguagem AMPL possui uma sintaxe bem próxima da descrição algébrica dos modelos de otimização e uma estrutura próxima das linguagens de programação estruturadas (C, PASCAL, FORTRAN). Além disso permite a inserção de algoritmos que manipulam as estruturas definidas para descrever o modelo, o que pode ser muito útil quando modelam-se classes de problemas de otimização linear inteira. A flexibilidade (em relação ao uso de diferentes algoritmos) e sintaxe simples a tornam ideal para rápida modelagem de problemas de programação linear inteira, enquanto a sua velocidade e opções de controle a fazem uma escolha especialmente eficiente para a execução repetida de simulações.

O CPLEX [ILOG] é um pacote de software de otimização. Seu nome advém do uso do método simplex e da linguagem de programação C, embora hoje ele tenha métodos e interfaces de linguagens como C ++, C # e Java. CPLEX resolve problemas de programação inteira e problemas de programação quadrática.

A fase de desenvolvimento do protótipo do SUTIL foi dividida nas etapas listadas a seguir, com as respectivas ferramentas utilizadas em cada uma delas:

- Concepção do esquema de integração entre o Matlab, Simulink, AMPL e CPLEX para implementação do protótipo;
- Confecção de modelo Simulink baseado nas equações da seção 3.4.2.
- Confecção de modelo em linguagem AMPL do problema de maximizar a função utilidade de forma a estabelecer o máximo de conexões possíveis entre os usuários e as respectivas redes-alvo preferenciais (ver Seção 3.4.3).
- Confecção de um *script MatLab para*:
 - Ler as entradas providas pelos dispositivos móveis dos usuários(serviços utilizados, energia residual e tipo de interface de rede);
 - Inicializar as variáveis de configuração
 - Carregar o modelo Simulink e ter como saída a matriz utilidade;
 - Com base nos dados de redes e MSs do cenário, na matriz utilidade e no modelo em AMPL fazer as melhores escolhas com o solucionador CPLEX.

5.2 Implementação

A Figura 17 apresenta o diagrama de atividades que descreve a implementação do protótipo no ambiente MATLAB. O primeiro passo é definir o número de usuários, o número de redes e os perfis de QoS (coeficientes de calibração entre relevância e tempo de vida α e β)(tarefas desempenhadas pelo operador); as variáveis do cenário são inicializadas em função da configuração dos parâmetros feita pelo operador e então é executada a primeira fase do SUTIL, onde são eliminadas as redes inadequadas, construídos os Serviços Representativos e o cenário produzido é salvo na matriz_cenário (tarefas desempenhadas por script MATLAB). O cenário então é lido de forma sequencial pelos subsistemas de energia e relevância e tem como saída uma matriz utilidade relacionando usuários e redes, onde cada usuário tem uma utilidade (preferência) para cada rede (tarefas desempenhadas pelo Simulink). Finalmente o solucionador recebe essa matriz utilidade, as solicitações de banda dos usuários, a banda disponível das redes e o modelo do problema de maximização das utilidades, e gera em sua saída uma matriz usuários x redes com as redes escolhidas para os usuários (tarefas desempenhadas pelo CPLEX/AMPL).

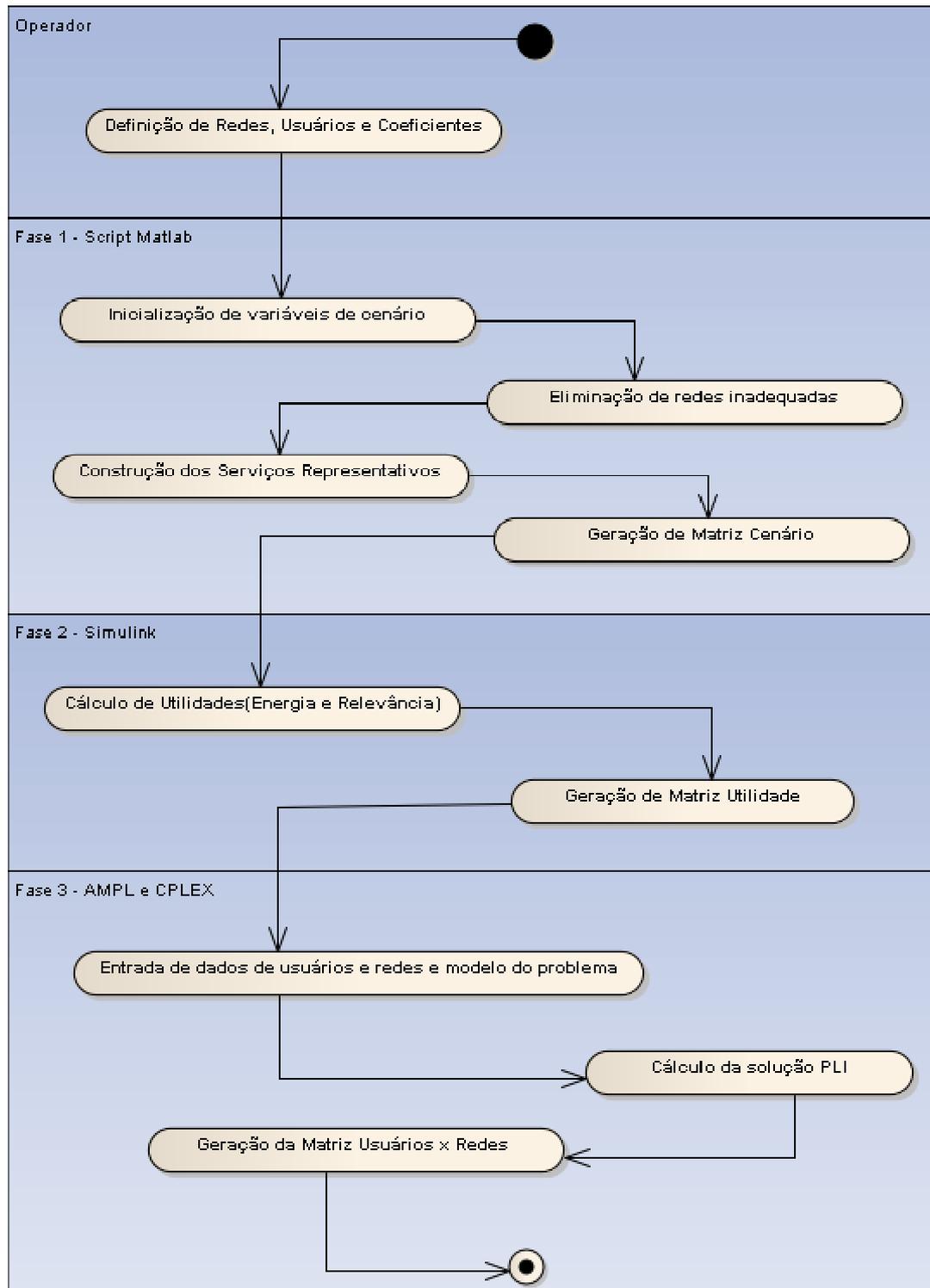


Figura 17. Diagrama de atividades de implementação

5.2.1 Construção dos Serviços Representativos (Script M)

A primeira fase do SUTIL implementou-se utilizando a linguagem própria do ambiente MATLAB. O *script inicializa.m* ficou responsável pela inicialização do cenário e pela primeira fase do SUTIL, com as seguintes funções: (i) perguntar ao operador quantos usuários, redes e os valores dos coeficientes de calibração de relevância (α) e tempo de vida (β) que deseja para o cenário; (ii) gerar os valores userID, atraso de rede, perda de rede, ID da rede e energia Residual, (iii) Eliminar redes inadequadas; (iv) inicializar os valores da matriz de Serviços Representativos; (v) Gerar matriz cenário; e (vi) instanciar o modelo Simulink Funcao_Utilidade.mdl.

O *script inicializa.m* realiza também a inicialização de todas as variáveis necessárias ao modelo *Simulink*, recalculando-as a cada rodada de simulação. A criação de relatórios (matrizes) contendo os valores obtidos pelos principais módulos do modelo *simulink* foi de suma importância nas fases de validação e testes de desempenho do protótipo. Para o armazenamento das variáveis e arquivos, o ambiente MatLab possui basicamente três áreas de memória. A primeira delas é o **diretório work** onde são armazenados em disco todos os arquivos oriundos das implementações e a serem utilizados nas simulações (*scripts*, modelo *Simulink*, e arquivos *.mat*). Um arquivo “*.mat*” armazena o estado de determinada variável ou grupo de variáveis após a execução de um *script*, para utilização posterior. A segunda área de memória é o **workspace** onde são armazenadas as variáveis inicializadas e/ou geradas nos *scripts*, em tempo de execução. Essas variáveis contidas no *workspace* também podem ser usadas pelo modelo *Simulink*, durante as simulações, através do seu carregamento para a **área de memória do Simulink** e/ou escritas dessa área para o *workspace*, após o término da execução do modelo. A área de memória do *Simulink* é a terceira, e é de uso exclusivo do modelo *Simulink*. Nesse local, ficam armazenadas as variáveis carregadas (lidas) do *workspace* e as criadas em tempo de simulação. Após o término da execução do modelo *Simulink*, essas variáveis são apagadas ou podem ser salvas no *workspace*, conforme já descrito. Durante a execução do *script* para controle da simulação, diversas variáveis são inicializadas. Um grupo dessas é de utilidade do próprio *script*, outras são para a utilização do modelo *Simulink* e outras são para o cálculo dos resultados estatísticos das simulações.

5.2.2 Cálculo de utilidades (Simulink)

A segunda fase do SUTIL foi implementada como um modelo *Simulink* que contém um conjunto de subsistemas responsáveis por ler as entradas advindas da "matriz_cenário", criada pelo *script inicializa.m*, e executar os cálculos para gerar uma matriz utilidade relacionando usuários e redes, baseada nas função utilidade e nas equações formuladas na Seção 3.4.2. A Figura 18 mostra o diagrama de componentes que representa a arquitetura do modelo Simulink proposto.

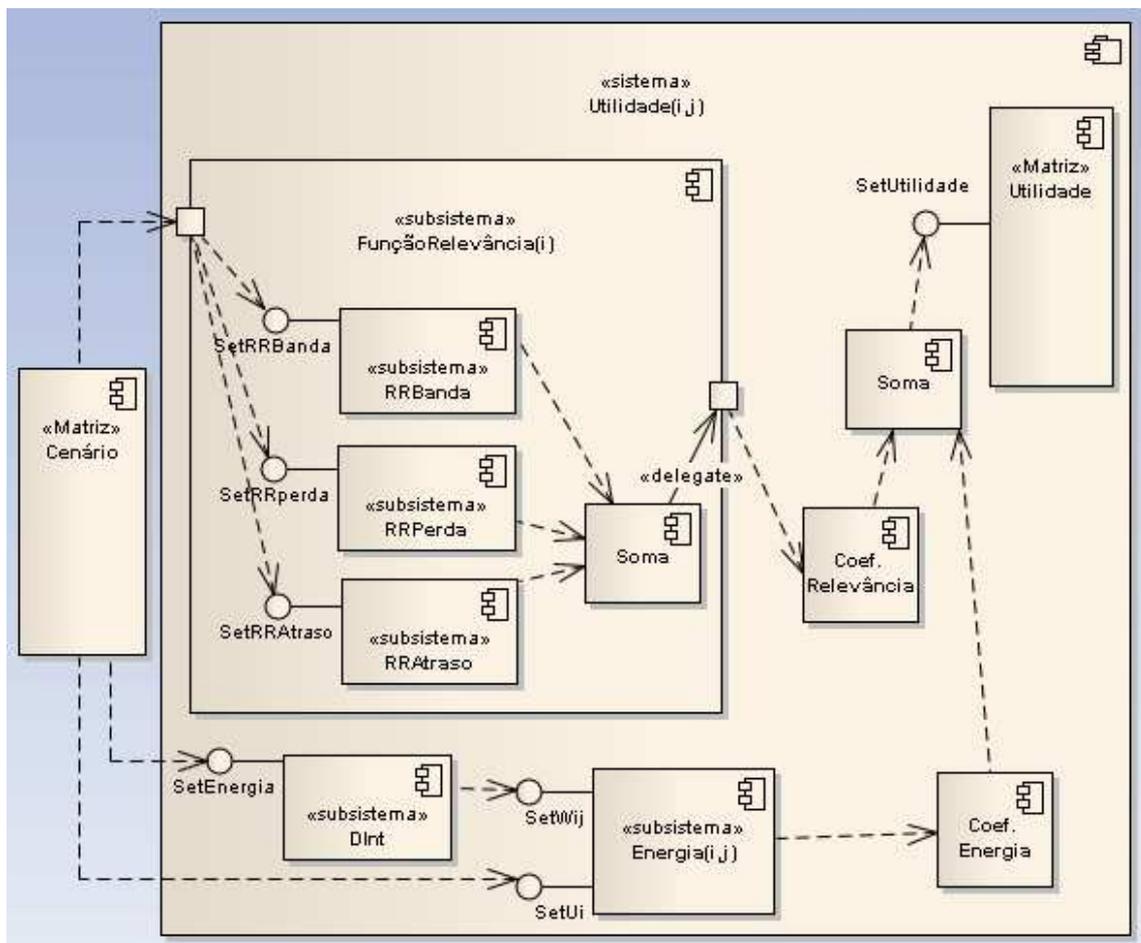


Figura 18. Componentes do modelo Simulink proposto

Na Figura 18 pode ser observada a *matriz_cenario* que foi gerada pelo *script inicializa.m* e serve de entrada para o *sistema Utilidade(i,j)*. O sistema *Utilidade(i,j)* é composto de outros subsistemas que implementam sua funcionalidade, tendo como saída uma *matriz utilidade* que será usada na terceira fase do SUTIL - Maximização PLI. A Figura 18 representa os subsistemas do sistema *Utilidade(i,j)*, onde pode-se observar que a

matriz_cenario tem seus valores multiplexados e cada valor é passado para os cálculos seguintes utilizando as respectivas interfaces disponibilizadas pelos subsistemas RRBanda, RRPerda, RRAtraso, DInt e Energia. O subsistema *DispositivosInterfaces(Dint)* tem os valores de consumo de energia de cada tipo de rede simulada (inicializados pelo script *inicializa.m*) e através da interface *SetEnergia* um valor de consumo é escolhido para cada MS. A seguir observam-se mais dois subsistemas, de *FunçãoRelevância* e *Energia* que implementam as equações descritas na Seção 3.4.2. Finalmente vem a saída que, como já citado, tem como produto uma *Matriz Utilidade*. O subsistema de *FunçãoRelevância* se divide em mais três que implementam os cálculos de requisitos representativos e recebem como entrada os serviços utilizados (banda disponível, atraso e perda de cada rede) pelo usuário através das respectivas interfaces *SetRRBanda*, *SetRRAtraso* e *SetRRPerda*. É importante ressaltar que a *matriz_cenario* depois de multiplexada apresenta os usuários e redes de forma serial, ou seja, cada usuário é confrontado com cada uma das redes, um por vez. O subsistema *Energia* implementa a equação relativa a energia $(U_i - W_{ij}) / U_i$ (Seção 3.4.2), recebendo os valores de U_i e W_{ij} através das respectivas interfaces *SetUi* e *SetWij*. Os valores calculados pelos subsistemas *RRBanda*, *RRPerda*, *RRAtraso* são somados (componente *Soma*) e o resultado é multiplicado pelo valor do coeficiente de relevância (componente *Coef. Relevância*). A saída do componente *Energia* da mesma forma tem o seu valor multiplicado pelo coeficiente de energia (Componente *Coef. Energia*). Em seguida, os valores de relevância e energia são somados (componente *Soma*) e armazenados na *Matriz Utilidade* usando a interface *SetUtilidade*.

5.2.3 Maximização PLI (AMPL/Cplex)

A *matriz utilidade* gerada na segunda fase do SUTIL, as solicitações de banda dos usuários, a banda disponível das redes e o modelo do problema de maximização da utilidade (AMPL) são utilizados como entradas para a maximização usando Programação Linear Inteira (PLI). A *matriz utilidade* e os dados de usuários são gerados em um arquivo texto nomeado *cenario.dat* que obedece a formatação usada pelo CPLEX. Um outro arquivo texto nomeado *max.mod* contém o modelo para o problema de maximizar a função utilidade a ser usado pelo solucionador CPLEX e pode ser visto na figura 19. A saída do solucionador CPLEX é uma matriz com as redes escolhidas pelo SUTIL para cada usuário.

```

1 set REDES;
2 set USUARIOS;
3
4 param custo {USUARIOS} >= 0;
5
6 param recursos {REDES} >= 0;
7
8 param utilidade {REDES,USUARIOS} >= 0, <=1;
9 param penalidade {REDES,USUARIOS} >= -1 integer, <= 0 integer;
10
11 var X{REDES,USUARIOS} >= 0 integer, <= 1 integer;
12
13 maximize Util:
14     sum{i in REDES} (sum {j in USUARIOS}
15         (utilidade[i,j] + penalidade[i,j]) * X[i,j] );
16
17 subject to umasoh {u in USUARIOS}:
18     sum {r in REDES} X[r,u] = 1;
19
20 subject to recmax {r in REDES}:
21     sum {u in USUARIOS} custo[u] * X[r,u] <= recursos[r];
22
23 subject to user_rede {r in REDES}:
24     sum {u in USUARIOS} X[r,u] <= 10;

```

Figura 19. Modelo para o problema de maximizar utilidades

Depois de expor detalhadamente as três fases do SUTIL, na próxima Seção passa-se a analisar os resultados obtidos nas simulações, começando por um estudo de caso.

5.3 Estudo de Caso 1: Aplicando SUTIL em Cenário próximo do real

5.3.1 Descrição

A fim de avaliar o desempenho do mecanismo proposto em um cenário que simula o que pode ser encontrado em uma situação real, nesta seção propõe-se um estudo de caso. Considerou-se o cenário de uma equipe de resgate destinada a cobrir a via federal Dutra que liga Rio de Janeiro a São Paulo. Determinados trechos dessa via são cobertos por redes celulares de diferentes operadoras, enquanto outros trechos também possuem redes Wimax mantidas pelas prefeituras de algumas das cidades locais, além de provedores de acesso e/ou indústrias localizadas na região. Além disso, alguns trechos específicos podem contar com redes Wi-fi de restaurantes, hotéis e cafés ao longo da via. Quando ocorre um acidente, as ambulâncias mais próximas são acionadas pela polícia via celular ou rede Wimax dependendo da disponibilidade dessas redes. Durante o percurso entre a central de resgate e o local do

acidente, normalmente a equipe de paramédicos recebe nos seus palmtops informações sobre o local, tipo e gravidade do acidente e imagens e vídeos das câmeras posicionadas ao longo da estrada, utilizando-se para tal de redes Wimax ou de redes celulares instaladas na região. Essas imagens irão auxiliar os paramédicos, por exemplo, na preparação do kit de equipamentos necessários para atender os feridos. Depois de chegar ao local do acidente, os paramédicos podem fazer videoconferência com as equipes médicas de plantão dos hospitais da região de forma a receber orientações, além de mandar imagens de exames já efetuados no local do acidente usando seus palmtops.

Em geral, os acidentes ocorridos na via Dutra provocam congestionamentos nas redes celulares. Tal congestionamento se deve ao uso simultâneo de um grande número de celulares por parte dos usuários presos nesse engarrafamento. Um grupo de usuários usa seus celulares para entrar em contato com seus familiares ou locais de trabalho, enquanto outros aproveitam o tempo em que se encontram retidos no engarrafamento para assistir algum programa na TV ou, ainda, participam de reuniões via videoconferência.

Os paramédicos e todos os usuários retidos no engarrafamento deslocam-se lentamente na via Dutra, e seus dispositivos de comunicação precisam efetuar transições de uma rede para outra, ou seja, efetuar operações de *handover*, de forma a garantir a conectividade e a qualidade de serviço dos serviços em uso. Este cenário é fundamentalmente composto por: (i) paramédicos e motoristas utilizando as mais variadas aplicações com diferentes necessidades de QoS; e (ii) um conjunto de redes apresentando disponibilidade de recursos e cargas de utilização variadas.

Considerando-se um acidente em particular ocorrido na via Dutra e uma ambulância onde o paramédico João, chefe da equipe de resgate, inicia no Palm o aplicativo **Resgate** para receber imagens e relatos do acidente. Ao mesmo tempo, João faz uma **videoconferência** com 3 hospitais da região passando informações do número de feridos, bem como da necessidade dos recursos que os hospitais teriam que oferecer para atendê-los. Analisando as imagens recebidas pelo aplicativo **Resgate**, João reporta aos dirigentes dos hospitais sobre a necessidade de realizar ressonância magnética e de equipamentos específicos para atender queimados. Por outro lado, os dirigentes informam a capacidade de atendimentos de seus hospitais naquele momento. Com base nessas informações, João pode distribuir os feridos entre os hospitais.

Nesse cenário, a partir do ponto de vista da MS de João, o SUTIL deve, antes de iniciar um procedimento de *handover*: (i) obter os requisitos de QoS dos aplicativos **Resgate**

(Imagens e Textos) e **Videoconferência** (Voz e Dados); (ii) criar o serviço representativo dessas duas aplicações (Voz, Video e Dados); (iii) criar os requisitos representativos desse serviço representativo; (iv) agrupar os usuários de acordo com o seu padrão de mobilidade (direção); (v) verificar se todas as redes candidatas têm capacidade para atender aos requisitos de todos os serviços em uso pela MS de João, bem como de cada MS em particular do grupo de MSs prontas para o *handover*; se a MS de João não for muito exigente, ela pode ser associada a uma rede que não foi aprovada para outra MS (Seção 3.4.1) ou ela pode permanecer no processo de seleção; (vi) calcular a utilidade de cada uma das redes disponíveis para cada um dos usuários móveis candidatos a efetuar o processo de *handover* junto com o João; (vii) selecionar as melhores redes possíveis para o João e demais usuários móveis candidatos ao processo de *handover*, segundo as utilidades das redes obtidas na etapa anterior; e finalmente (viii) passar o resultado para o módulo responsável pela execução do processo de *handover* do Palm do João e dos demais usuários envolvidos.

O serviço que representa as aplicações em uso por João são de voz, dados e imagens. Os requisitos representativos de banda, atraso e perda desse serviço representativo, conforme Tabela 5, são respectivamente: 167-692 kbps, 5-120 ms e 2-3 %.

As redes celulares responsáveis pela cobertura do trecho da via onde ocorreu o acidente estão todas congestionadas devido ao uso intensivo dos celulares por parte de seus assinantes. Conforme a ambulância se desloca para chegar ao local do acidente, o Palm de João, que está conectado a rede Bright, precisa efetuar a transição para outra rede, garantindo a conectividade e a qualidade de serviço necessárias para os aplicativos **Resgate e videoconferência**.

Neste momento, a BS da rede celular Bright está hospedando João e mais 6 usuários que se deslocam em direção ao local do acidente, além de 3 usuários que se deslocam em direção oposta. Dentre todas as redes que dão cobertura para o trecho em questão, o mecanismo SUTIL detectou que as redes celulares Live(N11) e Tones(N12) não atendiam naquele momento aos requisitos dos aplicativos do paramédico João. Ou seja, a quantidade de banda de 92 e 150 kbps para as respectivas redes celulares Live e Tones são insuficientes para atender os serviços em uso pelo João. Assim, para cada uma das redes, o mecanismo SUTIL associa a primeira MS dentre as MSs que são menos exigentes do conjunto de MSs em transição (ver seção 3.4.1). Entretanto, o SUTIL também conseguiu detectar que 3 redes WiMax e 7 redes Wi-Fi estão localizadas na mesma direção do deslocamento de João e, por conseguinte, do seu PALM. Os recursos de banda para as 3 redes wimax, N1, N2 e N3, são

respectivamente, 530 Kbps, 2Mbps, e 923kbps e os recursos para as redes Wi-Fi são 215, 425, 333, 468, 925, 92 e 150 Kbps. Na direção oposta ao deslocamento de João e, por conseguinte, do seu PALM, foram detectadas 2 redes Wi-Fi com banda disponível respectivamente de 789 e 655 Kbps . Dentre essas redes, a N3 é a rede que naquele momento poderia fornecer o melhor conjunto de valores de banda, atraso e perda para atender aos requisitos dos aplicativos em uso pelo João.

5.3.2 Avaliação

Considera-se que o Cenário do Estudo de Caso proposto é composto de: (i) doze redes candidatas ao processo de *handover* com os respectivos requisitos de banda disponível, atraso e perda de pacotes iguais a (N1, 923, 86, 3), (N2, 530, 73, 2), (N3, 2000, 5, 2), (N4, 215, 30, 2), (N5, 425, 83, 3), (N6, 333, 42, 3), (N7, 789, 61, 2), (N8, 655, 53, 3), (N9, 468, 110, 2), (N10, 925, 120, 2), (N11, 92, 150, 8) e (N12, 150, 150, 8 voz); (ii) o usuário João que usa vídeo e dados; dois usuários que utilizam voz e vídeo; e sete usuários que utilizam apenas voz, com os seus limites superiores de banda, atraso e perda de pacotes iguais a (João, 692, 120, 3), (U2, 64, 150, 8), (U3, 64, 150, 8), (U4, 64, 150, 8), (U5, 192, 150, 3), (U6, 192, 150, 3), (U7, 64, 150, 8), (U8, 64, 150, 8) e (U9, 64, 150, 8) e (U10, 64, 150, 8). O SUTIL analisa os usuários (João, U2, U3, U4, U5, U7 e U10) que seguem em direção ao local do acidente (isto é, os usuários no sub-conjunto MS-F, de acordo com a divisão explicada na Seção 4.4.1) e também as redes (conjunto AF) localizadas nessa área (N1, N2, N3, N4, N5, N6, N9, N10, N11, N12). SUTIL também analisa os usuários (MS-T) que se movem na direção oposta (U6, U8 e U9), bem como as redes (AT), localizadas nessa (N7 e N8).

Como as redes N11 e N12 não foram capazes de atender aos requisitos da MS de João, foram-lhe atribuídos, respectivamente, os usuários U2 e U3, uma vez que essas redes possuem requisitos suficientes para atendê-los. Portanto, estas redes (N11 e N12) e estações (U2 e U3) não participam das fases seguintes do SUTIL.

Tabela 5. Matriz Utilidade (ÁREA da FRENTE)

	CN	N1	N2	N3	N4	N5	N6	N9	N10
João	0	0,65	0,69	0,75	0,31	0,63	0,36	0,44	0,6
U 4	0	0,67	0,75	0,62	0,63	0,72	0,63	0,66	0,63

U 5	0	0,67	0,75	0,62	0,49	0,72	0,49	0,53	0,63
U7	0	0,67	0,75	0,62	0,63	0,72	0,63	0,66	0,63
U10	0	0,67	0,75	0,62	0,63	0,72	0,63	0,66	0,63

Tabela 6. Matriz Utilidade (ÁREA de TRÁS)

	CN	N7	N8
U6	0	0,54	0,69
U8	0	0,67	0,68
U9	0	0,67	0,68

Nas matrizes das Tabelas 5 e 6, cada posição F_{ij} fornece a utilidade da rede j em relação a um usuário i e seu valor reside no intervalo $[0,1]$. Quanto mais próximo de um (1) é o valor da F_{ij} , maior é a utilidade (preferência) da rede j em relação a um usuário i , em comparação com as outras redes. É importante notar que, em um dado instante de tempo, cada usuário móvel candidato ao processo de *handover* é associado com uma única rede (a rede atual - CN)). Note na Tabela 5 que a utilidade da atual rede (CN) é zero (0).

De acordo com a matriz da Tabela 5, o mecanismo SUTIL atribuiu a rede N3 para João, a rede N2 para os usuários U4, U5, U7, e U10. As outras redes (da ÁREA da FRENTE) não foram atribuídas a nenhum usuário, constatando que os usuários foram realmente alocados para as redes de suas máximas preferências(utilidades). De acordo com a matriz da Tabela 5, o SUTIL atribui a rede N8 (da ÁREA de TRÁS) para os usuários U6, U8 e U9. O CPLEX foi executado 30 vezes, atingindo um intervalo de confiança de 95%. A ordem de apresentação dos usuários ao processo de *handover* foi aleatória e todos os usuários foram avaliados simultaneamente (solução Global). Como esperado, o valor de preferência (utilidade) calculado para a rede N3 em relação ao usuário João foi de 0,75 (o maior valor), enquanto as utilidades obtidas para as outras redes em relação ao usuário João variaram de 0,31 a 0,69.

5.4 Estudo de Caso 2: Aplicando SUTIL em Cenário Reduzido

5.4.1 Descrição

Novos experimentos foram conduzidos em um novo cenário com o objetivo de avaliar qual é a melhor solução: Global ou Individual. Na solução Global o SUTIL faz a seleção de redes para todos os usuários simultaneamente, enquanto na solução Individual os usuários são apresentados um por vez ao SUTIL. Os itens avaliados são : (i) o desempenho das soluções Global e Individual do SUTIL em encontrar as preferências dos usuários; (ii) a eficiência da distribuição de carga entre as redes preferenciais; (iii) o tempo gasto para realizar a solução quando o número de redes e usuários é aumentado (escalabilidade da solução em termos de tempo de CPU); (iv) o comportamento da solução quando o número de usuários e recursos (redes/ banda) é aumentado; e (v) o efeito da utilização de diferentes perfis de QoS.

Para confrontar as soluções Global e Individual, vamos considerar um novo cenário composto por: (i) quatro redes candidatas ao processo de *handover* com as seguintes disponibilidades de banda: (N1, 1132), (N2, 1128), (N3, 1199) e (N4, 853); (ii) nove usuários com as seguintes solicitações de banda: (U1, 692), (U2, 628), (U3, 628), (U4, 692), (U5, 128), (U6, 128), (U7, 692), (U8, 192) e (U9, 192); e (iii) a *matriz utilidade* descrita na Tabela 7. Para ambas as soluções global e individual, o CPLEX foi executado 30 vezes, adotando um intervalo de confiança de 95%. A ordem de apresentação dos usuários ao processo de *handover* foi aleatória. Novamente, em um dado instante de tempo, cada candidato usuário móvel para o processo de *handover* é associado com uma única rede (a rede atual - CN).

Tabela 7. Matriz Utilidade

	CN	N1	N2	N3	N4		CN	N1	N2	N3	N4
U1	0	0,65	0,57	0,65	0,60	U6	0	0,82	0,59	0,69	0,62
U2	0	0,82	0,65	0,67	0,71	U7	0	0,82	0,63	0,67	0,69
U3	0	0,83	0,46	0,63	0,47	U8	0	0,83	0,65	0,70	0,70
U4	0	0,80	0,47	0,63	0,47	U9	0	0,83	0,62	0,69	0,67
U5	0	0,80	0,58	0,69	0,62						

Considerando esses objetivos, as métricas utilizados para avaliar o desempenho do mecanismo SUTIL foram: **Índice Global de Satisfação do Usuário (IGS)**, **Eficiência de Alocação de Recursos (EAR)** e **Tempo Total da Solução Global (TTS)**. O Índice Global de Satisfação do Usuário (IGS) é definido como sendo a soma das utilidades das redes selecionadas como alvo para cada usuário no final do processo de *handover*. O valor máximo de IGS para um determinado cenário é conseguido através da adição para cada usuário dos valores das utilidades das redes que têm o valor mais elevado (ver Tabela 7). O Tempo Total da Solução Global (TTS) é o tempo de CPU consumida pelo CPLEX para alcançar uma solução Global. A métrica Eficiência de Alocação de Recursos (EAR) é definida pela equação XIII.

$$\frac{[\sum_{i=1}^N \sum_{j=1}^M (z_{ij} F_{ij} \cdot BRRUpLim_j)]}{\sum_{i=1}^N \cdot BRRUpLim_i},$$

© $i \in \{1, \dots, N\}, j \in \{1, \dots, M\}$ XIII

onde: M é o número de redes, N é o número de usuários, F_{ij} é a utilidade da rede j em relação ao usuário i descrito na Seção 3. Quanto mais próximo de um (1) é o valor de EAR, maior é o número de redes preferenciais ocupados pelos usuários.

5.4.2 Avaliação Numérica dos Resultados para o Cenário Reduzido

Passa-se agora a avaliar os resultados das simulações realizadas com o auxílio das métricas de desempenho IGS, EAR e TTS definidas na subseção anterior.

- **Índice Global de Satisfação do Usuário (IGS)**

O IGS é capturado usando o modelo de otimização global que visa satisfazer simultaneamente as exigências de todos os usuários candidatos a *handover*, enquanto que ao mesmo tempo maximiza suas preferências, e garante que os recursos disponíveis para as redes não sejam ultrapassados. Sempre que não é possível atender simultaneamente todos os usuários candidatos a *handover*, devido à disponibilidade de recursos suficientes nas redes avaliados, alguns desses usuários são mantidos em suas redes atuais.

Para tal experimento, em uma situação ideal, o valor do IGS deve ser igual a 7,20. Este valor é facilmente obtido pela soma dos valores mais elevados de preferências (utilidade), conforme ilustrado na Tabela 7, considerando-se as redes de cada usuário (0,65 + 0,82 + 0,83 + 0,80 + 0,80 + 0,82 + 0,82 + 0,83 + 0,83) já que em uma situação ideal, todos os usuários

são atendidos em suas preferências máximas. Uma vez sabendo do caso ideal (hipotético, limite máximo teórico para esse cenário), vamos analisar as soluções fornecidas pelo SUTIL. Ao aplicar a solução global SUTIL, o IGS é igual a 5,97, valor obtido pela soma dos valores das preferências (utilidade) das redes selecionadas por SUTIL para cada usuário ($0,65 + 0,71 + 0,83 + 0 + 0,80 + 0,82 + 0,63 + 0,70 + 0,83$), tendo em conta as restrições (ver Seção 3.3). Quando se utiliza a solução global, resultados idênticos para a métrica IGS foram encontrados em todas as 30 rodadas de CPLEX. Esses resultados já eram esperados em todas as rodadas usando as mesmas preferências, independente da ordem em que os usuários foram avaliados por uma solução Global. Além disso, é importante observar que o valor obtido IGS é devido ao cenário restritivo escolhido, isto é, a soma das bandas de rede disponível é menor do que a soma das solicitações dos usuários, a fim de forçar a solução a falhar em satisfazer todos os usuários em suas redes preferenciais.

Em relação à solução Individual SUTIL, o valor obtido para IGS é igual a 5,50, em média, variando de 5,43 a 5,57 em 95% dos casos. Assim, para todas as situações simuladas (com diferentes ordens de apresentação de usuários) a solução Global alcançou os valores mais elevados de IGS, comparativamente à solução Individual. Esta última solução não fornece a melhor escolha possível para todo o grupo de usuários e é muito sensível à ordem com que os usuários são apresentados. Ou seja, os primeiros usuários têm suas melhores preferências satisfeitas e os últimos são servidos em suas piores opções.

- **Eficiência de Alocação de Recursos (EAR)**

A métrica EAR procura estabelecer uma relação entre os recursos alocados para os usuários nas redes escolhidas e a preferência (utilidade) das respectivas redes para esses usuários, ou seja, quanto maior o seu valor, maior a satisfação global dos usuários. Nos experimentos realizados com o objetivo de comparar a solução Global com a solução Individual, o EAR foi conseguido através da aplicação da equação XIV aos valores das preferências (utilidades) das redes selecionadas pelo SUTIL para cada usuário, (Tabela 7) e aos valores dos respectivos Limites Superiores de Banda para esses usuários (692, 628, 628, 692, 128, 128, 692, 192, 192). Como pode ser visto na figura 20.

$$\frac{(0,65 * 692) + (0,71 * 628) + (0,83 * 628) + (0 * 692) + (0,8 * 128) + (0,82 * 128) + (0,63 * 692) + (0,7 * 192) + (0,83 * 192)}{692 + 628 + 628 + 692 + 128 + 128 + 692 + 192 + 192}$$

Figura 20. Cálculo do EAR

Assim, o resultado normalizado é dado pela divisão do somatório do numerador (2354) com a soma dos Limites Superiores de Banda dos usuários (3972), alcançando um valor de 0,59 para uma solução Global. Os mesmos resultados foram obtidos em todas as 30 rodadas do CPLEX com a solução Global, já para a solução Individual, um valor de 0,42 foi obtido na pior das hipóteses e 0,58 no melhor dos casos (ocorreu apenas duas vezes nas 30 rodadas).

É interessante notar que, ao usar a solução Global, o valor de utilidade alcançado para o quarto usuário foi igual a zero uma vez que este usuário permaneceu na sua rede corrente (CN). Considerando a exigência de atraso, em 78% dos casos, a solução global escolheu redes com o atraso menor ou igual a pior solução "individual" e em 67% dos casos, a solução global escolheu redes com o atraso menor ou igual a melhor "solução individual". Considerando o requisito de perda de pacotes, os resultados foram 78% e 89%, respectivamente (Tabela 8).

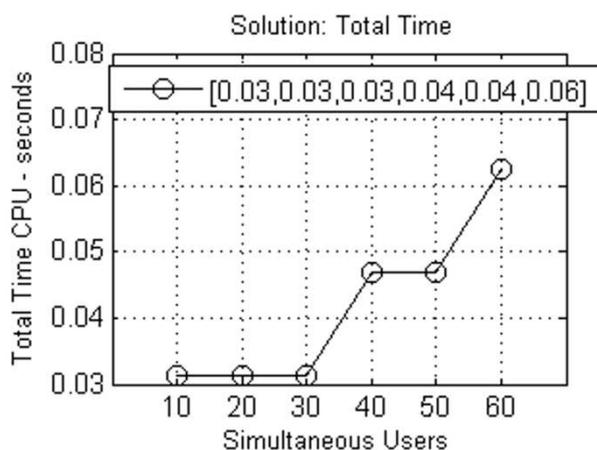


Figura 21. TTS Global

Os resultados de ambas as soluções garantem que a capacidade máxima da rede escolhida não foi excedida. No entanto, apenas a solução Global reúne um número maior de usuários considerando suas preferências. Sempre que os recursos disponíveis nas redes são insuficientes, a solução Global procura manter em sua rede atual aqueles usuários que, em tal

situação, contribuirão para a melhor solução para todo o conjunto de usuários. Além disso, ambas as soluções impedem que muitos usuários móveis, candidatos para realizar o processo de *handover* e localizados na mesma BS, migrem para a mesma rede preferencial em um determinado instante de tempo (sincronização de *handover*). Este resultado é alcançado através da aplicação da restrição R2 em ambas as soluções, assegurando que não haverá deterioração da nova rede, devido à carga súbita imposta pela migração desses usuários móveis.

Tabela 8. Comparação entre os atrasos e perdas de pacote

Solução	U1	U2	U3	U4	U5	U6	U7	U8	U9
Global	70; 6	4; 7	134; 7	none	134; 7	134; 7	115; 5	70; 6	134; 7
Pior Indiv.	none	4; 7	115; 5	70; 6	134; 7	134; 7	none	134; 7	134; 7
Melhor Indiv.	115; 5	4;7	134; 7	none	134; 7	134; 7	70; 6	134; 7	4; 7

- **Tempo Total da Solução Global (TTS)**

O objetivo principal do terceiro conjunto de experimentos é investigar o tempo gasto pela solução Global quando o número de usuários é variado proporcionalmente à quantidade de banda disponível nas redes. Para este conjunto de experiências, um novo cenário hipotético foi simulado, que é composto por quatro redes candidatas ao processo de *handover*, com os seguintes valores de banda disponível: (N1, 4000), (N2, 5800), (N3, 1000) e (N4, 700). Além disso, esse cenário também inclui 10 usuários com a seguinte requisição de largura de banda (U1, 2800), (U2, 1500), (U3, 1500), (U4, 1100) (U5, 1000), (U6, 1000), (U7, 900), (U8, 600), (U9, 500) e (U10, 300). A metodologia adotada para aumentar proporcionalmente o número de usuários e de banda foi dobrar, triplicar, e assim por diante, as banda disponível das redes, bem como o número de usuários (inicialmente igual a 10).

Portanto, a solução foi feita para atender 10, 20, 30, 40, 50, 60 usuários candidatos para realizar o processo de *handover* e localizada em um determinado BS, considerando 4 redes com suas respectivas bandas disponíveis, que também foram proporcionalmente variadas. O gráfico da Figura 21 mostra os resultados obtidos. Cabe salientar que o SUTIL atua antes do processo de *handover* e leva em consideração apenas os usuários que irão realizar o processo de *handover* simultaneamente. Além disso, de acordo com [BECVAR 2007], o tempo necessário para atualizar as informações durante o processo de *Soft Handover*

entre MS e BS pode chegar a 5,12 segundos. Considerando esses fatores, os valores dos tempos obtidos nas simulações realizadas podem ser considerados baixos e, portanto, a solução fornecida pela SUTIL tem um bom desempenho para os cenários avaliados.

5.5 Comparando SUTIL com trabalho em [CHEN 2005]

A fim de realizar um estudo comparativo do SUTIL, foram realizados experimentos confrontando os resultados da solução utilizando o SUTIL com os obtidos durante a execução da solução proposta em [CHEN 2005]. A descrição detalhada de [CHEN 2005] pode ser encontrada no Capítulo 2. Uma vez que em [CHEN 2005] uma abordagem baseada em escolhas individuais é adotada, e por isso, não se considera a melhor solução para o conjunto total de usuários e/ou redes, escolheu-se este trabalho para comparar com SUTIL, a fim de reforçar o nosso argumento de que uma abordagem global para a seleção de rede é melhor do que uma solução individual. Os experimentos foram realizados com o objetivo de avaliar: (i) o desempenho da solução global SUTIL e da solução proposta em [CHEN 2005] em atender às preferências dos usuários; e (ii) a eficiência da distribuição de carga entre as redes preferenciais, tanto para SUTIL como para o trabalho em [CHEN 2005].

Considerou-se o mesmo cenário descrito na Seção 5.4.1 e a matriz de utilidade descrita na Tabela 7. A solução Global SUTIL e a solução proposta em [CHEN 2005] foram executadas 30 vezes, adotando um intervalo de confiança de 95%. A ordem de apresentação dos usuários ao processo de *handover* foi aleatória. Novamente, em um dado instante de tempo, cada candidato usuário móvel para o processo de *handover* é associado com uma única rede (a rede atual - CN)).

Resultados idênticos para a métrica IGS foram encontrados em todas as 30 rodadas da utilização da solução global. Para tal experimento, em uma situação ideal, o valor do IGS deve ser igual a 7,20. Ao aplicar a solução global SUTIL, o IGS é igual a 5,97, valor obtido pela soma dos valores das preferências (utilidade) das redes selecionadas por SUTIL para cada usuário. Quanto à solução descrita em [CHEN 2005], o valor obtido para IGS é igual a 4,77, em média, variando de 4,34 to 5,19 em 95% dos casos. Assim, para todas as situações simuladas (com diferentes ordens de apresentação dos usuários) a solução Global SUTIL alcançou os valores mais elevados de IGS, comparativamente a solução [CHEN 2005]. Para a métrica EAR, o mesmo valor de 0,59 foi alcançado em todas as 30 rodadas com a solução

global SUTIL. Nos experimentos realizados com a solução [CHEN 2005], um valor de 0,41 é obtido na pior das hipóteses e 0,54 no melhor dos casos.

É interessante notar que, ao usar a solução Global, apenas um usuário permaneceu na sua rede corrente (CN). No pior caso de [CHEN 2005] dois usuários permaneceram em suas redes correntes. Os resultados obtidos são uma consequência direta da solução [CHEN 2005] não fornecer a melhor escolha possível para todo o grupo de usuários e ser muito sensível à ordem de apresentação dos usuários, devido à abordagem individual do algoritmo (em que cada MS faz sua escolha). Em [CHEN 2005], a MS escolhe, entre as redes capazes de responder às suas necessidades, aquela com a menor quantidade de recursos disponíveis.

5.6 Perfis de QoS

Nesta subseção, os efeitos do uso de diferentes perfis considerados neste trabalho são avaliados. Para o **perfil baseado em desempenho** (no qual a prioridade é dada à relevância da rede), o valor do coeficiente α foi definido para 0 e β foi definido como 1. Para o **perfil baseado em tempo de vida** (a prioridade é dada para minimizar o consumo de energia), o valor de α foi definido como 1 e β a 0. Para o **perfil com baseado na relação entre o desempenho e vida útil**, os valores dos coeficientes α e β foram definidos como 0,5.

Considerando-se as mesmas configurações do cenário descrito na Seção 5.4.1, o efeito do uso de diferentes perfis foi avaliado. Para o **perfil com baseado na relação entre o desempenho e vida útil**, um valor de 5,97 para a métrica IGS foi obtido através da solução Global, (quando o valor máximo para IGS em uma situação ideal é igual a 7,20). Para o **perfil baseado em desempenho**, que privilegia a preferência com base na relevância em relação a preferência baseada em energia, as escolhas feitas pela solução Global permaneceram as mesmas, atingindo um valor de 5,97 para a métrica IGS.

Por outro lado, para o **perfil baseado em tempo de vida**, que privilegia a preferência com base na energia em relação a preferência com base na relevância, uma mudança significativa na escolha da rede é notada. Os usuários migraram para outras redes que oferecem o menor consumo de energia para realizar o processo de *handover*, causando um aumento de usuários nas redes de alto valor de utilidade no requisito economia de energia. Assim, para o **perfil baseado em tempo de vida**, atendendo à expectativa, foi obtido o valor de **7,22** usando a solução Global para a métrica IGS, maximizando as preferências de energia.

Estes resultados mostram que o coeficiente α e β podem ser usados para efetivamente priorizar o desempenho ou o tempo de vida, já que o resultado da solução Global converge para um valor próximo da máxima satisfação dos usuários, sem violar os requisitos, ou seja, acomodando todos os usuários dentro dos limites dos recursos de redes.

5.7 Estudo de Caso 3: Aplicando SUTIL em Cenários Maiores

A fim de analisar o comportamento do SUTIL em diferentes situações, o cenário (Estudo de Caso 2) foi aumentado em até 5 vezes de duas maneiras, cada uma caracterizando um cenário novo e distinto. Em ambos os cenários, foram aumentados o número de usuários e recursos disponíveis, além de utilizar diferentes configurações de restrição R3 (P igual a 10, 20, 30 e 40), uma vez que essa restrição está diretamente relacionada ao número de utilizadores alocados por rede. No primeiro cenário, foram utilizados os valores de 9, 18, 27, 36 e 45 como número de usuários e 4, 8, 12, 16 e 20 como número de redes. No segundo cenário, foram utilizadas as quantidades de 9, 18, 27, 36 e 45 usuários e o número de redes foi fixado em 4, mas aumentou-se a largura de banda dessas quatro redes em 2, 3, 4 e 5 vezes. Além disso, este último cenário foi testado com e sem a definição da restrição R3.

No primeiro cenário, com o aumento proporcional do número de usuários e redes, a solução Global SUTIL alcançou IGS igual a 5,97, 11,94, 17,91, 23,88 e 29,85. Além disso, o número de usuários atribuídos a nenhuma rede por falta de largura de banda disponível suficiente foi, respectivamente, 1, 2, 3, 4 e 5. Os valores de IGS aumentaram respectivamente 2, 3, 4 e 5 vezes. No entanto, a solução Global produzida por SUTIL era exatamente a mesma em todos os casos já que o valor IGS aumentou na mesma proporção que o número de usuários. Como nenhuma rede teve seu limite de número de usuários atingido, a restrição R3 não foi ativada, não tendo nenhum impacto nos resultados (Figura 22).

No segundo cenário, em vez de aumentar o número de redes, aumentou-se proporcionalmente a banda das 4 redes. Com $P = 10$ os resultados obtidos neste cenário para a métrica IGS foram 5,97, 13,14, 19,47, 25,55, 28,31, indicando, em primeiro lugar, uma notável melhora em relação ao primeiro cenário. No entanto, essa melhora diminuiu com o aumento do número de usuários. A razão para este comportamento é que a restrição R3 começou a atuar a partir de 27 usuários. Os resultados do segundo cenário são melhores do que os obtidos no primeiro cenário com até 36 usuários, uma vez que todos os usuários foram alocados em uma rede. Contudo, com 45 usuários os resultados do segundo cenário são o

piores que no primeiro cenário, mostrando 5 usuários que não foram alocados a nenhuma rede. Isto aconteceu devido ao fato de que as 4 redes atingiram o seu limite de 10 usuários cada (imposto por R3). Com $P = 20$, a restrição R3 começa a atuar em 45 usuários causando uma pequena diminuição no IGS (32,96 vs 33,16). Para melhor compreensão, em um cenário com 45 usuários sem usar a restrição R3, N1 tinha 23 usuários alocados, já com $P = 20$, 3 usuários foram transferidos para redes de menor utilidade para eles. Com $P = 30$ e $P = 40$ nenhuma das redes teve seu limite de número de usuários atingido, a R3 restrição não foi ativada, não tendo nenhum impacto nos resultados.

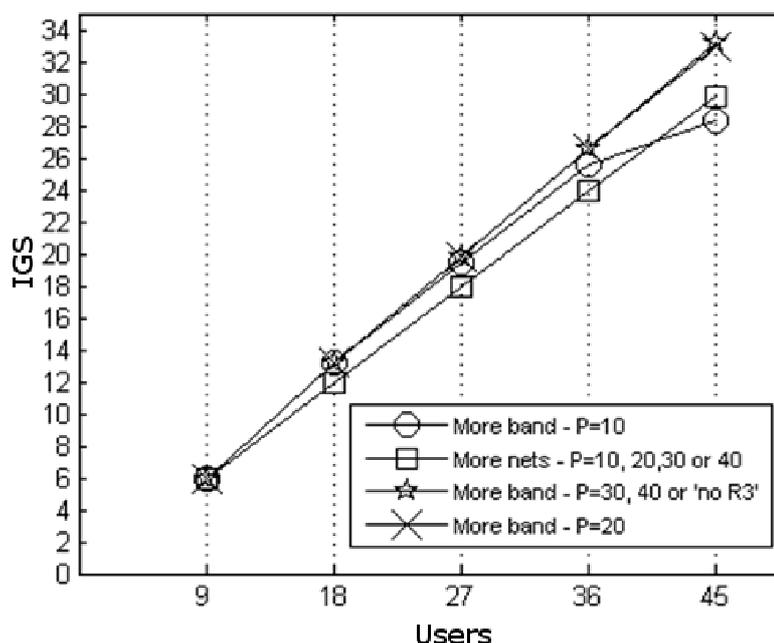


Figura 22. Escalabilidade e a Restrição R3

O desempenho melhor inicialmente alcançado pelo segundo cenário em relação ao primeiro cenário pode ser explicado pelo fato de que a solução inicial com 9 usuários e 4 redes já deixou um usuário sem alocação (devido à restrição R2) já que a banda disponível no primeiro cenário não foi suficiente para atender as necessidades do usuário. Com o aumento no número de redes esta situação se manteve. Entretanto, o aumento de banda mantendo o mesmo número de redes, conduz a uma situação em que a banda disponível em cada rede é suficiente para atender a todos os usuários. Para observar o impacto de R3, o segundo cenário foi simulado, sem a influência de tal restrição. Neste caso, os valores obtidos para IGS foram 5,97, 13,14, 19,83, 26,62 e 33,16 confirmando a influência da R3 uma vez que este cenário

tem melhor desempenho a partir de 27 usuários para $P = 10$ e de 45 usuários para $P = 20$, o ponto exato onde R3 começa a atuar neste segundo cenário.

A partir destas experiências é possível concluir que SUTIL tem um desempenho satisfatório em cenários maiores, mantendo a sua eficiência quando o número de usuários e redes são aumentados na mesma proporção e possui desempenho ainda melhor quando a banda da rede é aumentada em vez do número de redes. A restrição R3 impede que as redes tenham os seus valores de atraso e perda de pacotes degradados devido ao aumento excessivo de usuários conectados. No entanto, tal restrição deve ser ajustada com cuidado. Se o seu valor é demasiado restritivo, pode influenciar negativamente os resultados como acontece no caso de 4 redes, 45 usuários e $P = 10$ no segundo cenário. É importante notar que a solução fornecida pelo SUTIL está condicionada pelas limitações intrínsecas de qualquer solução centralizada. Como o SUTIL é executado na BS, ele é limitado pelas capacidades da mesma, como o poder de processamento e memória. Portanto, o SUTIL escala bem, enquanto o custo do processo de solução em um determinado cenário não ultrapassar as capacidades da BS.

5.8 Considerações finais

O Capítulo 5 apresentou uma descrição detalhada do processo de implementação do SUTIL e a infra-estrutura do ambiente necessário a execução das simulações, mostrando todas as etapas que as compuseram. A cada modelo implementado, o mesmo era submetido a testes de validações, a fim de se verificar a correção do mesmo. O próximo capítulo apresenta os resultados das simulações e análises do algoritmo ARPRAP.

6 Implementação e Análise de Resultados do uso do ARPRAP

Nesse capítulo são apresentados os resultados e as respectivas análises do uso do algoritmo ARPRAP.

Este capítulo está organizado em quatro seções. A Seção 6.1 descreve as ferramentas utilizadas para a simulação e implementação do ARPRAP. Na Seção 6.2 é apresentada a organização das simulações do ARPRAP. Na Seção 6.3 são apresentadas a simulação e análises dos resultados do ARPRAP. A última seção traz as considerações finais do capítulo em questão.

6.1 O ambiente de Implementação e Simulação do ARPRAP

O ambiente MATLAB foi escolhido para a implementação e simulação do ARPRAP. O MATLAB foi escolhido porque é uma linguagem de alto nível que facilita a implementação, tem um ambiente interativo que permite executar tarefas de computação intensiva rapidamente e incorpora ferramentas para geração de gráficos a partir dos dados. Foi usado o editor de códigos do MATLAB na confecção dos scripts necessários aos testes e às simulações. A linguagem utilizada nos scripts foi a do próprio ambiente MATLAB.

A fase de desenvolvimento do protótipo do ARPRAP foi dividida nas seguintes etapas:

- Confecção de três scripts MATLAB para implementação do algoritmo;
- Confecção de dois scripts MATLAB para mostrar graficamente a evolução da simulação ;
- Confecção de dois scripts MATLAB para coleta dos resultados e geração de gráficos;

6.2 Esquema das simulações

A Figura 23 apresenta um diagrama de atividades descrevendo a simulação no ambiente MATLAB. O primeiro passo da simulação é escolher o número de rodadas; em seguida são inicializadas as variáveis com a definição das coordenadas da área do cenário, características das BSs (quantidade, banda disponível aleatória e coordenadas), perfil do usuário, tamanho do caminho guardado no perfil (4 ou 8 áreas), trajetória a ser seguida pela MS. A simulação é iniciada com o movimento da MS que ao sair do alcance da BS corrente e ao aproximar-se de outras BSs executa o *handover* com as regras do ARPRAP ($isPremium = 1$) ou sem elas dependendo se a rodada de simulação foi configurada para ser executada, com

ou sem o ARPRAP ($isPremium = 0$); Têm-se então a plotagem na tela do cenário com as BS posicionadas, o deslocamento da MS, a inclusão de uma marcação com um triângulo das BS escolhidas para *handover*, com um círculo com o alcance das BSs, para acompanhamento da simulação em tempo real. Os dados gerados durante a simulação são guardados em variáveis e usados para o cálculo das métricas de desempenho **Probabilidade de Falha de Handover**, **Melhoria na Transferência de Dados** e **Eficiência na Alocação de Recursos**. O script na sequência gera os gráficos das métricas usando os valores calculados no passo anterior.

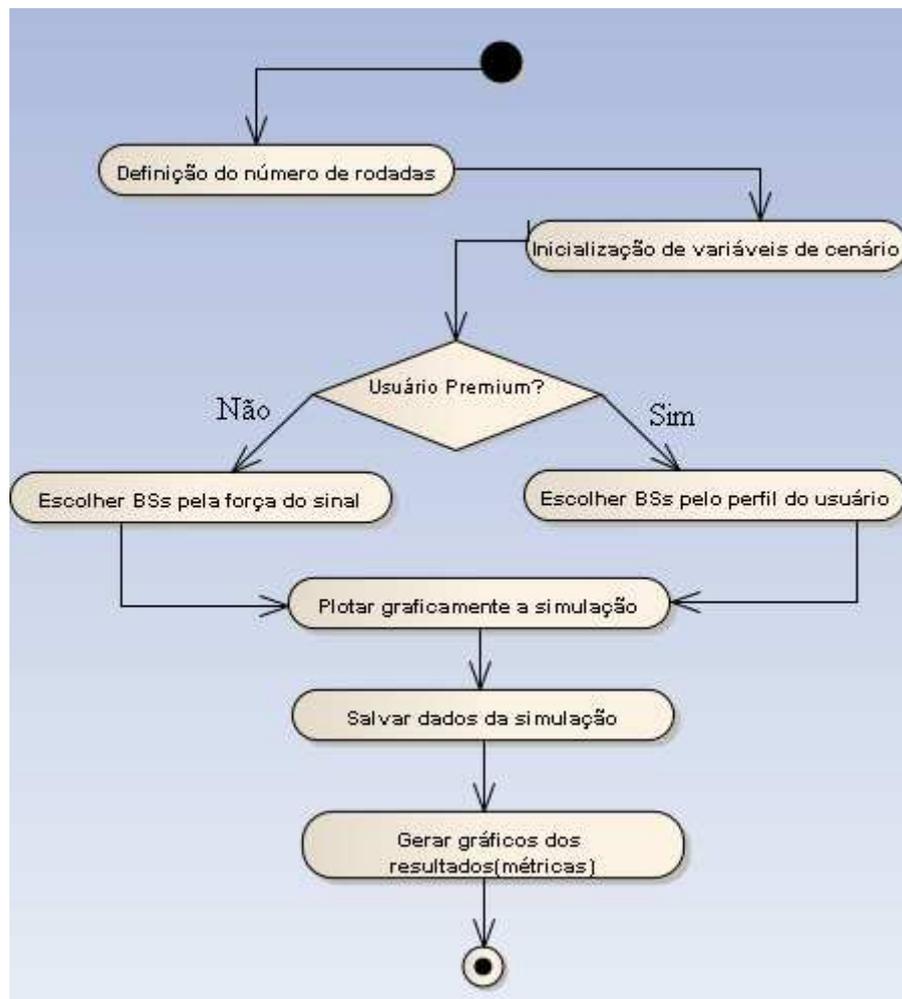


Figura 23. Diagrama de atividades de simulação

6.2.1 Implementação

Os scripts e funções implementados utilizam a linguagem própria do ambiente MATLAB (*M-files*) e são: (i) um *script* para inicialização das simulações – *Simula.m*; (ii) Um *script* e duas funções que implementam o cenário, o algoritmo de seleção de BSs por força de

signal e o algoritmo de seleção de BSs baseado no perfil do usuário ARPRAP, respectivamente – *arprap.m*, *bsProx.m* e *bsProxPremium.m*; (iii) Duas funções responsáveis por plotar os dados gerados em forma de gráficos para visualização – *plot_MS.m*, e *circle.m*; e (iv) Duas funções que calculam os resultados das métricas e geram o gráfico com os resultados *CreateFigure.m* e *int_confianca.m*. As funções do *script Simula.m* são: (i) Inicializar o cenário a ser simulado e o número de rodadas; e (ii) Inicializar parâmetros dos gráficos a serem gerados no final da simulação. O *script arprap.m* têm as funções de: (i) Montar o cenário, definir as posições das BS, gerar bandas aleatórias e comandar o movimento da MS; (ii) Chamar as funções que implementam os algoritmos de *handover* *bsProx.m* ou *bsProxPremium.m* de acordo com o que foi configurado na inicialização da simulação; (iii) passar os dados de simulação para plotagem em tempo real (*plot_MS.m*); e (iv) ao terminar a simulação passar os dados para o geração dos gráficos das métricas de simulação (*CreateFigure.m* e *int_confianca.m*). A geração dos relatórios estatísticos inclui o cálculo das médias dos valores das métricas escolhidas na simulação, o cálculo dos seus respectivos desvios padrões, variâncias e intervalos de confiança.

6.3 Simulações e Análises dos resultados do uso do ARPRAP

Essa seção descreve as simulações realizadas para avaliação e verificação da validade da estratégia que foi adotada na concepção do ARPRAP, conforme descrito no Capítulo 4. É também apresentada a análise dos resultados das simulações.

Assim, a seção está organizada em 4 subseções. Na Seção 6.1 tem-se os aspectos gerais da implementação e simulação do ARPRAP. Na Seção 6.2 é apresentado diagrama de atividades descrevendo a simulação e detalhes da implementação no ambiente MATLAB. Na Subseção 6.3.1 são descritos o ambiente e os cenários desenvolvidos para as simulações. Na Subseção 6.3.2 são apresentadas as métricas utilizadas para validar o desempenho do ARPRAP. Nas Seções 6.3.3, 6.3.4 e 6.3.5 são analisadas respectivamente as métricas **Probabilidade de Falha de Handover**, **Melhoria na Transferência de Dados** e **Eficiência na Alocação de Recursos**. A última seção finaliza com as considerações finais do capítulo.

6.3.1 Cenários das Simulações

Os benefícios da utilização do algoritmo proposto (ARPRAP) foram verificados através de simulações realizadas no MATLAB [MATLAB – SIMULINK] para dois diferentes cenários. No primeiro cenário (cenário 1) o ARPRAP foi utilizado para escolher a BS alvo

que melhor atendesse os requisitos definidos no perfil do usuário móvel para um dado caminho e a partir de um conjunto de BSs vizinhas candidatas a *handover*. Nas simulações realizadas, o perfil do usuário é uma estrutura de dados contendo informações sobre o caminho do usuário, requisitos de aplicação e de rede. No segundo cenário (cenário 2), o ARPRAP não foi utilizado. Ao invés disso, entre o conjunto de BSs vizinhas da BS Âncora, a BS escolhida era aquela cuja distância da MS em questão era menor que a distância entre a MS e a BS corrente.

Em ambos os cenários, a área de simulação (região) é definida como um retângulo de 6km x 6km, dividida em 16 áreas retangulares de 1,5km x 1,5km. Em cada área, coexistem 4 redes WiMAX, isto é, estão posicionadas 4 BSs. Quanto ao critério de posicionamento das BS estabeleceu-se que cada área é dividida em 4 sub-áreas e cada BS é posicionada no centro de cada sub-área. Logo, o número total de células (raio da célula =1,5 km), ou BSs, é igual a 64 (16x4). Cada célula é composta por uma BS e o número total de canais disponíveis por BS é dado por uma variável C. Como o objetivo das simulações é mostrar o quão eficiente é o algoritmo de predição de áreas alvos usando o perfil de comportamento do usuário móvel, foi adotado o valor de $C = 1$, sem perda de generalidade para as simulações realizadas.

O objetivo das simulações é demonstrar a eficiência do algoritmo para predição de áreas-alvo usando o perfil de comportamento do usuário. Nesse trabalho, define-se eficiência como sendo a capacidade do algoritmo de escolher a BS alvo a ser usada no *Fast Handover*, resultando, conseqüentemente, em menos transmissões entre BSs, em um atendimento personalizado quanto ao quesito de QoS, e permitindo um processo de *handover* rápido pela assertividade ao escolher a BS Alvo. O valor de banda atribuído para cada BS é selecionado de forma aleatória a partir do intervalo de 0 a 64kbps. Como se tem apenas um canal por BS, esse canal terá banda igual a sua BS. Dessa forma, quando é usado o algoritmo de predição proposto, é possível simular uma situação futura de não existência de BS que atenda aos requisitos da aplicação, ocasionando falha de *handover*.

No presente trabalho, as simulações foram realizadas usando-se uma única MS na região de simulação, impedindo assim que o tráfego de outras MSs mascarasse os resultados. Essa única MS, no cenário 1 é controlada por um usuário *Premium* enquanto que no cenário 2 é controlada por um usuário normal. Os requisitos de QoS das aplicações acordados por um usuário *Premium* correspondem aos da classe UGS e os de um usuário normal correspondem aos da classe *best effort*. É importante notar que os parâmetros de configuração da camada física não afetam os resultados da simulação (em ambos os casos) já que o algoritmo proposto

foi definido para garantir um requisito da camada de enlace. A trajetória de uma MS é definida como sendo uma seqüência de áreas. A seqüência (A_1, A_6, A_{11}, A_{15}) (ver seção 4.2.1) é um exemplo de trajetória da MS. O modelo de mobilidade adotado seguiu a abordagem apresentada em [PALCHAUDHURI 2005]. Tal abordagem procura modelar uma situação do mundo real, onde pessoas se deslocam de formas previsíveis na ida ao trabalho e na volta aos seus lares, usando uma das grandes vias que cortam as cidades e que levam das suas residências localizadas na periferia da cidade ao centro e vice-versa. Assim, durante a semana, pela manhã, haverá um grande fluxo de pessoas no sentido periferia-centro e, ao final da tarde, um grande fluxo de pessoas no sentido contrário, centro-periferia. Da mesma forma que pode ser modelado o perfil de comportamento das pessoas durante os dias da semana, é possível também fazer o mesmo nos finais de semana. O modelo de mobilidade baseado em perfil é ideal para rede WiMAX, que prevê cobertura sem fio em áreas livres, de periferia, rurais e cidades. Segundo o modelo adotado, a MS percorre o trajeto arbitrado em velocidades veiculares (40 a 60 Km/h). As trajetórias da MS são guardadas na Tabela de perfil do usuário de forma que seja possível prever o trajeto do usuário nesta mesma faixa de horário em outra oportunidade. Sabendo a área para a qual a MS vai se mover, o mecanismo de predição proposto nesse trabalho deve escolher uma BS que possui a melhor oferta de QoS dentre as BS existente na respectiva área.

Quanto mais regular for o comportamento do usuário, mais precisa a predição realizada. Entretanto, em um caso raro, o mecanismo de predição proposto pode ter resultados negativos quando ocorrer a combinação de três situações: (i) o usuário está perto o suficiente da BS alvo para o ARPRAP iniciar o *handover*; (ii) existe outra BS fora da área predita com sinal forte e que atende aos requisitos de QoS; e (iii) logo depois do *handover* o usuário rapidamente muda seu caminho em direção a esta célula fora da área predita. Neste caso um *handover* extra ocorre. Caso nenhuma das BSs na área predita atendam aos requisitos de QoS, uma falha de *handover* (indireta) ocorre ao invés de um *handover* extra.

6.3.2 Métricas de Desempenho

Em cada cenário de simulação, as métricas de desempenho avaliadas são: (i) Probabilidade de Falha de HO (PFH); (ii) Melhoria na Transferência de Dados (MTD); e (iii) Eficiência de Alocação de Recursos (EAR). Realizam-se 10 rodadas de simulações para cada cenário, ao final das quais a probabilidade de falha do HO e o número total de *handovers* no trajeto são obtidos.

A **Probabilidade de Falha de Handover** devido ao não atendimento pela BS da QoS solicitada (PFH) é um dos parâmetros da camada de enlace que mede a QoS de uma rede móvel sem fio e pode ser calculada segundo a equação XIV.

$$PFH = \frac{NFHO}{NFHO+NOHOE} * 100\% \quad \text{XIV}$$

onde: *NOHOE* = No total de ocorrências de *handover* efetivadas e *NFHO* = No de falhas de *handover* devido ao não atendimento pela BS da QoS solicitada.

A **Melhoria na Transferência de Dados** (MTD) compara a quantidade em bytes de dados transmitidos com e sem o uso do ARPRAP.

Quanto à eficiência da alocação de recursos feita pelo ARPRAP, ela foi avaliada em função de uma métrica **Eficiência na Alocação de Recursos** (EAR) que compara o número total de solicitação de *handovers* (NSHO) quando se usa ou não o algoritmo. A métrica foi definida como sendo, para uma mesma MS e uma mesma trajetória, a diferença entre o número total de solicitações de *handovers* (efetivados e os não efetivados) obtidas usando o FBSS - valor denominado *NSHON* - e número total de solicitação de *handovers* obtidas usando o ARPRAP (usuário **Premium**) – denominado *NSHOP* - dividida pelo número total de solicitação de *handover* usando o FBSS (equação XV). É importante mencionar que essa métrica mostra o quão eficiente (valores próximos a 100) é o algoritmo proposto no presente trabalho em relação ao desperdício de recursos.

$$EAR = \frac{NSHON-NSHOP}{NSHON} * 100\% \quad \text{XV}$$

6.3.3 Probabilidade de Falha de Handover

A métrica Probabilidade de Falha de *Handover* foi capturada com simulações utilizando trajetos com 4 e 8 áreas de extensão. As Figuras 24 e 25 apresentam gráficos que mostram a evolução das probabilidades de falha de *handover* quando se usam trajetos curtos(4 áreas) e longos(8 áreas) no perfil. Conforme esperado, os resultados das simulações provaram que a

probabilidade de falha de *handover* quando é usado o algoritmo ARPRAP (usuário *Premium*) é significativamente menor do que quando não é utilizado o algoritmo (usuário *Normal*), indicando que, caso haja uma BS com requisitos de QoS desejado pela aplicação dentre as BS vizinhas da âncora, o algoritmo proposto escolherá essa BS, ou seja, só haverá falha de *handover* no ARPRAP se não existir BS com tais requisitos. Observa-se ainda que quanto mais restritivos são os cenários em termos de recursos disponíveis, mais evidentes são os benefícios do algoritmo proposto. Os gráficos mostram que com 75% das BSs tendo recursos suficientes para atender os requisitos do usuário, a probabilidade de falhas para o usuário *Normal* é de 26% contra 0% para usuários *Premium*. Com somente 25% das BSs tendo recursos suficientes, esta diferença entre *Normal* e *Premium* cresce para 75% de probabilidade de falha contra 3% respectivamente.

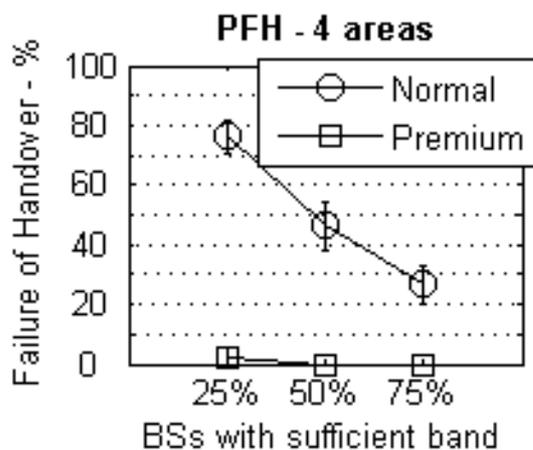


Figura 24. PFH em 4 áreas

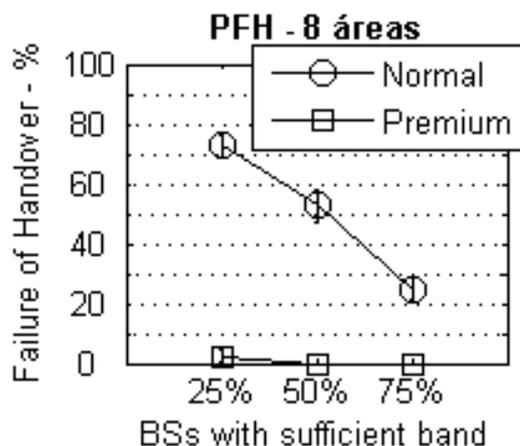


Figura 25. PFH em 8 áreas

O tamanho do trajeto do perfil do usuário a ser armazenado pode variar de acordo com o comportamento do usuário. Por exemplo, um usuário mais previsível se mantém em vias expressas enquanto um usuário menos conservador muda com frequência suas rotas, passando por pequenas ruas. Para o primeiro caso o perfil do usuário deve guardar trajetos mais longos (8 áreas) enquanto para o segundo caso trajetos curtos (4 áreas) são mais indicados. Trajeto com mais áreas aumentam a probabilidade da BS Âncora usar o cache ao invés de ler o perfil do usuário pela Internet, ou seja, usar as consultas anteriores salvas e assim melhorar a eficiência geral. Pode-se observar nas Figuras 24 e 25 que o algoritmo proposto realmente reduz a probabilidade de falha de *handover* de um usuário **Premium** quando comparado com a probabilidade de falha de *handover* de um usuário **Normal**. Isso ocorre pois o ARPRAP consegue, através do uso do perfil, prever as necessidades futuras de recursos e do trajeto. Um FBSS normal não tem como saber o trajeto a ser seguido, nem as necessidades de recursos futuras e faz *handover* para BSs que não são necessariamente a melhor escolha para o trajeto.

6.3.4 Melhoria na Transferência de Dados

A Melhoria na Transferência de Dados é capturada pela métrica MTD (Figura 26) que compara a quantidade de dados transmitidos em bytes com e sem o uso do ARPRAP, em um contexto onde 25%, 50% e 75% das BSs possuem banda suficiente para a realização de um *handover* com sucesso. As simulações mostram ganhos variando entre 22% e 52%. Estes ganhos são muito similares para os cenários com 4 e 8 áreas. A eficiência para os usuários

Premium em comparação com os usuários normais cresceu proporcionalmente ao nível de restrição do cenário.

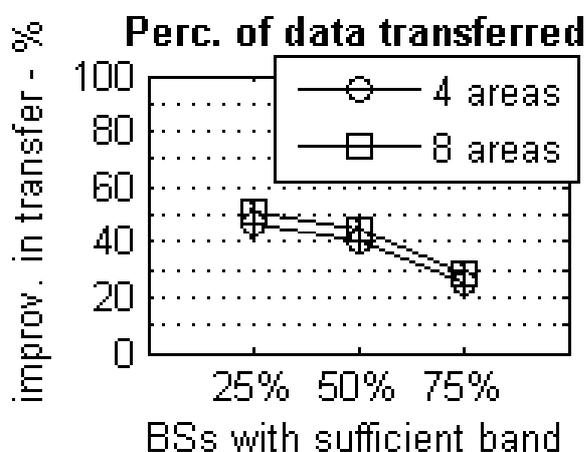


Figura 26. Melhoria na Transferência de Dados

6.3.5 Eficiência na Alocação de Recursos

Conforme mencionado, a eficiência da alocação de recursos feita pelo ARPRAP foi avaliada em função da métrica Eficiência na Alocação de Recursos (EAR) que compara o número total de solicitação de *handovers* (NSHO) quando se usa ou não o algoritmo.

O valor obtido para EAR foi de 67% para trajetos com 4 áreas e 50% para trajetos com 8 áreas (ver Figura 27). Adicionalmente, pode-se notar que esta métrica independe do número de BSs com recursos abundantes. Isso significa que, escolhendo a melhor BS em função da QoS e dos trajetos preditos, o ARPRAP é capaz de reduzir em 50% ou 67% o número de *handovers* necessários para completar o trajeto do usuário em comparação com o FBSS original. Esta redução reflete diretamente em desperdício de recursos: quanto menor a quantidade de solicitações de *handover*, menor será a quantidade de mensagens e contextos trocados (e, por consequência, menor a transferência de dados).

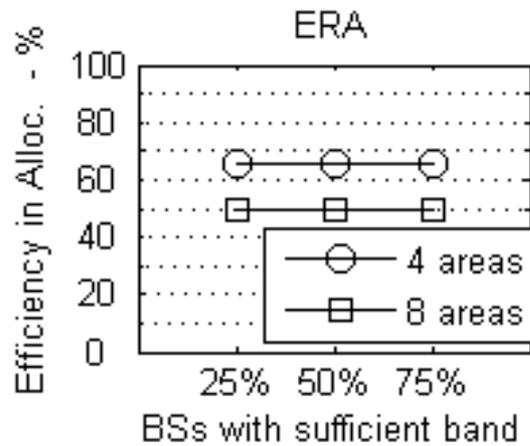


Figura 27. Eficiência na Alocação de Recursos

6.4 Considerações finais

O Capítulo 6 apresentou uma descrição detalhada do processo de implementação do ARPRAP e a infra-estrutura do ambiente necessário a execução das simulações, mostrando todas as etapas que as compuseram. A cada modelo implementado, o mesmo era submetido a testes de validações, a fim de se verificar a correção do mesmo. O próximo capítulo traz as conclusões deste trabalho.

7 Conclusão e Trabalhos Futuros

Esta dissertação teve como objetivo geral melhorar os processos relacionados ao serviço de gerência de *handover*, para maximizar a eficiência da rede e prover melhor qualidade de serviço em presença de mobilidade. Nesse sentido, o presente trabalho propôs o mecanismo SUTIL e o algoritmo ARPRAP, duas melhorias relacionadas ao serviço de gerência de *handover*. O SUTIL consiste em um mecanismo capaz de escolher a(s) melhor(es) redes em presença de *handover* considerando vários critérios para a tomada de decisão relacionados ao usuário, a rede e ao dispositivo. Já o ARPRAP é um algoritmo de predição de área alvo que leva em conta o perfil de comportamento do usuário, além do contexto do serviço e o da rede, para obter uma lista mais apurada de BSs a ser usada no *Fast Handover* e para dinamicamente alocar recursos para as aplicações de usuários considerados *Premium* na célula alvo.

Este capítulo está organizado em três seções. As Seções 7.1 e 7.2 apontam respectivamente as principais contribuições do mecanismo SUTIL e do algoritmo ARPRAP propostos. A Seção 7.3 enumera alguns trabalhos futuros.

7.1 Principais Contribuições do SUTIL

No presente trabalho foi apresentado o SUTIL, um mecanismo para seleção de redes em ambiente de redes de próxima geração (NGN) que prioriza redes com maior relevância para aplicação e menor gasto de energia para os dispositivos, contribuindo assim para que os dispositivos móveis e aplicações dos usuários estejam permanentemente conectados e, por conseqüência, para a viabilidade de sistemas ubíquos. Os resultados obtidos com a implementação do SUTIL demonstraram ganhos em termos de distribuição de recursos nas redes candidatas, do número de usuários alocados em redes capazes de atender a todos os requisitos de suas aplicações e da economia de energia de acordo com o dispositivo em uso pelo usuário móvel. Ainda, os resultados demonstraram que o SUTIL apresenta tempos de solução na casa dos centésimos de segundo, enquanto que o tempo gasto com a sinalização de um *handover* vertical pode variar de alguns segundos até várias dezenas de segundos [AKYILDIZ 2005], portanto a solução fornecida pelo SUTIL tem um bom desempenho para os cenários avaliados. Pode-se portanto considerar que o SUTIL consiste em um efetivo mecanismo de seleção de redes em ambientes NGN, dessa forma contribuindo para suplantando alguns dos importantes desafios existentes na exequibilidade de tais ambientes, e assim

indiretamente promovendo a viabilização da computação ubíqua conforme vislumbrada pelo seu preconizador Mark Weiser [WEISER 1991].

Um importante diferencial do SUTIL é sua abordagem global, que leva em consideração o conjunto de usuários e redes simultaneamente candidatos a *handover*, enquanto os trabalhos relacionados se utilizam de uma abordagem individual, que não permite uma visão holística onde a melhor solução para o sistema NGN como um todo pode ser encontrada. Os resultados obtidos na comparação entre a solução Global e a Individual (ver Seção 5.4.1), além da análise comparativa feita com o trabalho em [CHEN 2005] (ver Seção 5.5) comprovam as vantagens da abordagem global. A abordagem global permitiu encontrar a melhor distribuição possível para o conjunto de usuários e redes candidatos a *handover*, obtendo os melhores índices de usuários satisfeitos, ou seja, alocados segundo suas preferência em termos de energia e requisitos de suas aplicações como atraso, banda e perda de pacotes.

Na fase de testes do SUTIL, percebeu-se a necessidade de acrescentar dois novos procedimentos ao mecanismo. O primeiro procedimento tratava da introdução de um controle configurável que permitisse limitar o número máximo de usuários que uma única rede candidata pode receber. Percebeu-se que algumas redes possuem índices de perda de pacotes e atraso deteriorados em virtude do número de usuários conectados. O segundo procedimento consistia em alterar o SUTIL de forma a levar em consideração a direção e sentido do movimento do usuário, já que não fazia sentido o usuário fazer a transição (*handover*) para uma rede candidata da qual o usuário estava se afastando (BS em posição oposta ao movimento do usuário). Este procedimento ainda pode ser melhorado, já que no momento o SUTIL distingue apenas áreas de BSs candidatas a frente e atrás da BS corrente, podendo no futuro incluir divisões mais precisas com relação ao movimento da MS, sem contudo dividir excessivamente os grupos de MSs e BSs candidatas e tender para uma solução individual. Ou seja, pretende-se no futuro explorar o compromisso entre uma divisão maior ou menor das áreas consideradas no processo de decisão do SUTIL, analisando-se os prós e contras de cada cenário.

7.2 Principais Contribuições do ARPRAP

No presente trabalho apresentou-se também o ARPRAP, a proposta de um algoritmo que explora o perfil do comportamento do usuário em movimento para maximizar a eficiência da rede e prover melhor QoS para diferentes classes de usuários em um cenário que integra redes WiMAX.

Uma grande contribuição do ARPRAP é a sua capacidade de reservar os recursos antecipadamente nas BSs de destino do usuário móvel, utilizando-se de mecanismos de fluxo de serviço previstos na norma IEEE 802.16e, isto é, sem alterar o padrão. Esta capacidade está intimamente ligada ao perfil de comportamento do usuário que pode ser populado dinamicamente com os caminhos e recursos comumente utilizados pelo usuário ou por escolha prévia destes caminhos e recursos. A utilização do algoritmo proposto pode ser um diferencial oferecido pelos provedores de acesso para usuários do tipo *Premium*.

Em comparação com os trabalhos relacionados observou-se que a capacidade de prévia reserva de recursos do ARPRAP pode ser complementar a muitos dos trabalhos estudados, inclusive porque o ARPRAP é uma solução de camada 2 podendo ser utilizada em conjunto com as demais soluções de camada 3. O fato de ser de camada 2, por outro lado, torna a solução fortemente ligada a tecnologia de rede sem fio para que foi desenhada, no caso do ARPRAP, o WIMAX. Esta limitação de portabilidade existe porque é necessária uma adequação às características específicas de sinalização de handover das outras tecnologias de rede sem fio.

Resultados obtidos com a implementação do algoritmo proposto demonstraram a redução no número de falhas de *handover* e uma melhoria na transferência de dados e na alocação de recursos. A chave para estes resultados foram a eficiência demonstrada pelo algoritmo para predição de áreas-alvo usando o perfil de comportamento do usuário. Nesse trabalho, define-se eficiência como sendo a capacidade do algoritmo de escolher a BS alvo a ser usada no *Fast Handover*.

É importante ressaltar que quanto mais fiel ao comportamento do usuário for o perfil, tanto melhor será o desempenho da predição de BSs alvo e das reservas de recursos. Da mesma forma os trajetos e recursos que o usuário pode fixar previamente com a operadora de rede. Por este motivo, é importante aprimorar a forma de criar e manter estes perfis de usuários com estudos adicionais.

Além disso, os resultados demonstraram que o refinamento do conjunto de células previstas para o próximo *handover* promovem uma diminuição das transmissões entre as BSs

vizinhas. Demonstraram também que o usuário aumenta a sua chance de executar menos transições (*handovers*) para um mesmo trajeto, uma vez que se conhece antecipadamente a próxima área que o usuário se destina pode-se evitar *handovers* desnecessários.

7.3 Trabalhos Futuros

Futuramente, na linha de investigação relacionada ao SUTIL, pretende-se pesquisar a interação de múltiplas instâncias do mecanismo de forma que as decisões em uma BS levem em consideração as decisões das demais evitando que várias BSs escolham uma mesma rede preferencial. Uma solução que possibilita a existência de múltiplas interações no SUTIL consiste em fazer com que a matriz com as redes escolhidas para cada usuário resultante do processo de seleção do SUTIL seja enviada para uma entidade central, responsável por gerenciar o processo de *handover* no sistema NGN. Um exemplo desta entidade pode ser o NIA, descrito em [AKYILDIZ 2005], que fica localizado em um ponto central na Internet. Neste ponto central um novo módulo do SUTIL seria responsável por fazer checagens para evitar que uma BS candidata se torne congestionada por ser simultaneamente escolhida por várias instâncias do SUTIL.

Com relação aos trabalhos futuros do ARPRAP, pretende-se utilizar um simulador que implemente as camadas físicas e MAC do IEEE 802.16e a fim de investigar o impacto do uso do ARPRAP em termos de perda de pacotes e atraso. Além disso, averiguar a adoção de técnicas de Inteligência Artificial para “aprender” o comportamento do usuário e assim obter perfis de comportamento mais refinados.

Por fim, espera-se no futuro unir as melhores características do ARPRAP e do SUTIL em um único mecanismo de forma a tirar vantagem do perfil de comportamento do usuário e da alocação prévia de recursos também em *handovers* verticais e enriquecer o algoritmo de seleção de BSs numa rede homogênea com os conceitos de utilidade, além de fazê-lo considerando o contexto de usuários e redes simultaneamente candidatas a *handover*, ou seja, de forma global.

8 Referências Bibliográficas

AHSON, S.; ILYAS, M. **Wimax: standards e security**. CRC Press, 2007.

AKYILDIZ, I. F.; MOHANTY, S.; XIE, J. **A ubiquitous mobile communication architecture for next-generation heterogeneous wireless systems**. In: *IEEE Communications Magazine*, vol. 43, no. 6, pp. 29-36, Jun. 2005.

BECVAR, Z.; ZELEKA, J. **Implementation of Handover Delay Timer into WiMAX**. In: 6th Conference on Telecommunications, Peniche, Portugal, May 2007.

CARVALHO JR, J. C et al. **SUTIL - Network Selection based on Utility Function and Integer Linear Programming**. In: (COMNET), 2009, Revision 1.

_____. **User's Profile-Driven Resource Reservation for WiMAX networks**. In: 4th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), 2008.

CARVALHO JR, J. C. **Qualidade de serviço em redes IEEE 802.16: Como agendar tráfego para cumprir os requisitos QoS de classes heterogêneas de tráfego**. Orientador: Luci Pirmez. Rio de Janeiro, 2005. Monografia. (Pós-Graduação Latu Senso) - Universidade Federal do Rio de Janeiro, 2005.

CHALMERS, R. C.; KRISHNAMURTHI, G.; ALMERTH, K. C. **Enabling Intelligent Handovers in Heterogeneous Wireless Networks**. In: *Mobile Networks and Applications* 11, pp. 215-227, 2006.

CHANG, C. **A Mobile-IP Based Mobility System for Wireless Metropolitan Area Networks**. In: ICPPW'05, pp. 429-435, Oslo, Jun. 2005.

CHEN, W.; SHU, Y. **Active Application Oriented Vertical Handoff in Next-Generation Wireless Networks**. In: IEEE Communications Society / WCNC 2005.

CHOI, S. et al. **Fast handover scheme for real-time downlink services in IEEE 802.16e BWA systems**. In: IVTC'05, pp. 2028-2032 Volume: 3, Jun. 2005.

CHOU, J. et al. **MAC and PHY MIB for WirelessMAN and WirelessHUMAN BS and SS. IEEE 802.16 Broadband Wireless Access Working Group, IEEE C802.16mgt-04/04, pp.1-115, May 2004.**

CPLEX – **Mathematical Programming Optimizer**. Disponível em : <http://www.ampl.com>. Outubro 2009. Acesso em: Oct. 2009.

DELICATO, F. et al. **Redes WiMAX: Arquitetura, Protocolos, Segurança e QoS**. In.: Simpósio Brasileiro de Redes de Computadores, 26, 2008, Rio de Janeiro **Anais ...**, Rio de Janeiro: SBC, 2008.

FOURER, R.; GAY, D. M.; KERNIGHAN; B. W. **A Modeling Language, for Mathematical Programming**. Boyd & Fraser Publishing Company, 1993.

GUSTAFSSON, E.; JOHNSON, A. **Always Best Connected**, IEEE Wireless Communications, February 2003, vol.10, pp. 49-55.

HARRINGTON, D.; PRESUHN, R.; WIJNEN, B. **An architecture for Describing Simple Network Management Protocol (SNMP) Management Frameworks**, IETF RFC 3411, December, 2002.

HEYAIME-DUVERGE, C. ; PRABHU, V. K. **Traffic-Based Bandwidth Allocation for DOCSIS Cable Networks**. In: Proc. IEEE International Conference on Computer Communications and Networks, pp. 586-590, Oct. 2002.

HORRICH, S.; JAMAA, S. B.; GODLEWSKI, P. **Adaptive vertical mobility decision in heterogeneous networks**. In: IEEE Third International Conference on Wireless and Mobile Communications (ICWMC'07), 2007.

IEEE Std 802.11 - 1999, **Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications**, 1999.

IEEE Std 802.16 - 2004, **IEEE standard for local and metropolitan area networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems**, October 2004.

IEEE Std 802.16e - 2005, **IEEE standard for local and metropolitan area networks - Part 16: Air interface for fixed and mobile broadband wireless access systems**, February 2006.

ILOG. Disponível em: <http://www.ilog.com/products/cplex/>. Oct. 2008. Acesso em: Oct. 2008.

ITU-R. Disponível em: <http://www.itu.int/ITU-T/worksem/ictdev/200507/presentations/KeynoteAddress-CSLee.pdf>. Setembro 2005. Acesso em: Sep. 2009.

ITU-T. Disponível em: <http://www.itu.int>. Sep. 2009. Acesso em: Sep. 2009.

JIA, H. et al. **Study on network selection for next-generation heterogeneous wireless networks**. In: 17th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'06).

KANG, T. et al. **A Context-aware Handoff Management for Seamless Connectivity in Ubiquitous Computing Environment**. In: PSC 2006, pp. 128-134.

KIM, D.; GANZ, A. **Architecture for 3G and 802.16 Wireless Networks Integration with QoS Support**. In: IEEE Second International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks, Aug. 2005.

KNIGHTSON, K.; MORITA, N.; TOWLE, T. **NGN architecture: generic principles, functional architecture, and implementation**, *IEEE Communications Magazine*, vol. 43, no. 10, pp. 49–56, Oct. 2005.

LATVAKOSKI, E.J.; LAURILA, P.J. **Application based access system selection concept for all IP mobile terminals.** In: GLOBECOM 2002, vol. 3, pp. 2984-2988, Nov. 2002.

LEE, D.H.; KYAMAKYA, K.; UMONDI, J.P., **Fast Handover Algorithm for IEEE 80.16e Broadband Wireless Access System.** In: 1st International symposium on Wireless Pervasive Computing, 2006 1st International Symposium, Jan. 2006.

MATLAB – SIMULINK- **The Mathworks**, Inc. MATrix LABoratory and Simulink. Disponível em: <http://www.mathworks.com>. Setembro 2009. Acesso em: Sep. 2009.

MCNAIR, J.; ZHU, F. **Vertical handoffs in fourth-generation multinetwork environments.** In: Wireless Communications, IEEE , vol.11, no.3, pp. 8-15, Jun. 2004.

MOHANTY, S.; AKYILDIZ, I. F. **A Cross-Layer (Layer 2 + 3) Mobility, Management for Next Generation Wireless Systems.** In: IEEE Transactions on Mobile Computing archive Volume 5 , Issue 10 (Oct. 2006). pp. 1347-1360.

MOTOROLA PARTNER. Disponível em: <http://www.transceiver.com.br/>. Sep. 2008. Acesso em: Sep. 2008.

PANDEY, V.; GHOSAL, D.; MUKHERJEE, B. **Exploiting User Profiles to Support Differentiated Services in Next-Generation Wireless Networks**, Nortel Networks and University of California, IEEE Network Sep./Oct. 2004.

PANDEY, V.; GHOSAL, D.; MUKHERJEE, B. **Exploiting User Profiles to Support Differentiated Services in Next-Generation Wireless Networks** IEEE v.18, n.5, pp. 40-48, Sept-Oct 2004.

SHAABAN, S.; BADAWEY, H.; HASHAD, A. **Performance Evaluation of the IEEE 802.11 Wireless LAN Standards.** In: World Congress on Engineering 2008 Vol IWCE 2008, Jul. 2 - 4, 2008, London, U.K.

WEISER, M. **The Computer for the Twenty-First Century**, Scientific American, v267, n. 3, Sep. 1991, pp. 94-100.

ZAHNAN, A. H.; LIANG, B.; SALEH, A. **Signal threshold adaptation for vertical handoff in heterogeneous wireless networks**, Journal Mobile Networks and Applications, v. 11, n. 4, 2006.

ZHANG, Y.; CHEN, H. **Mobile WiMAX: Toward Broadband Wireless Metropolitan Area Networks.** Auerbach Publication, 2007.

ZHU, F.; MCNAIR, J. **Mutiservice vertical handoff decision algorithms.** In: EURASIP Journal on Wireless Communication and Networking, v. 2006, issue 2, pp. 1-13, 2006.