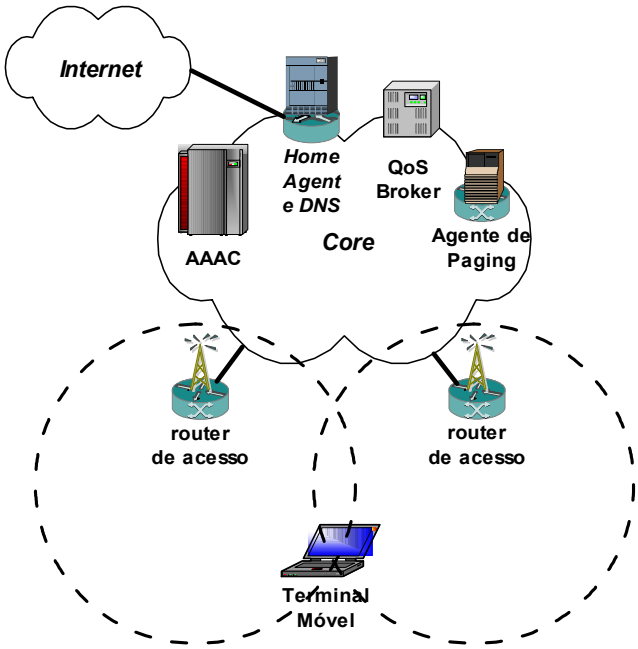
Universidade de Aveiro Departamento de Electrónica e Telecomunicações
2005

Victor Manuel Letra Macedo Marques

Serviços Multimédia Sobre Redes Heterogéneas




Universidade de Aveiro Departamento de Electrónica e Telecomunicações
2005

Victor Manuel Letra Macedo Marques

Serviços Multimédia Sobre Redes Heterogéneas

Tese apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Engenharia Electrotécnica, realizada sob a orientação científica do Professor Doutor A. Manuel de Oliveira Duarte, Professor Catedrático do Departamento de Electrónica e Telecomunicações da Universidade de Aveiro e co-orientação do Professor Doutor Rui Luís Andrade Aguiar, Professor Auxiliar do Departamento de Electrónica e Telecomunicações da Universidade de Aveiro.

O júri

Presidente

Doutora Ana Maria Vieira da Silva Viana Cavaleiro
Professora Catedrática da Universidade de Aveiro, por delegação da Reitora da Universidade de Aveiro

Doutor Augusto Júlio Domingues Casaca
Professor Catedrático do Instituto Superior Técnico da Universidade Técnica de Lisboa

Doutor Aníbal Manuel de Oliveira Duarte
Professor Catedrático da Universidade de Aveiro (Orientador)

Doutor Joaquim Arnaldo Carvalho Martins
Professor Catedrático da Universidade de Aveiro

Doutor Alexandre Júlio Teixeira dos Santos
Professor Associado da Universidade do Minho

Doutor Rui Luís Andrade de Aguiar
Professor Auxiliar da Universidade de Aveiro (Co-Orientador)

Agradecimentos

Ao Professor Manuel Oliveira Duarte pela oportunidade que me deu, pelas condições que me proporcionou para a realização deste trabalho e pelo apoio e ensinamentos prestados durante toda a minha vida académica.

Ao Professor Rui Aguiar pelo excelente relacionamento e amizade, discussões produtivas, conselhos e orientações, e por todo o restante apoio prestado na revisão atenta e cuidada desta tese.

Ao Professor Amaro de Sousa por todo o apoio e conhecimentos prestados, em especial na fase inicial do programa de doutoramento, pelo excelente relacionamento e amizade e ainda pela motivação transmitida.

Ao Doutor Francisco Fontes pela amizade e confiança em mim depositada, que permitiram conduzir o trabalho de doutoramento num ambiente favorável.

Ao Eng.º Pedro Gonçalves, pelo apoio prestado em especial na consolidação de um demonstrador para extracção de resultados.

Aos Eng.ºs Diogo Gomes, Nuno Duarte, Nuno Sénica e João Barraca pela preparação do demonstrador e pela ajuda na obtenção dos resultados.

A todo o consórcio do projecto Moby Dick, em especial ao Eng. Hans Einsiedler pelas condições favoráveis ao desenvolvimento do trabalho.

Ao meu amigo Eng.º Ricardo Cadime pela grande ajuda e apoio, em especial na fase inicial deste trabalho. Ao meu amigo Eng. Paulo Salvador pelo apoio e disponibilidade.

Ao Departamento de Electrónica e Telecomunicações da Universidade de Aveiro e ao Instituto de Telecomunicações, pólo de Aveiro, pelas condições que me permitiram concretizar parte deste trabalho.

À Fundação para a Ciência e Tecnologia pela bolsa de doutoramento que me atribuiu no âmbito do programa PRAXIS XXI.

À PT Inovação, em especial ao Eng.º Paulo Nordeste, pela possibilidade concedida de participar em projectos de investigação, importantes no desenvolvimento do trabalho e pela criação de condições favoráveis à prossecução do trabalho de doutoramento num ambiente industrial. Agradeço ainda o apoio financeiro concedido no âmbito desta formação.

À NEC (Heidelberg), em especial ao Eng. Xavier Perez Costa pelo contributo dado em termos de suporte no código de simulação.

A todos os meus colegas do Grupo de Sistemas de Banda Larga, do Instituto de Telecomunicações e da Portugal Telecom Inovação, pelo excelente ambiente de trabalho e amizade que sempre me ofereceram.

A todos os meus grandes amigos pela paciência que sempre manifestaram quando, nas partes mais difíceis do trabalho, não pude estar presente.

Aos meus Pais, irmãs e restantes familiares, pela constante motivação, apoio e confiança, carinho e paciência, e pelo pouco tempo e atenção que lhes pude dispensar durante estes anos.

À Susana por todos os sacrifícios, pela minha falta de disponibilidade, pelos fins-de-semana, feriados e noites "perdidos", por todo o apoio e motivação transmitidos, pelo exemplo de trabalho e dedicação e por todo o amor e carinho sempre presentes.

Resumo

Esta Tese aborda a problemática do fornecimento de serviços multimédia em redes heterogéneas. A heterogeneidade é aqui apresentada tanto na sua vertente de diferentes tecnologias de acesso como de vários fornecedores de serviço.

Na génese do trabalho apresentado nesta tese está a convergência de redes e serviços que hoje em dia se observa. Esta convergência de acessos fixos e móveis sobre uma infraestrutura comum, baseada na utilização do protocolo IP acarreta consigo alguns problemas que urge analisar e resolver. Esta tese centra-se na provisão transparente de serviços, num ambiente "all-IP", em cenários com suporte de mobilidade transparente (para utilizadores e aplicações), inter e intra tecnologias, num ambiente com vários fornecedores de serviço.

Esta tese está dividida em três partes distintas.

Na primeira parte (Capítulo 2) é abordada a problemática do fornecimento de serviço nas suas diversas vertentes, sendo feito um resumo histórico do fornecimento do serviço de telecomunicações desde as suas origens. É analisada a problemática do suporte de mobilidade, de *Authentication*, *Authorisation*, *Accounting* and *Charging* (AAAC), de *Service Level Agreements* (SLAs) e *Service Level Specifications* (SLSs), de qualidade de serviço (QoS), de gestão e ainda alguns aspectos comerciais. São apresentados alguns desafios que surgem quando se juntam todos estes componentes com vista a operacionalizar o fornecimento de serviço em ambientes heterogéneos integrados.

Na segunda parte (Capítulo 3 e Capítulo 4) são apresentadas as principais tecnologias existentes hoje em dia que servem de base para a construção de arquitecturas de fornecimento de serviço em ambientes futuros.

Na terceira parte (Capítulo 5 e Capítulo 6) é apresentada e avaliada uma arquitectura para o fornecimento de serviços multimédia, num ambiente de heterogeneidade de tecnologias de acesso e de fornecedores. Esta arquitectura suporta a provisão de qualidade de serviço (fulcro desta tese) mesmo enquanto os terminais se deslocam entre diversas redes com serviços activos. A arquitectura baseia-se na utilização do protocolo IPv6 como elemento integrador, e toda a sinalização é feita sobre este protocolo, tornando-a assim independente das tecnologias de nível 2 utilizadas. É proposta a sinalização entre as diversas entidades, de modo a garantir o funcionamento da arquitectura e a total transparência de *handover*, quer do ponto de vista do utilizador, quer do ponto de vista das aplicações.

São identificados os cenários críticos aos quais a arquitectura e sinalização propostas devem responder por forma a ser validada. São depois apresentados os bons resultados obtidos num ambiente de simulação e em testes numa implementação real. Estes resultados reforçam a validade da concepção teórica realizada.

Abstract

This thesis addresses multimedia service provisioning in heterogeneous environments. Heterogeneous here refers both to access technologies and service providers.

The motivation of the work here presented is the current convergence of networks and services. The convergence of fixed and wireless on one single network infrastructure has several problems that must be analysed and solved. This thesis is centred in the service provisioning, on an “all-IP” environment, supporting seamless mobility (both for users and applications), inter and intra technologies, on a multi-provider environment.

This thesis is divided into three different parts.

The first part (chapter 2) starts by presenting an historical view of the telecommunications service provision and addresses the several problems associated with service provisioning. Several issues, such as Authentication, Authorisation, Accounting and Charging (AAAC), Service Level Agreements (SLAs) and Service Level Specifications (SLSs), quality of service (QoS), management and marketing are analysed. It finalises by presenting some challenges of merging all these components in order to enable service provision on integrated heterogeneous environments.

The second part of this thesis (chapter 3 and 4) presents several existing technologies that are used as building blocks for next generation network architectures.

The third part of the thesis (chapter 5 and 6) presents a next generation network architecture and its evaluation. This architecture supports quality of service provision (the main topic of this thesis) even while the terminals move with active services across different access technologies. This architecture is based on IPv6 and all signalling is done at the IP level. This way, the architecture and signalling are independent of the layer 2 technologies. The signalling between the architecture entities is presented. The signalling was defined in a way to allow seamless *handovers* both to the users and applications.

Critical scenarios to validate both the architecture and signalling are identified and presented. The good evaluation results obtained (both in a simulation environment and on a real test bed) are then presented.

Tabela de Conteúdos

Lista de figuras

Lista de tabelas

CAPÍTULO 1

INTRODUÇÃO

Na última década assistimos a um crescimento exponencial da *Internet* e das redes de comunicação celulares. Foi uma década muito activa também na procura de uma tecnologia integradora, que permitisse aos operadores manterem apenas uma infra-estrutura de rede que desse suporte a todos os tipos de serviço de telecomunicações. O aumento da capacidade de processamento, substanciado na microelectrónica, torna agora realizáveis estes desejos antigos dos operadores.

O estado actual das redes de telecomunicações reflecte uma evolução histórica, nem sempre perfeitamente estruturada. As motivações de evolução são quase sempre fortemente condicionadas por aspectos económicos, introduzindo várias condicionantes negativas [Borr02]. Presentemente temos ao nosso dispor serviços de telecomunicações distintos, suportados em redes também elas distintas. Nem sempre estes serviços são prestados sobre a infraestrutura ou tecnologia mais adequada (por exemplo o acesso à *Internet* sobre GSM - *Global System for Mobile communications* ou PSTN - *Public System Telephone Network* com recurso a *modems*). Contudo, a tendência actual aponta para uma uniformização e convergência de redes e tecnologias (*Internet*, telefonia fixa e celular, *broadcast*, etc.). Esta convergência necessita de uma estreita colaboração e combinação

entre as tecnologias "*Internet*" e as telecomunicações tradicionais para atingir o acesso ubíquo, disponibilidade e robustez. Nas décadas passadas, cada serviço era suportado pela sua própria rede de transporte. Hoje em dia, há já redes que suportam a provisão de serviços distintos (por exemplo as redes de televisão por cabo, que para além do tradicional serviço de difusão de televisão, fornecem também serviços de dados interactivos e serviços de voz). No entanto, na maioria destes casos, apesar da infra-estrutura física ser a mesma, os serviços são suportados em infra-estruturas lógicas distintas.

Presentemente, associam-se os estágios evolutivos das redes e serviços (no geral) à evolução da rede móvel. As várias fases evolutivas, com marcos bem identificados, designam-se de "gerações". Actualmente situamo-nos num período entre as chamadas "segunda" (2G) e "terceira" (3G) gerações, que vulgarmente de designa por "2,5G". É um ambiente misto. Nas redes celulares a situação presente apresenta diferentes estruturas lógicas: existe um serviço de dados e um serviço de voz, com infraestruturas lógicas separadas. As redes de terceira geração representam um avanço no sentido de uma maior convergência. O serviço de voz será tendencialmente suportado em comutação de pacotes, levando à convergência de dois mundos: voz e dados. Esta é uma problemática que abrange todas as zonas da rede: cliente, acesso, fronteira e núcleo.

O desenvolvimento observado ao nível das capacidades tecnológicas torna possível que se concentrem num mesmo dispositivo diversas interfaces de acesso a redes de tecnologias distintas. Contudo, hoje em dia ainda não é possível fazer um uso transparente das várias tecnologias que um terminal tem ao seu dispor, devido a esta separação das interfaces lógicas.

As redes de "próxima geração" serão redes em que haverá a total convergência de serviços sobre uma única infra-estrutura. Esta infra-estrutura incluirá todas as tecnologias de acesso disponíveis. Todos os serviços serão suportados com base em apenas um protocolo de transporte, o *Internet Protocol* (IP) [IP81]. Será possível usar qualquer das tecnologias de acesso disponíveis para aceder a determinado serviço, e será possível manter um serviço activo mesmo havendo alteração no meio (tecnologia) de acesso. É então necessário definir uma arquitectura de rede que satisfaça estes requisitos.

Esta tese versa o tema do fornecimento de serviços multimédia em redes heterogéneas, considerando as suas diferentes vertentes, ou seja, heterogeneidade de tecnologias, heterogeneidade de fornecedores e heterogeneidade de serviços. O trabalho

realizado acompanhou de perto o percurso das tendências evolutivas nas redes de telecomunicações. Até perto do final da década de 90, assumiu-se que o *Asynchronous Transfer Mode* (ATM) [ATMF] seria a tecnologia eleita para fazer a integração dos mundos dos circuitos e dos pacotes. Foi assim com naturalidade que o tema de trabalho inicial versava a provisão de serviços multimédia em redes ATM. Neste âmbito, para além do ATM, foram estudados protocolos como o MPEG4 (*Moving Picture Experts Group* – versão 4) [MPEG4a], nomeadamente a sua parte 6, o *Delivery Multimedia Integrated Framework* (DMIF) [MPEG4b] e foram implementados *drivers* que permitiam que aplicações multimédia estabelecessem sessões ATM com determinados requisitos de qualidade de serviço, controlados ao nível da aplicação. Foram também realizados estudos de mapeamento de níveis de qualidade de serviço em parâmetros ATM. Todavia, o sucesso da *Internet* e a sua associação ao protocolo IP, motivaram o estudo de outras estratégias (vide capítulo 3). Assim, o trabalho realizado no âmbito desta tese, foi redireccionado para o estudo do protocolo IPv6 (*Internet Protocol version 6*), das suas características e da sua utilização em cenários de rede de próxima geração. Os maiores contributos do trabalho aqui apresentado acabaram por se situar neste contexto das redes de próxima geração, com a integração de conteúdos e serviços, num ambiente "*all-IP*". O facto do trabalho ter sido parcialmente realizado integrado num ambiente industrial, influenciou naturalmente as ideias desenvolvidas, enriquecendo-as com cenários reais de interesses de provisão de serviço por parte de um operador.

No decurso dos diferentes capítulos desta tese serão abordados temas distintos como a mobilidade, a autenticação e autorização, e o tema unificador da qualidade de serviço (QoS). Dentro da problemática do fornecimento de serviços, do ponto de vista da rede, o suporte de qualidade de serviço é um factor determinante para o sucesso dos serviços. Neste sentido, o trabalho realizado passou pela definição de uma arquitectura genérica (com ênfase na sua componente de QoS), incluindo funcionalidades e interacções (incluindo sinalização) entre diferentes entidades, suportando de uma forma transparente a mobilidade de utilizadores entre diversas tecnologias de acesso. A arquitectura apresentada visa assim responder à problemática do fornecimento de serviços heterogéneos em redes também elas heterogéneas. A arquitectura proposta é totalmente baseada em IP, sendo no decurso da tese, apresentados os motivos desta opção.

Este trabalho culminou com a simulação, implementação e teste das entidades e arquitectura definidas, que mostraram o seu bom desempenho e a viabilidade de se poder tornar um modelo de referência para redes futuras.

1.1 Estrutura da Tese

Nos próximos parágrafos é descrita a forma como esta tese está organizada.

O capítulo 2 versa a problemática do fornecimento do serviço nas suas diversas componentes. No início do capítulo é apresentada uma breve resenha histórica do fornecimento do serviço, seguida por uma discussão acerca das várias vertentes do fornecimento de serviço de telecomunicações. Este capítulo tem uma função enquadradora, apresentando uma reflexão sobre aspectos pertinentes para redes multiserviços da próxima geração – uma área em que a reflexão em si afecta as escolhas arquitecturais a efectuar. Assim a perspectiva das redes de comunicações reflectida nesta análise tem impacto directo em termos de especificações para o trabalho apresentado nas secções seguintes. A abordagem utilizada é de alto nível, tentando mostrar frequentemente como os requisitos de telecomunicações estão ligados à nossa sociedade.

O capítulo 3 foca-se no suporte de serviços em redes heterogéneas. Começa com uma breve discussão crítica da evolução tecnológica da última década, avançando depois para uma breve apresentação de diversas tecnologias de rede, de suporte ao fornecimento de serviço e de interligação multimédia, fundamentalmente do ponto de vista das tecnologias IP. De notar que esta discussão foi fundamentada em diversos casos por trabalho de investigação e desenvolvimento de diferentes aspectos técnicos (por exemplo o realizado sobre interfaces aplicacionais para sistemas multiserviços). Este capitulo pretende mostrar o historial tecnológico que dominará os desenvolvimentos arquitecturais futuros.

No capítulo 4 são apresentadas especificamente tecnologias e conceitos que determinam o controlo e gestão de qualidade de serviço em redes IP. No início é apresentada uma visão estruturada (em camadas) da qualidade de serviço, sendo depois apresentados os diversos mecanismos de oferta de qualidade de serviço em redes IP. Este capítulo apresenta ainda as estratégias de qualidade de serviço nos diversos pontos da rede, assim como algumas estratégias de controlo de qualidade de serviço extremo-a-extremo,

passando pela discussão de cenários de operador único ou de múltiplos operadores, desenvolvidos durante este trabalho de doutoramento.

No capítulo 5 é feita uma proposta de uma arquitectura genérica para provisão de serviços num ambiente de próxima geração. Neste sentido, é apresentada a arquitectura, são definidos os seus principais elementos, as suas componentes, e suas interacções. É ainda apresentada e discutida a sinalização nos cenários de registo, autorização e de *handover*. O capítulo termina com uma discussão sobre alguns dos aspectos mais relevantes em termos de fornecimento de serviço, do modo como foram abordados, das suas vantagens e eventuais lacunas.

No capítulo 6 é apresentada a formulação da validação da arquitectura proposta; são também apresentados resultados de simulação do processo de *handover*, assim como os resultados da implementação prática da arquitectura. O capítulo contem ainda várias considerações e a discussão dos resultados obtidos.

Finalmente, o capítulo 7 apresenta as principais conclusões do trabalho realizado, uma análise breve a respeito de arquitecturas alternativas, bem como os tópicos em aberto considerados mais importantes e desenvolvimentos industriais baseados no trabalho realizado.

1.2 Enquadramento do trabalho

Conforme foi referido atrás, o trabalho desenvolvido ao longo do percurso do programa de doutoramento foi em grande parte condicionado pelas tendências mundiais de evolução das redes de telecomunicações. O trabalho aqui apresentado surge apoiado numa sequência de participações em projectos de investigação europeus e também em projectos de desenvolvimento de produtos e soluções, que espelham essa mesma evolução das redes. De seguida são apresentados os projectos, por ordem cronológica, aos quais e dos quais o programa de doutoramento aqui apresentado deu e recebeu contributos:

- Jupiter II – Projecto Eurescom P807-GI [Jupi] – O trabalho realizado no âmbito deste projecto visou o desenvolvimento de *drivers* que fizessem o mapeamento de qualidade de serviço, definida de uma forma genérica ao nível da aplicação, em parâmetros de qualidade de serviço ATM [VMar01];

- Armstrong – Projecto Eurescom P1009 [Arms] – O trabalho realizado no âmbito deste projecto visou o estudo do protocolo IPv6, nas suas vertentes práticas, assim como a projecção de cenários de transição entre as redes IPv4 e as redes IPv6 e o estudo de soluções, utilizando os mecanismos de transição existentes, para que a migração seja o mais suave possível [Kram01].
- Moby Dick – Projecto Europeu IST-2000-25394 [Moby] – Foi no âmbito deste projecto que as principais componentes do trabalho de doutoramento foram realizadas e correspondeu aos três últimos anos desse trabalho. Este projecto visou a definição e implementação de uma arquitectura de rede de próxima geração, que possibilitasse a mobilidade transparente entre células de tecnologias distintas (heterogeneidade de acesso), em ambiente de operador de telecomunicações.
- Acções de consultoria a empresas do Grupo PT – no percurso do trabalho deste doutoramento, e no âmbito da actividade profissional paralela, a participação em acções de consultoria (a configurações de rede e estratégias de evolução) a diversas empresas do Grupo PT permitiu um maior conhecimento das redes dos operadores, do seu modo de operação e das preocupações de evolução. Este conhecimento permitiu a definição e o encaminhamento de determinadas soluções para caminhos que visam responder às necessidades e perspectivas dos operadores.
- End-to-end QoS – Este foi um projecto do Contrato de Inovação do Grupo Portugal Telecom que permitiu um conhecimento mais alargado das tecnologias de acesso fixas (por exemplo, o ADSL – *Asymmetrical Digital Subscriber Line*) e dos mecanismos de controlo e gestão de qualidade de serviço. Foi também possível alargar os conhecimento sobre a gestão e controlo de SLAs (*Service Level Agreements*) e SLSs (*Service Level Specifications*) e, simultaneamente, conhecer as estratégias dos operadores neste domínio.

1.3 Principais contribuições

Tendo este trabalho de doutoramento sido enriquecido pelo seu enquadramento em actividades arquitecturais de maior âmbito, é sempre de realçar o contributo individual do mesmo e como se interrelacionou com as restantes componentes das arquitecturas. As principais contribuições do trabalho realizado no âmbito do programa de doutoramento podem ser enumeradas do seguinte modo:

- Componente de Qualidade de Serviço (cenários intra e inter-operador, entidades funcionais e sinalização) de uma arquitectura de próxima geração, completamente desenvolvida no âmbito de um projecto europeu (Moby Dick) .
- Integração de QoS em mecanismos existentes de sinalização de autenticação, autorização e *handover* em cenários de mobilidade IP com QoS e AAAC (*Authentication, Authorization, Accounting and Charging*).
- Proposta de arquitectura, funcionalidades e modo de operação de um QoS *Broker* – o QoS *Broker* utilizado num cenário de próxima geração foi fundamentalmente definido no âmbito do trabalho deste doutoramento. De notar que a implementação desta entidade foi realizada num trabalho separado.
- Especificação e desenvolvimento de módulos de *software* no âmbito de projectos internacionais (nomeadamente interfaces DMIF para redes ATM).

O trabalho realizado foi reportado num conjunto de artigos em revistas e conferências:

- Victor Marques, Xavier Perez Costa, Rui L. Aguiar, Marco Liebsch, A. Manuel Oliveira Duarte, “Evaluation of a Mobile IPv6-based Architecture Supporting User Mobility QoS and AAAC in Heterogeneous Networks”, IEEE Journal on Selected Areas in Communications, Special Issue on Wireless Overlay Networks Based on Mobile IPv6 (aceite para publicação)
- A. Cuevas Casado, P. Serrano Yanez-Mingot, J. I. Moreno Novella, C. J. Bernardos, J. Jaehnert, R. L. Aguiar, V. Marques, "Usability and Evaluation of a Deployed 4G Network Prototype", Towards the Next Generation Mobile Communications, Journal of Communications and Networks (aceite para publicação)

- Juergen Jaehnert , Jie Zhou, Rui L. Aguiar, Victor Marques, *et al*, "*Moby Dick: A Pure-IP 4G Architecture"*, Computer Communications, Elsevier Computer Communications, Vol 28/9 pp 1014-1027, 2005
- P. Gonçalves, D. Gomes, V. Marques, R. Aguiar, "*QoS control support for heterogeneous networks*", CRC 2003, Bragança, Portugal, Setembro 2003.
- Victor Marques, Rui L. Aguiar, Carlos Garcia, Jose Ignacio Moreno, Christophe Beaujean, Eric Melin, Marco Liebsch, "*An IP-based QoS architecture for 4G operator scenarios*", IEEE Wireless Communications Magazine, Junho de 2003.
- Antonio Cuevas, José Ignacio Moreno, Rui Aguiar, Victor Marques, Carlos García, Ignacio Soto, "*Mechanisms for AAA and QoS Interaction*", ASWN 2003, Zurique, Suíça, Maio de 2003
- Jacinto Vieira, Victor Marques, Carlos Parada, Carlos Rodrigues, Francisco Fontes, "*A Glance of the Current IPv6 Status*", ConfTele 2003, Aveiro, Portugal, 18-20 Junho de 2003
- Isabel Borges, Carlos Rodrigues, Victor Marques, Francisco Fontes, "*Voice Over Next Generation Networks*", ConfTele 2003, Aveiro, Portugal, 18-20 Junho de 2003
- Victor Marques, Carlos Parada, Pedro Gonçalves, Rui L. Aguiar, Francisco Fontes, "*Next Generation Network Provider Architecture Demonstrator*", ConfTele 2003, Aveiro, Portugal, 18-20 Junho de 2003
- Christophe Beaujean, Nesrine Chaher, Victor Marques, Rui L. Aguiar, Carlos García, José Ignacio Moreno, Michelle Wetterwald, Thomas Ziegler, "*Implementation and Evaluation of an End-to-End IP QoS Architecture for Networks Beyond 3rd Generation*", IST Mobile Summit 2003, Aveiro, Portugal, Junho de 2003
- Janusz Gozdecki, Piotr Pacyna, Victor Marques, Rui L. Aguiar, Carlos Garcia, Jose Ignacio Moreno, Christophe Beaujean, Eric Melin, Marco Liebsch, "*An IP QoS architecture for 4G networks*", "Art-QoS 2003", Warsaw, Poland, Março de 2003
- Victor Marques, Rui L. Aguiar, Antonio Cuevas Casado, Jose Ignacio Moreno, Nesrine Chaher, "*A simple QoS service provision framework for*

beyond 3rd generation scenarios", "10th International Conference on Telecommunications - ICT'2003", Tahiti, Papeete, French Polynesia, 23 de Fevereiro a 1 de Março de 2003.

- Victor Marques, Francisco Fontes, Jacinto Vieira e Carlos Parada, "Uma Arquitectura para Integração de Mobilidade e Qualidade em Redes IPv6 de Operadores", revista "Saber e Fazer Telecomunicações", PT Inovação, 2003
- Francisco Fontes, Victor Marques, Jacinto Vieira e Carlos Parada, "Mobilidade e Qualidade em Redes IPv6", revista "Saber e Fazer Telecomunicações", Portugal Telecom Inovação, 2003
- A. Cardoso, P. Gonçalves, M. Fernandes, P. Mendes, A. Gamelas e V. Marques, "Serviços e Tecnologia de Banda Larga na Rede de Acesso", revista "Saber e Fazer Telecomunicações", PT Inovação, 2003
- Victor Marques, Rui L. Aguiar, Piotr Pacyna, Janusz Gozdecki, Christophe Beaujean, Nesrine Chaher, Carlos García, José Ignacio Moreno, Hans Einsiedler, "*An Architecture Supporting End-to-End QoS with User Mobility for Systems Beyond 3rd Generation*"; "IST Mobile Summit 2002", Tessalónica, Grécia, 16 a 19 de Junho de 2002
- Victor Marques, Rui Aguiar, Jürgen Jähnert, Karl Jonas, Marco Liebsch, Hans Einsiedler, Francisco Fontes; "*An Heterogeneous Mobile IP QoS-aware Network*"; "CRC 2001", 29 e 30 de Novembro de 2001
- Victor Marques, Rui Aguiar, Francisco Fontes, Jürgen Jähnert, Hans Einsiedler ; "*Enabling IP QoS in Mobile Environments*"; "IST Mobile Summit 2001", Barcelona, Espanha, 9 a 12 de Setembro de 2001
- Victor Marques, Ricardo Cadime, Amaro Sousa, A. Oliveira Duarte, "*DMIF based QoS Management for MPEG-4 Multimedia Streaming: ATM and RSVP/IP Case Studies*"; ConfTele 2001, Figueira da Foz, Portugal, Abril de 2001
- B. Júnior, V. Marques, A. Sousa, A. Oliveira Duarte, "Integração de uma plataforma ATM na rede local da Universidade de Aveiro – um caso de estudo de encaminhamento em redes heterogéneas", 1ª Conferencia em Redes de Computadores (CRC'98), Coimbra, Portugal, 9-10 Novembro 1998

CAPÍTULO 2

FORNECIMENTO DE SERVIÇO DE TELECOMUNICAÇÕES

2.1 Introdução

O fornecimento do serviço de telecomunicações tem pouco mais de um século de existência. Pode-se dizer que até há três décadas atrás, os serviços de telecomunicações podiam resumir-se ao serviço de telefonia (telefone e rádio), de distribuição de televisão, e um serviço de dados muito rudimentar. No entanto, nas três últimas décadas e em especial na última, ocorreram mudanças profundas no mundo das telecomunicações. Assistimos a uma verdadeira revolução, não tanto pelo desenvolvimento tecnológico, mas pelo agregar de conceitos e tecnologias em novas formas de utilização na sociedade. Estas mudanças foram por um lado motivadas por uma sociedade cada vez mais exigente, mas por outro lado estão a ser também as charneiras de modificação dessa mesma sociedade.

A mobilidade e o estar sempre contactável representam aspectos que se tornaram comuns no quotidiano social actual. É indiscutível o sucesso do serviço móvel terrestre de

voz (os populares telemóveis), que veio permitir a comunicação falada entre quaisquer pessoas, potencialmente localizadas em qualquer ponto do mundo, desde que possuindo um terminal e a rede de acesso adequados. O mesmo sucesso se pode constatar em relação à *Internet* que possibilita uma comunicação fácil, em tempo real ou não, entre duas ou mais pessoas, e possibilita ainda o acesso a uma quantidade de informação até há poucos anos inimaginável.

Já no respeitante à qualidade (dos diversos serviços e de vida), é um aspecto que suscita preocupações crescentes. Os consumidores são mais exigentes, querendo mais e melhor pelo mesmo (ou menor) preço. Como tal, os prestadores de serviço têm actualmente como principal objectivo satisfazer um número crescente de clientes, mantendo ou mesmo melhorando o nível e variedade de serviços prestados, ao mesmo tempo que tentam reduzir custos de operação.

A situação actual encontra-se balizada por uma evolução histórica da tecnologia e da sociedade, e o futuro terá que contemplar as necessidades sentidas actualmente. Na realidade, esta visão das redes de telecomunicações corresponde a um exercício de reflexão sobre as especificações adequadas para futuras redes de telecomunicações, tendo especial interesse para este trabalho de doutoramento o impacto destas especificações em termos de suporte de redes multiserviços.

2.2 Redes de telecomunicações: passado, presente e futuro

2.2.1 As tecnologias e as infraestruturas de suporte: do telégrafo às fibras ópticas

Em 1794, Claude Chappe inventou o "Tachygraphe", que consistia num sistema de semáforos com um conjunto de braços móveis com os quais se comunicava remotamente, em linha de vista. Eram precisos vários homens para operar cada um dos dispositivos, era altamente dependente das condições atmosféricas e a velocidade de transmissão era de cerca de 15 caracteres por minuto! Mesmo com estas limitações, este foi o primeiro serviço de telecomunicações a ser prestado, sob ordem do então actual ministro da guerra francês, que ordenou também que se chamasse "telégrafo" ao dispositivo. A palavra telégrafo deriva de duas palavras gregas, "tele" que significa "longe" e "graphein", que significa "escrever". Este serviço estava disponível para envio de mensagens entre Paris e Lille mas

apenas ao serviço do estado. Rapidamente estes dispositivos foram sendo instalados e espalharam-se pela Europa e Estados Unidos, mesmo tendo enormes limitações (não funcionavam de noite e eram problemáticos em dias chuvosos ou com nevoeiro, por exemplo).

Em 1837, William Fothergill Cooke (1806-1879) e o Professor Charles Wheatstone (1802-1875) desenvolveram e patentearam um telégrafo baseado na descoberta de 1832 do Barão Pavel Lvovitch Schilling (1780-1836) que consistia num sistema de 5 agulhas que, por efeito electromagnético, seleccionavam um conjunto de 20 caracteres disponíveis. É assim que em 1839 surge o primeiro serviço comercial de telecomunicações, transmitindo mensagens a cerca de 21 quilómetros entre Paddington e West Drayton. O público em geral podia pagar 5 xelins para ver o dispositivo em operação e podia também enviar as suas próprias mensagens. Em 1841 Wheatstone inventou o primeiro telégrafo com escrita, e propôs também pela primeira vez um sistema de multiplexagem temporal. Dois anos volvidos (em 1843), Alexander Bain (1810-1877) desenvolveu e patenteou a transmissão por Facsimile (Fax)! No entanto, esta tecnologia não teve na altura seguimento como serviço comercial.

Apesar dos diversos tipos de "telégrafos" inventados, a palavra telégrafo veio mais tarde a ser aplicada apenas ao dispositivo de transmissão de impulsos eléctricos desenvolvido e demonstrado por Samuel Finley Breeze Morse (1791-1872) em 1844, seguindo uma ideia inicialmente apresentada em 17 de Fevereiro de 1753 por um correspondente da publicação "Scots Magazine" que assinou apenas C M. A primeira mensagem enviada por Morse, entre Washington e Baltimore (cerca de 65 quilómetros) foi "*What hath God wrought*" ("O que Deus fez"). Morse utilizou um equipamento inteiramente desenvolvido por si, totalmente diferente do de Cooke e Wheatstone. Morse utilizou o código que ainda hoje é utilizado e ficou conhecido pelo seu nome. Em apenas 6 anos, foram dezenas as linhas de telégrafo instaladas e a entrar em funcionamento para fornecimento de serviço público.

O telégrafo revolucionou o conceito de serviço público de telecomunicações. Em 1850 foi instalado o primeiro cabo submarino e em 1858 foi instalado o primeiro cabo submarino transatlântico, permitindo pela primeira vez na história a comunicação expedita entre a Europa e a América. Em 1872 foi transmitida pela primeira vez uma mensagem de telégrafo entre Londres e Melbourne na Austrália, envolvendo a retransmissão em cada

uma das 18 estações intermédias. O tempo médio para transmissão de uma mensagem entre estas cidades era de 20 horas. Em 1882, W. H. Preece, fez uma experiência de telégrafo sem fios entre Southampton e Newport tendo em 12 de Dezembro de 1901 sido feita a primeira transmissão de telégrafo sem fios sobre o Atlântico, onde Marconi ouviu o sinal previamente combinado: a letra "S" em código Morse.

Outros desenvolvimentos ocorreram durante este período. Enquanto professor, Alexander Graham Bell (1847-1922) começou a experimentar vários dispositivos para ajudar surdos a falar. Durante o seu trabalho, surgiu-lhe a ideia do "telefone" . Em 1874 tinha o conceito básico desenvolvido e dois anos mais tarde, a 10 de Março de 1876, Bell conseguiu enviar pela primeira vez uma mensagem de voz por fios eléctricos (cerca de 30 metros) ao seu assistente que estava na sala do lado: "*Mr. Watson, come here; I want you.*". Curiosamente Bell submeteu a sua patente do telefone apenas 3 horas antes de Elisha Gray (1835-1901) que na altura trabalhava para a "*Telegraph Company*" – antecessora da "*Western Union*". Bell conseguiu a patente e ofereceu a sua venda à "*Telegraph Company*". A IEEE *Systems* "*Transactions on Man & Cybernetics Society*" numa das suas edições mais antigas publicou o seguinte artigo, que pela sua significância fica aqui transcrito na sua versão original [Bill]:

In 1876, Alexander Graham Bell and his financial backer, Gardiner G. Hubbard, offered Bell's brand new patent (No. 174,465) to the Telegraph Company - the ancestor of Western Union. The President of the Telegraph Company, Chauncey M. DePew, appointed a committee to investigate the offer. The committee report has often been quoted. It reads in part:

"The Telephone purports to transmit the speaking voice over telegraph wires. We found that the voice is very weak and indistinct, and grows even weaker when long wires are used between the transmitter and receiver. Technically, we do not see that this device will be ever capable of sending recognizable speech over a distance of several miles.

"Messer Hubbard and Bell want to install one of their "telephone devices" in every city. The idea is idiotic on the face of it. Furthermore, why would any person want to use this ungainly and impractical device when he can send a messenger to the telegraph office and have a clear written message sent to any large city in the United States?

"The electricians of our company have developed all the significant improvements in the telegraph art to date, and we see no reason why a group of outsiders, with extravagant and impractical ideas, should be entertained, when they have not the slightest idea of the true problems involved. Mr. G.G. Hubbard's fanciful predictions, while they sound rosy, are based on wild-eyed imagination and lack of understanding of the technical and economic facts of the situation, and a posture of ignoring the obvious limitations of his device, which is hardly more than a toy... .

"In view of these facts, we feel that Mr. G.G. Hubbard's request for $100,000 of the sale of this patent is utterly unreasonable, since this device is inherently of no use to us. We do not recommend its purchase."

O mais caricato desta carta, vista agora em retrospectiva, é o facto de apenas 4 anos mais tarde Bell ter conseguido o controlo da companhia à qual tentou vender a sua patente. Nesta era de constantes evoluções tecnológicas, esta é uma história cuja lição ainda não foi inteiramente compreendida.

O crescimento das redes telefónicas nos anos seguintes ao da invenção de Bell foi enorme, e mesmo hoje em dia novos utilizadores continuam a surgir todos os dias um pouco por todo o mundo.

Em 1903 foi introduzido pela primeira vez o conceito de horário económico nas chamadas telefónicas, permitindo que os clientes pudessem fazer uma chamada do dobro do tempo normal, pagando o mesmo.

Um inventor canadiano, de seu nome Reginald Fessenden, veio dar um outro contributo importante para a nossa realidade actual. Em 1900 ele transmitiu, via rádio, a primeira mensagem de voz. Durante os seis anos seguintes aperfeiçoou a sua invenção e na noite de Natal de 1906, fez a primeira teledifusão de rádio da história.

Em 1920, Sir Samuel Instone, a partir da sua residência em Londres, realizou uma conversa telefónica com um avião em voo para Paris.

Outro facto marcante foi o registo da patente da televisão em 1884 pelo alemão Paul Nipkow que culminou em 1926, com John Logie Baird a emitir a primeira imagem parada de um quarto para outro, dentro da mesma casa. Um ano mais tarde, Baird conseguiu enviar uma sequência de imagens de Londres para Glasgow, utilizando a rede telefónica, sendo realizada em 1928 a primeira transmissão televisiva transatlântica.

Em 1927 a Bell System inaugurou o primeiro serviço telefónico transatlântico, utilizando uma ligação de rádio bidireccional, entre os Estados Unidos e Inglaterra. Estas chamadas custavam 75 dólares por cada 3 minutos de conversação. Em 1935, dois executivos da Bell System fizeram uma chamada de um para o outro, a partir de dois gabinetes contíguos em Nova York. A particularidade da chamada foi o facto da ligação ter sido estabelecida à volta do mundo. No entanto, o serviço via rádio estava longe de ser ideal e em 1956 o serviço telefónico entre os Estados Unidos e a Europa passou a ser fornecido através de um cabo submarino.

Em 1947, nos Estados Unidos, é utilizado o primeiro sistema de transmissão (com capacidade para 2.400 chamadas simultâneas) por microondas entre duas cidades (Boston e Nova York). A capacidade de transmissão por microondas instalada cresceu consideravelmente nas três décadas seguintes, quer pela quantidade de antenas instaladas, quer pelo aumento de capacidade de cada ligação (cerca de 19.200 chamadas simultâneas). Os sistemas via rádio, nos Estados Unidos passaram a transportar cerca de 70% do tráfego de voz. Os sistemas terrestres (cabo coaxial) eram utilizados para a quase totalidade do tráfego de televisão. Uma mudança substancial ocorreu em 1962 quando a Bell System colocou em órbita o primeiro satélite comercial de comunicações, o Telstar I. Este passou a constituir uma alternativa especialmente adequada a comunicações de voz internacionais.

Em 1983 foi introduzido o primeiro sistema de transmissão por fibra óptica, que relegou os sistemas de microondas e coaxiais para situações de falha (*backup*) do sistema de fibra. Após a separação da Bell System, em 1984, em diversas companhias (a AT&T para chamadas de longa distância, investigação e desenvolvimento e diversas "*Regional Bell Operating Companies*" - ROCS - para as chamadas regionais), os preços das chamadas de longa distância, nos Estados Unidos, caíram cerca de 40% nos seis anos seguintes. Isto deveu-se à utilização da métodos de transmissão mais económicos (especialmente devido à utilização de fibra óptica), mas também à existência de concorrência para este tipo de chamadas. Na Europa, só no final do século é que acabaram os monopólios dos operadores de telecomunicações.

Em 1988 foi colocado ao serviço o primeiro cabo óptico submarino entre a Europa e os Estados Unidos. Tinha capacidade de transportar cerca de 40.000 chamadas simultâneas, 10 vezes mais do que os anteriores cabos de cobre. Os cabos ópticos utilizados hoje em dia têm uma capacidade de cerca de 1.000.000 de chamadas

simultâneas. Sem dúvida que em 250 anos as telecomunicações evoluíram muito desde o telégrafo original.

2.2.2 Comutação e sinalização nas redes de operadores telefónicos

Até 1926, para estabelecer qualquer chamada, o cliente falava com um operador, que depois de anotar toda a informação necessária, desligava, indicando ao cliente que lhe voltaria a ligar quando a ligação estivesse concluída. Este processo envolvia directamente diversos operadores, tendo cada um que analisar qual o percurso que a chamada deveria efectuar. Este processo demorava cerca de 7 minutos para uma chamada normal e as linhas ficavam dedicadas a essa chamada até esta terminar. Em 1926 este processo foi simplificado, sendo o primeiro operador com quem o cliente falava a estabelecer todo o circuito. O tempo de estabelecimento de chamada caiu para cerca de 2 minutos. Em 1929 a rede telefónica dos Estados Unidos foi reestruturada de acordo com um sistema hierárquico, composto por 8 centros regionais, 140 centros primários e mais de 2000 escritórios onde os operadores comutavam as chamadas para a próxima estação. Esta estrutura foi mantida, com algumas alterações, até aos anos oitenta. Em 1943 foi introduzido o primeiro comutador que permitia aos operadores instruírem o dispositivo através de sequências numéricas de códigos de área. Com o recurso a este sistema, o tempo médio de estabelecimento de chamada caiu para cerca de 10-20 segundos! As redes mais primitivas utilizavam esquemas de sinalização próprios para o estabelecimento e terminação de chamadas, ponto-a-ponto. Devido às limitações dessa sinalização (basicamente permitia apenas abrir e fechar lacetes) e das tecnologias de transmissão, o tipo de serviços oferecido nessas redes era bastante limitado.

Com a evolução das tecnologias de transmissão, as infraestruturas de transmissão analógicas foram substituídas por sistemas de transmissão digitais. Pela primeira vez, em 1970, foi dada aos utilizadores a capacidade de eles próprios fazerem a marcação nas chamadas internacionais (disponível entre Manhattan e Londres). Foi em 1976 que chegou a tecnologia de comutação digital, e no ano seguinte, surgiu a primeira rede de sinalização de chamada separada. Pela primeira vez, a chamada e a sinalização utilizavam caminhos distintos, o que permitiu reduzir os tempos de estabelecimento de chamada para cerca de 1-2 segundos. Através da utilização de uma rede de sinalização separada os comutadores digitais estão sempre em contacto, permitindo a avaliação da disponibilidade de circuitos em tempo real. Em meados dos anos 80, o encaminhamento hierárquico foi substituído por

encaminhamento dinâmico, decidido em tempo real pelos comutadores envolvidos na ligação. Esta alteração resultou numa melhor qualidade de serviço oferecida e num menor custo para os fornecedores. Inicialmente estes sistemas de transmissão eram baseados em tecnologia PDH (*Plesiochronous Digital Hierarchy*). No entanto, devido aos desenvolvimentos nas tecnologias de comutação de alta velocidade e de transmissão óptica, os sistemas PDH foram sendo substituídos por sistemas SDH (*Synchronous Digital Hierarchy*). Desta forma, as redes de transmissão foram altamente simplificadas e foi possível atingir a interoperabilidade de equipamento de grandes capacidades de largura de banda. A sinalização evoluiu também desde o CAS (*Channel Associated Signalling*) para o CCS (*Common Channel Signalling*) chegando ao SS7 (*Signalling System number 7*). Em relação ao CAS, o CCS trouxe diversas vantagens, tais como a facilidade de implementação, o controlo centralizado e a redução do custo dos equipamentos. O SS7 trouxe uma maior fiabilidade e uma maior rapidez na execução das diversas funções, o que provou ser um grande avanço evolutivo. O SS7 permitiu ainda que os utilizadores pudessem especificar à rede a QoS de que necessitavam (em termos de largura de banda). As chamadas digitais passaram a poder ser efectuadas e terminadas de uma forma dinâmica e passou a ser possível o fornecimento de serviços mais evoluídos para além do tradicional serviço de voz.

No entanto, a tecnologia de comutação no *core* das redes continuou a ser baseada na comutação de circuitos, o que limitava o espectro de serviços que poderiam ser oferecidos de uma forma eficiente. Isto porque, numa ligação baseada em circuitos, a largura de banda da ligação está dedicada àquela ligação durante todo o tempo da ligação. Os fornecedores de serviço incorriam em custos elevadíssimos para fornecer serviços "*premium"*, tais como vídeo e acesso remoto de alta velocidade para computação distribuída. Nessa altura, uma opção para o fornecimento deste tipo de serviços era a instalação de uma rede separada (por exemplo, *Frame Relay*) para o transporte desse tipo de tráfego. No entanto, do ponto de vista dos fornecedores, esta solução não era de todo eficiente pois significava manter duas redes quase completamente distintas. Os custos associados a tal esquema são substanciais pois obrigam a duplicar instalações e infraestruturas de transmissão e de gestão. Estes problemas incentivaram a busca de uma tecnologia multiserviço que pudesse, de uma forma eficiente, transportar e cobrir todos os tipos de tráfego, com um ritmo de transmissão elevado e utilizando uma infraestrutura única. Surgiu desta forma o ATM

como uma das tecnologias multiserviço chave. Todos os grandes operadores e muitas outras organizações instalaram ATM nas suas redes públicas e privadas. O ATM permite o transporte de virtualmente quase todo o tipo de informação, requerendo qualquer quantidade de recursos, permitindo assim um espectro de serviços muito alargado e a custos muito competitivos. Sob este ponto de vista, o ATM constituiu um marco importante na convergência dos mundos da voz e dados, que até há poucos anos falavam linguagens complemente distintas.

2.2.3 Desde a transmissão simples de dados até à *Internet*

Desde o início do fornecimento do serviço telefónico que os circuitos foram utilizados para suportar serviços de telégrafo privados. Durante a década de 20, os mesmos circuitos passaram a ser utilizados também para a transmissão de programas de rádio e mesmo de fotografias (através de uma forma rudimentar de fax) para redacções de jornais. Nos anos 30, máquinas de escrever eléctricas permitiram a transmissão digital de texto. Nos anos 40, os circuitos telefónicos passaram também a ser utilizados para o transporte de televisão a longa distância. Uma década mais tarde, foi apresentado o primeiro *modem* comercial que permitiu que os recém inventados computadores pudessem comunicar.

Em 1969, nos Estados Unidos é criada a ARPANet, por decisão do Departamento da Defesa, antecessora da *Internet*, ligando na altura computadores da Universidade de Stanford e da UCLA. A *Internet* cresceu de uma forma continuada, mas apenas após a "invenção" da *World Wide Web* (WWW) [WWW] em 1992, pela mão do físico Tim Berners-Lee do CERN [CERN], é que atingiu a dimensão e globalidade que lhe conhecemos hoje em dia. O ano 2000 constituiu um marco importante pois nesse ano, pela primeira vez, o volume das comunicações de dados nos Estados Unidos ultrapassou o das de comunicações de voz. A afirmação da *Internet* como uma rede global veio trazer novas visões de serviços, correspondendo a novas visões da sociedade. Assim, assistimos actualmente ao "fenómeno *Internet*", suportado pelo protocolo IP (*Internet Protocol*) [IP81], sendo a sua face mais visível e a principal origem desse sucesso, o serviço WWW. A WWW veio facilitar e promover o acesso indiscriminado à informação, agora globalmente tornada pública. Como consequência deste fenómeno, assistimos a um crescimento exponencial do número de dispositivos com capacidade de acesso a esta rede, com especial relevo para os que utilizam a camada de rádio como interface de nível físico.

Neste último caso, a noção de mobilidade surge de uma forma inerente, abrindo um grande leque de opções (e levantando muitos problemas) para o fornecimento de serviços.

2.2.4 A integração de redes e serviços: o utilizador no centro

Podemos considerar que desde muito cedo houve integração de diversos tipos de serviços sobre uma mesma infraestrutura. Observámos que o serviço telefónico foi inicialmente transmitido sobre linhas até então utilizadas para a transmissão de telégrafo. Poucos anos mais tarde observámos que a situação se inverteu. Quando o serviço de voz passou a ser dominante, o serviço de telégrafo ainda continuou a ser utilizado durante muitos anos, mas agora sobre a infraestrutura de telefone. Outros exemplos foram a televisão e o rádio. Inicialmente foi utilizada a rede telefónica para a sua distribuição. Mais tarde, a tendência foi para que cada serviço tivesse a sua infraestrutura, optimizada para o serviço fornecido. Mas cedo os operadores perceberam que os benefícios que advêm de ter essas infraestruturas dedicadas e optimizadas, depressa se diluem quando ponderados os encargos de operação e manutenção de diversas infraestruturas em paralelo. É então com naturalidade que se verifica um certo retorno às origens, com uma nova vaga, desta vez global, de integração de todo o tipo de serviços, mas agora, sobre qualquer tipo de infraestrutura. Não se pretende acabar com as infraestruturas existentes em detrimento de outras, mas sim rentabilizá-las, utilizando cada uma delas para dar um suporte efectivo a quase todos os tipos de serviços. Por exemplo, hoje em dia, são cada vez mais os operadores de televisão por cabo a fornecerem os serviços de dados e telefónico sobre a sua infraestrutura de distribuição.

Há no entanto um factor diferenciador nesta nova abordagem de integração. Não se trata apenas de colocar diversos serviços sobre uma mesma infraestrutura utilizando, por exemplo no caso de tecnologias rádio, frequências distintas para cada um dos serviços. A abordagem actual é para uma verdadeira uniformização do transporte e suporte dos serviços. A última década foi uma década de crescimento explosivo da *Internet* por um lado, e da telefonia celular, nomeadamente o GSM (*Global System for Mobile communications*), por outro. Estes serão os pilares da convergência de serviços e plataformas, aproveitando as sinergias de desenvolvimento de ambos os modos de comunicação. É nesse sentido que todos os desenvolvimentos estão a ser orientados. Tendo como referência a pilha protocolar OSI (*Open Systems Interconnection*), pretende-se criar ao nível da rede uma camada comum a todo o tipo de infraestruturas e a todo o tipo de

serviços. É esta a evolução característica das redes celulares. De um modo simplista, podemos dizer que a primeira geração (1G) se caracterizou pelo fornecimento exclusivo do serviço de voz. A segunda geração (2G) coincidiu com o aparecimento do WAP (*Wireless Access Protocol*) e com a possibilidade da transmissão de dados de e através dos telemóveis e que com a introdução do GPRS (*Generic Packet Radio Service*) [GPRS00][Regi02] se estendeu à 2,5G. A terceira geração (3G) é basicamente definida pela possibilidade dos terminais poderem utilizar o protocolo IP para o estabelecimento de chamadas de voz, ou seja, é a geração da integração de dados e voz ao nível dos terminais de acesso. Quanto à quarta geração (4G) passa por uma unificação de conceitos e de tecnologias de transmissão. O maior desafio de hoje em dia é a junção dos mundos dos serviços fixos e móveis, numa plataforma comum.

Hoje em dia existe uma panóplia de tecnologias disponíveis para suporte à criação e fornecimento de serviços. De modo a atingir esta tão procurada universalidade do serviço haverá que, previamente, uniformizar as características das diferentes tecnologias, ou seja, criar uma camada capaz de abstrair a tecnologia do tipo de serviço. É necessária uma tecnologia, apoiada em vários sistemas e protocolos, capaz de ser o suporte a este objectivo. Devido aos custos reduzidos de acesso à *Internet* e à facilidade e globalidade do acesso, a tecnologia IP tem actualmente uma grande abrangência de serviços, cobrindo serviços anteriormente apenas possíveis de fornecer com tecnologias baseadas em comutação de circuitos. A voz e o vídeo são os exemplos mais visíveis, apetecíveis e populares dos serviços agora suportados em IP. Existe uma grande indústria de conteúdos multimédia a tirar partido da *Internet* tal como ela existe hoje, e com apetência e vontade para explorar tecnologias ainda mais interactivas. O desafio que o IP enfrenta é enorme. Pretende-se não só uma uniformização ao nível das redes fixas, mas também uma uniformização com as tecnologias sem fios, nomeadamente com as redes celulares. O GPRS e o UMTS (*Universal Mobile Terminal System*) [UMTS00][Kaar01] constituíram os pontos de partida desta visão. No entanto, apesar de todos os desenvolvimentos já atingidos, há ainda muitos obstáculos de caracter técnico, filosófico e comportamental que têm de ser transpostos antes do cidadão comum poder utilizar serviços multimédia, em qualquer lado, a qualquer hora, e do modo mais conveniente e sem restrições.

Nos últimos anos houve também uma mudança de filosofia no mundo das telecomunicações. Até há algum tempo, os "serviços" estavam no centro das atenções e

comandavam o funcionamento e a lógica de operação. Na última década, a situação modificou-se, sendo que o utilizador (cliente) ocupa agora o centro do paradigma. Os serviços devem adaptar-se à vontade do utilizador, ao contrário da visão tradicional na qual o utilizador se adaptava aos serviços existentes e suas características. A visão actual centrada no utilizador e nas suas necessidades conduz a um ambiente em que os meios de transmissão sem fios desempenhem um papel cada vez mais importante na sociedade e, consequentemente, na investigação. Conferências, jornais, revistas científicas, e até mesmo o *marketing* dão cada vez maior ênfase e destaque às redes e sistemas sem fios (*wireless*). De entre as diversas mudanças que ocorreram no paradigma da comunicação, o principal ingrediente foi a mudança de uma comunicação fixa ponto-a-ponto (sistema a sistema) para uma filosofia de comunicação móvel, pessoa-a-pessoa. Na última década o número de terminais móveis cresceu exponencialmente, chegando, em alguns países a ter taxas de penetração superiores a 90%. Por exemplo, número total de clientes anunciados pelos três operadores móveis nacionais é já equivalente ou superior à totalidade da população portuguesa. Este facto é mais relevante se atentarmos ao facto de as comunicações entre terminais da rede fixa serem ainda economicamente mais vantajosas. Isto revela as tendências e alterações comportamentais actuais. Hoje em dia é já possível comunicar de qualquer lado e a qualquer hora, incluindo locais impraticáveis até há pouco tempo (como por exemplo os aviões – existem algumas companhias aéreas que em alguns voos já facultam acesso sem fios à *Internet*). Os *hotspots* de acesso sem fios surgem um pouco por toda a parte, respondendo às necessidades dos utilizadores de acesso à informação, constituindo mais uma componente de uma rede global que se está a formar, integrando todas as tecnologias disponíveis.

Outra alteração relevante é o tipo de informação que é transmitida. Os conteúdos multimédia e a comunicação audiovisual são componentes cada vez mais procuradas e utilizadas, e que são bastante exigentes nos recursos de que necessitam. Pode considerar-se que o maior obstáculo ao fornecimento deste tipo de serviços, os serviços multimédia, é a escassez de recursos nas actuais redes. Isto não obstante nos últimos anos termos assistido a um aumento significativo das capacidades de transmissão em redes sem fios. Contudo, este crescimento na oferta de recursos foi acompanhado de perto quer pelo aumento dos requisitos das aplicações emergentes, quer pelo aumento do número de utilizadores. Desta forma, o facto de existirem mais recursos disponíveis, não significa que o utilizador os terá

sempre que os desejar utilizar, como é o caso dos recursos relativamente escassos do espectro de rádio. Assim, especialmente nas redes sem fios, os recursos devem ser geridos e controlados de um modo muito eficiente e terão de existir mecanismos que permitam que utilizadores "seleccionados" (i.e., clientes dispostos a pagar por um melhor serviço) recebam um serviço melhor. Esta exigência acarreta a necessidade de integrar o suporte diferenciado do tráfego de diferentes aplicações, de diferentes utilizadores e/ou serviços.

2.3 Aspectos tecnológicos do fornecimento de serviço de telecomunicações

O fornecimento de serviço está revestido, de um modo geral, de múltiplos factores e graus de complexidade, de acordo com o tipo de serviço, público alvo e tecnologia de suporte. Esta secção expõe em traços gerais as várias vertentes da problemática do fornecimento dos novos serviços de telecomunicações. A maioria das considerações apresentadas são válidas para o fornecimento de um qualquer serviço sobre uma qualquer tecnologia, mas o caso do fornecimento de serviço em redes heterogéneas IP será o mais debatido pela sua previsível relevância futura.

Se por um lado é essencial existir uma motivação para os avanços tecnológicos, por outro, essa mesma tecnologia impõem restrições àquilo que são as motivações para o fornecimento de serviço.

Existem as limitações tecnológicas que só por si balizam aquilo que pode (ou não) ser criado e oferecido. Por exemplo, as limitações de largura de banda das ligações sem fios restringem desde logo o tipo de serviço que pode ser fornecido. As limitações (e custo) da tecnologia fazem sentir-se também a outros níveis, tais como as dimensões físicas: um telefone celular não poderá ter dimensões demasiado grandes. Por exemplo, um grande desafio para os terminais UMTS foi colocar todas as funcionalidades requeridas em terminais cujas dimensões e consumo de potência permitissem uma fácil portabilidade. Um outro factor importante é o legislativo, que pode influenciar de forma decisiva a tecnologia, no sentido em que pode impor limitações (como por exemplo, os níveis de potência e de radiação electromagnética de determinado dispositivo). Ainda a acrescentar às limitações do fornecimento de serviço existem os motivos de ordem humana. Um serviço só se torna útil e só se poderá impor se for fácil de utilizar. Um serviço poderá abranger várias faixas etárias e vários grupos profissionais, daí que a sua forma de utilização deverá ser o mais

simples possível para poder ser utilizado por todos. Nem tudo o que a tecnologia disponibiliza é facilmente adaptável à criação de serviços, tal como nem tudo o que é pretendido de um serviço pode ser realizado.

O fornecimento de serviço não está apenas dependente das possibilidades tecnológicas de o fazer. Factores como o custo, a ergonomia e a utilidade do serviço comandam o processo de selecção e criação de serviços. O problema não se coloca como uma mera discussão teórica acerca das inúmeras possibilidades e funcionalidades que a tecnologia permite, mas sim acerca dos pontos de equilíbrio entre o possível, o necessário e o interessante.

Deste ponto de vista, do equilíbrio potencialmente interessante, as redes de nova geração devem estar dotadas de características que permitam endereçar aspectos fundamentais como a QoS (*Quality of Service* – Qualidade de Serviço), com contratos de serviço representados pelos SLAs (*Service Level Agreements*) e SLSs (*Service Level Specifications*); suportar Mobilidade com Autenticação, Autorização, Contabilização e Taxação (AAAC) associada, permitir Gestão eficiente, e serem suficientemente flexíveis para poderem ser adaptadas de acordo com decisões dos departamentos comerciais.

Todos estes aspectos, fulcrais para o projecto das futuras redes de telecomunicações, são apresentados nas próximas secções, de um ponto de vista muito conceptual, focando as preocupações e motivações dos operadores de telecomunicações.

2.3.1 Mobilidade

O século XX foi um século de mudanças profundas. A mobilidade humana cresceu exponencialmente devido ao surgimento de meios de transporte mais rápidos e cómodos. No entanto, a mobilidade acarreta a necessidade humana de manutenção de comunicação. O desenvolvimento tecnológico verificado ao nível das comunicações permitiu que hoje seja possível estar contactável, em tempo real, em quase toda a parte. Este acesso ubíquo presentemente está disponível apenas nas comunicações de voz, e quer agora estender-se a outros tipos de comunicação e informação. Pretende-se atingir a universalidade de acesso ao universo dos conteúdos. Tudo isto, num ambiente em que o utilizador não necessite conhecer os detalhes técnicos, não necessite preocupar-se com a sua localização física e nem sequer preocupar-se em saber qual o operador que o está a servir em determinado instante. Assim, atingir-se-á a universalidade de acesso transparente à informação.

No entanto a mobilidade tem alguns problemas associados. Veja-se por exemplo o caso de um utilizador em *roaming*, que, apesar de estar fora da rede do operador com quem tem contrato, deve manter o mesmo número telefónico. Este processo implica a existência de sinalização entre os sistemas de gestão do operador com quem tem contrato e do operador onde se encontra no momento. Ainda a outro nível, a mobilidade pode também levar à portabilidade do número ou identificador pessoal. Isto já acontece hoje em Portugal entre os diversos operadores telefónicos, em que é possível o cliente mudar de fornecedor e manter o mesmo número. Quando transpomos estas características de funcionamento das redes telefónicas para as redes IP, colocam-se desafios tão grandes ou maiores do que os que existem nestas redes, que utilizam sinalização dedicada. Ou seja, estas funcionalidades terão de ser reproduzidas ao nível do IP, à custa da introdução de novos elementos e de uma sinalização mais complexa.

Grupos de normalização, como o 3GPP [3GPP], têm vindo a desenvolver esforços para dotar as redes celulares com os serviços actualmente disponíveis na *Internet*. Após a possibilidade de utilizar o telemóvel com funcionalidades de *modem* e servindo de interface para recepção e envio de dados através de um computador, estes grupos avançaram em direcção a algo mais ambicioso: levar a *Internet* aos terminais móveis. O GPRS trouxe já essa possibilidade, embora ainda de uma forma algo limitada em termos dos serviços que podem ser oferecidos. Com o UMTS, faseado em termos de funcionalidades e capacidades disponibilizadas, pretende-se alargar o espectro de utilizações possíveis de um terminal móvel, bem como o modo como o acesso aos serviços será feito. Com efeito, as *release* 5 e seguintes do UMTS [BR5U04] revêem que todos os serviços, incluindo o serviço de voz, sejam "*all-IP*", atingindo a convergência de serviços e tecnologias.

As possibilidades de mobilidade nas actuais redes celulares são a de um terminal se mover de célula em célula, utilizando sempre a mesma tecnologia de suporte. Ainda a este nível, a mobilidade entre diferentes operadores é suportada na sua vertente de *roaming*, não sendo possível a mudança de fornecedor, em tempo real, durante uma sessão. Um outro exemplo de mobilidade ainda mais restrita é aquela que se consegue ao aceder à *Internet* através de um *Hotspot Wireless LAN* (*Local Area Network*), em que os terminais têm possibilidade de se deslocarem apenas dentro do *Hotspot*, normalmente de cobertura reduzida. Pretende-se no futuro poder ir um passo mais além e permitir que as mudanças se

efectuem entre várias tecnologias distintas, envolvendo vários fornecedores em tempo real, durante o decurso de uma qualquer chamada telefónica ou de dados. Poder-se-á alegar que é desvantajoso para um operador permitir a mobilidade de uma tecnologia em que o utilizador paga mais (por exemplo, UMTS) para uma em que pagará menos (por exemplo, *Ethernet* ou *Wireless* LAN). No entanto, isso não é verdade. Existem vários motivos que justificam o interesse e vantagens do operador. O primeiro, não necessariamente o mais importante, é que a recusa de um operador em fazer isso poderá ser aproveitada comercialmente por outro: este suporte apresenta-se como algo inevitável e quem se adiantar terá vantagens, conseguindo mais clientes. Um segundo motivo prende-se com o facto de também a mobilidade em sentido oposto ser suportada, isto é, também haverá situações em que o cliente passa da tecnologia "mais barata" para a "mais cara" (quando por exemplo sai do raio de cobertura de um *Hotspot Wireless LAN*, tem em curso uma sessão que não quer terminar e utilizará os recursos UMTS que terá à sua disposição no momento). Por último, o facto de o cliente pagar menos pelo seu acesso, utilizando uma tecnologia "mais barata", não quer dizer que o operador ganhe menos. Isto é, as tecnologias que saem menos dispendiosas ao cliente, também são aquelas em que o operador tem de investir menos para ter uma maior capacidade e são precisamente nestas que a margem de ganho é maior (por exemplo, é muito mais barato a um operador colocar 5 x 11 Mbps num acesso *Wireless LAN*, do que 11 Mbps num acesso UMTS – e aquilo que cobrará ao cliente por um e outro acesso não reflectirá potencialmente esta diferença do serviço).

Num futuro próximo pretende-se ainda expandir o conceito de mobilidade a níveis mais elevados, que permitam a transferência de uma sessão de um terminal para outro, sem cortes nem erros de comunicação. Por exemplo, o cliente desloca-se no seu veículo, com o seu filho no banco de trás (após o ter ido buscar à escola) jogando um jogo interactivo com um seu colega (e tendo também uma sessão de vídeo-telefonia com o seu amigo) a partir da consola disponibilizada pelo próprio veículo, e chega a casa. Ao chegar, o filho que ainda leva o jogo a meio, decide continuar o jogo através do seu terminal multimédia interactivo (o mesmo que utiliza na escola para receber os conteúdos disponibilizados pelo professor) e desloca-se para o interior da casa, para o seu quarto, onde por fim pode comutar toda a sessão para a sua estação multimédia fixa. Neste processo, o rapaz terá utilizado os recursos UMTS (através do veículo) enquanto viajava, comutou para o acesso

Wireless LAN que cobre toda a sua casa e área envolvente, e por fim terá comutado para um acesso "*Multi Gigabit Ethernet*" que tem no seu quarto. Tudo isto, sem quebra na sua sessão e ganhando pontos ao seu adversário.

Contudo, estas mudanças, evoluções e novos serviços têm custos elevadíssimos para os fornecedores de serviço e só com uma base de sustentação para os investimentos realizados é que isto pode ser tornado realidade. Desta forma, um factor determinante para o sucesso é a capacidade de os fornecedores conseguirem convencer os clientes a utilizarem e pagarem por estes novos serviços. Como o pagamento é feito de acordo com a utilização, as redes dos fornecedores devem estar preparadas para fazerem adequadamente essa contabilização aos utilizadores com quem têm contrato, mesmo em cenários em que os clientes estejam a mover-se constantemente.

2.3.2 AAAC (*Authentication, Authorisation, Accounting and Charging*)

A prestação de um serviço exige, salvo casos excepcionais, uma contrapartida. Assim, é vital que exista uma relação estreita entre o fornecedor e o cliente. Num ambiente em que este relacionamento se efectua num espaço "virtual" e sem contacto físico, torna-se necessário impor mecanismos que garantam segurança para ambas as partes. A Autenticação do cliente perante o fornecedor é essencial para garantir ao fornecedor que está a prestar o serviço à entidade com quem tem um acordo (podendo assim cobrar à entidade certa) e, por outro lado, garantir ao cliente que mais ninguém irá usufruir dos recursos por que está a pagar no âmbito do seu contrato. Portanto, a autenticação é o primeiro passo fundamental para que se estabeleça essa relação de confiança entre cliente e fornecedor, de modo a um ou mais serviços poderem ser prestados. Após a fase de autenticação, o fornecedor do serviço sabe qual o cliente a quem irá fornecer o serviço e desta forma sabe também quais os serviços que esse cliente em particular poderá usufruir. Esses serviços dependem e fazem parte do contrato que necessariamente existe entre ambas as partes.

Para que um serviço possa ser fornecido é necessária também uma Autorização. Isto é, um determinado fornecedor de serviços poderá ter um portfólio de serviços muito extenso, mas um determinado cliente poderá ter apenas contratado alguns e, nesse caso, só estará autorizado a utilizar esses. Existem casos em que a autorização de determinado(s) serviço(s) é feita logo após a fase de autenticação, mas este processo poderá ser independente. Vejamos como exemplos distintos o que se passa num acesso à *Internet* via

modem e no caso de uma chamada telefónica via rede GSM. No primeiro caso, logo na fase de autenticação do cliente (envio de nome de utilizador e palavra chave), o fornecedor autoriza-o de imediato a aceder ao serviço contratado (acesso à *Internet*). No segundo caso, a fase de autenticação dá-se quando o cliente liga o telemóvel e introduz o seu PIN (*Personal Identifier Number*) e se desencadeia o processo de autenticação com o fornecedor de serviço. Após esta fase o cliente poderá realizar chamadas telefónicas mas será autorizado para a realização de cada uma apenas quando a iniciar. Ou seja, a fase de autorização realiza-se apenas no acto do fornecimento do serviço. Desta forma, é possível ao fornecedor, por exemplo, impedir que utilizadores em falta (que tenham contas em atraso) realizem chamadas, ficando apenas habilitados a receber. Este é um exemplo típico em que a autenticação e a autorização de serviço são processos independentes.

Ainda associado ao fornecimento do serviço está a capacidade do fornecedor o poder cobrar. Desta forma, é essencial que o fornecedor possa contabilizar os gastos de cada um dos seus clientes, em cada um dos serviços utilizados, para lhe poder exigir o respectivo pagamento. A contabilização de serviços diferentes poderá ser feita de modos distintos, e mesmo a contabilização de um determinado serviço poderá ser distinta dependendo do contrato existente entre fornecedor e cliente. Por exemplo, um determinado ISP (*Internet Service Provider*) poderá contabilizar o acesso à *Internet* por tempo de utilização, por volume (número de *bits*), ou mesmo numa combinação de ambos. Poderá também haver serviços que não estejam sujeitos a cobrança, como por exemplo o acesso ao número de emergência, no caso das redes telefónicas.

A tarifação dos serviços é algo que está também subjacente à grande generalidade dos serviços prestados. Esta tarifação depende não só do serviço, mas também do contrato entre o cliente e o fornecedor. Um determinado cliente, pelo facto de ter mais serviços subscritos ou por ter subscrito um serviço mais dispendioso, poderá ter contrapartidas. Estas questões estão todas intimamente relacionadas e são denominadas de questões de AAAC (*Authentication, Authorisation, Accounting and Charging*), que se apresentam como uma evolução do agora vulgarizado conceito de AAA (*Authentication, Authorisation and Accounting*). O cliente, uma vez firmado o contrato, terá um perfil de utilização que estará guardado no sistema de AAAC.

A complexidade acrescida que advém do fornecimento de uma maior multiplicidade de serviços nas redes IP com mobilidade, força a que os sistemas de AAA tenham funcionalidades e capacidades acrescidas, como veremos na secção 3.3.

2.3.3 SLAs e SLSs

Sempre que exista um ambiente de fornecimento de serviço, é forçoso existirem regras e meios para garantir do seu cumprimento. Em todas as situações da vida quotidiana, quando existe o fornecimento de um bem ou serviço, existem subjacentes um conjunto de garantias que especificam as características desse bem ou serviço, ou seja, existe um contrato. Esse contrato pode ser implícito ou explícito.

Podem existir situações em que haja violação desse contrato por diversos motivos. O contrato tem normalmente associado um nível de serviço e prevê contrapartidas para o caso de incumprimento (reposições ou indemnizações). Por exemplo, quando uma empresa contrata um serviço de 2 Mbps, caso o operador não consiga cumprir a disponibilidade desse serviço por um período de X horas, terá direito a um eventual reembolso e compensação pelo tempo de indisponibilidade. Isto não é válido apenas nas telecomunicações, mas em tudo o que é passível de ser transaccionado. No mundo das telecomunicações, ao acordo e especificação de características de serviço entre as partes envolvidas, dá-se o nome de SLA (*Service Level Agreement*). É de notar que um SLA não existe apenas entre cliente e fornecedor; podem existir SLAs entre operadores (fornecedores), em que cada um pode ser visto simultaneamente como cliente e fornecedor.

Em termos de redes IP, existe já algum trabalho efectuado na definição de SLAs e SLSs (*Service Level Specifications*). De acordo com o QoS Forum [QoS Forum], um SLA pode ser definido da seguinte forma:

"Um SLA é um contrato de serviço entre um fornecedor de serviço e os seus clientes que define as responsabilidades do fornecedor em termos de níveis da rede e disponibilidade, método de medição, consequências se os níveis de serviço não forem alcançados ou se os níveis de tráfego definidos forem excedidos pelo cliente e todos os custos envolvidos".

Uma especificação de nível de serviço (SLS) é um subconjunto do SLA, orientado para os aspectos da qualidade de serviço. O SLS descreve as características operacionais do SLA em termos de níveis do serviço. O QoS Forum dá a seguinte definição de SLS:

"O SLS pode consistir no *throughput* esperado, probabilidade de perda, atraso, perturbações nos pontos de entrada e saída nos quais o serviço é fornecido, adesão a padrões de tráfego para que o serviço desejado seja fornecido, tratamento do tráfego submetido em excesso relativamente ao padrão especificado, e marcação e formatação dos serviços fornecidos". O SLS define portanto, de uma forma genérica, os parâmetros de rede que o serviço terá que cumprir. No entanto, os SLSs devem ser independentes da tecnologia de nível 2 que está a ser usada para o fornecimento do serviço, até porque um serviço poderá depender e atravessar áreas com diferentes tecnologias de nível 2.

Existem várias maneiras de mapear os parâmetros de SLS em parâmetros de rede pelo que a alteração da tecnologia de rede subjacente deve ter um impacto mínimo no SLS: este deve ser portável. O SLS pode, teoricamente, ser monitorizado e modificado por meios electrónicos. Em contrapartida, o SLA pode possuir outras condições que não sejam passíveis de ser monitorizadas nem modificadas por meios electrónicos. A especificação de SLA deve ser independente de fabricantes, deve garantir a interoperabilidade e ser escalável. Um exemplo é o acordo no qual, em caso de conflito, o cliente e o fornecedor de serviços aceitam o julgamento num tribunal de comarca específico.

A base de dados com os valores limites dos parâmetros constantes do SLS é derivada do SLA. Nessa base de dados são definidos quais os parâmetros do SLS que podem ser modificados e quais os valores máximos ou mínimos que esses parâmetros podem ter. Exemplos de possíveis parâmetros SLS são:

- Largura de banda máxima para envio de informação;
- Largura de banda máxima para recepção de informação;
- Atraso máximo aceitável;
- Variação do atraso (*jitter*) máximo aceitável;
- Percentagem mínima de disponibilidade;
- Percentagem máxima de perda média de pacotes;
- Parâmetros de segurança.

O fornecedor pode planear os seus recursos com base nesses limites. O cliente é forçado a aderir a esses limites de forma a evitar situações indesejáveis nos nós da rede.

Existem algumas áreas diferentes de restrição de tráfego:

- Restrições na quantidade total, média e instantânea de largura de banda que um cliente pode utilizar ao mesmo tempo, incluindo tráfego de entrada e saída. Isto permite a existência e fornecimento de serviços assimétricos;
- Restrições nas classes de serviço (CoS - *Classes of Service*) que podem ser seleccionadas. Podem ser construídas classes de serviço especiais dependendo do tipo de cliente;
- Restrições na utilização de filtros de tráfego definidos pelo utilizador. Os filtros de tráfego permitem ao cliente especificar regras que determinam como a rede deve tratar tipos diferentes de tráfego.

No processo de gestão de SLA, o administrador define os SLSs para a rede (ou servidor) e a ferramenta de gestão traduz os SLSs em parâmetros de controlo de débito para os dispositivos específicos da rede. Os parâmetros específicos de cada elemento de rede são distribuídos pelos dispositivos correspondentes. Os agentes nos dispositivos recebem os parâmetros da ferramenta de gestão e processam-nos de modo a obterem configurações exactas para aquele tipo de dispositivo.

2.3.4 Qualidade de Serviço

O problema dos SLAs está assim associado ao facto do fornecimento de um serviço impor o cumprimento de determinadas regras. O serviço prestado poderá ter uma melhor ou pior qualidade e isto pode ser dependente de factores distintos, tal como, por exemplo, os encargos que o cliente está disposto a ter com o serviço (quando vamos a um restaurante "mais caro", normalmente a comida e o serviço têm "melhor qualidade"). De uma forma muito generalizada podemos dizer que, a cada serviço está subjacente uma determinada Qualidade de Serviço (QoS). Em cenários de provisão de serviço, é importante o fornecedor poder dar garantias estritas de QoS, mas é também importante do ponto de vista do operador conseguir diferenciar a QoS que atribui a diferentes clientes e diferentes serviços (que pelas suas características intrínsecas o necessitem). Numa lógica puramente comercial pode dizer-se que, em telecomunicações, quem paga mais, deverá ter, à partida, melhor qualidade.

A necessidade de controlar os recursos que cada cliente, utilizador, ou aplicação utiliza, deve-se sobretudo ao facto de os recursos serem escassos. Numa situação limite em que a quantidade de recursos disponíveis seja ilimitada, quase deixa de fazer sentido o controlo de QoS tirando, claro, a vertente comercial (numa situação extrema, poderá haver

interesse de um operador em degradar intencionalmente um serviço mais barato para que os utilizadores sintam necessidade de optar por um mais dispendioso). Assim, e de um modo geral, pode dizer-se que o controlo de recursos (e de QoS) é tão mais crítico quando mais limitados são os recursos disponíveis. Neste aspecto em particular, as redes sem fios são aquelas onde este controlo deverá ser mais necessário, dada a sua (relativamente) baixa largura de banda.

A definição de qualidade de serviço é fortemente dependente de como é vista e por quem. Isto é particularmente visível quando pessoas do mundo tradicional dos circuitos comutados e pessoas do mundo da comutação de pacotes se encontram para discutir e acordar parâmetros de QoS.

No mundo das telecomunicações, de um ponto de vista técnico a QoS é dependente do nível a que seja observada, isto é, existem vários níveis a partir de onde se pode observar e "medir" QoS e que são quantificáveis de formas distintas. Para exemplificar, atente-se à Figura 1 onde se exemplifica os diferentes significados de QoS, dependentes das entidades e pontos onde esta é observada. Por exemplo, uma falha momentânea de uma ligação na rede, é interpretada ao nível do desempenho de rede como uma sequência de pacotes perdidos, enquanto que ao nível da aplicação, pode representar a perda de uma trama de uma transmissão de vídeo e o utilizador, que está a ver o filme, interpreta como uma "distorção na imagem".

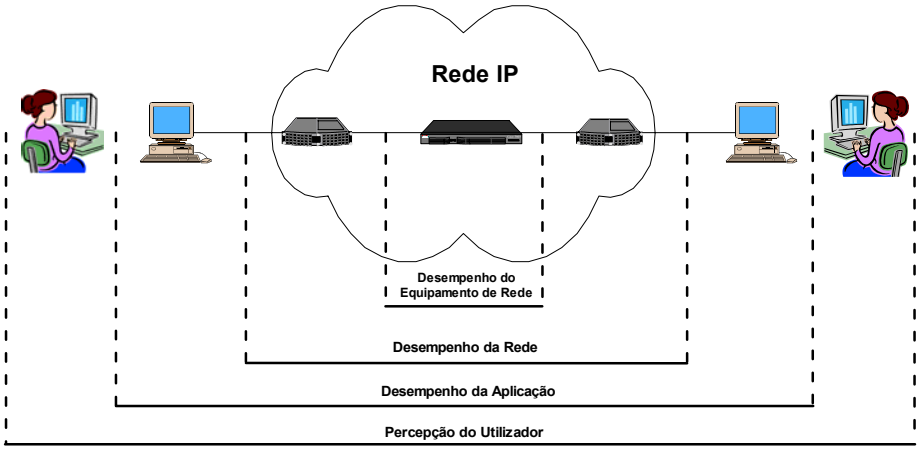


Figura 1: Qualidade de Serviço dependente do ponto de "observação"

A Figura 2 mostra como a noção de QoS apresentada na Figura 1 é mapeada para um modelo de camadas.

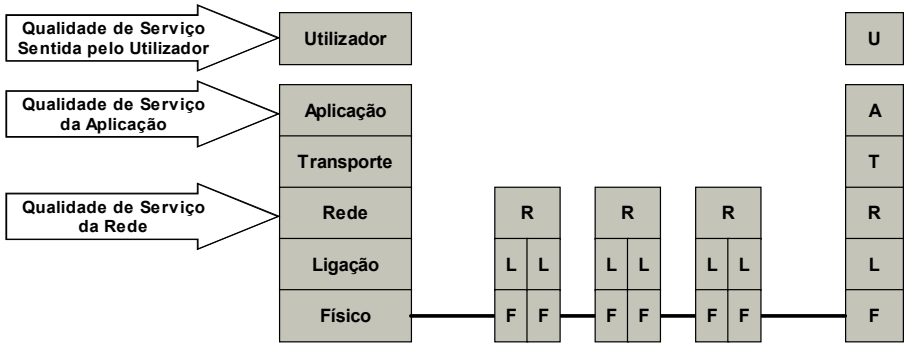


Figura 2: Várias camadas de Qualidade de Serviço

Há já algum tempo que as potencialidades das redes de transporte de pacotes são exploradas para o fornecimento de serviços diversificados e exigentes, mas estes exigem da rede algumas garantias de cumprimento de parâmetros de modo a serem praticáveis. Têm sido propostas várias alternativas para garantir essa diferenciação de QoS, alternativas essas que vão desde a definição de uma nova tecnologia de transporte à proposta de melhorias das já existentes. Por exemplo, o ATM (*Asynchronous Transfer Mode*) surgiu com o objectivo de vir a ser uma tecnologia integradora e com suporte de QoS. O ATM foi definido de tal forma que seria adequado quer à transmissão de informação com requisitos de tempo real (e.g. voz e vídeo), quer à transmissão de dados.

O IP, que agora se assume claramente como a tecnologia integradora de redes e serviços, não foi planeado para dar suporte de QoS. Neste sentido tem sido realizado muito trabalho de investigação de modo a dotar o IP com mecanismos que permitam gerir e controlar a QoS, pois é absolutamente necessário que a QoS atribuída a cada serviço seja a que satisfaz adequadamente as necessidades desse serviço. Por exemplo, os parâmetros de rede associados a um serviço de voz são bastante distintos dos parâmetros que deverão ser associados a um serviço de transferência de dados. O operador deverá ter a capacidade de fazer essa distinção e tratar cada serviço de acordo com as suas características. O IETF (*Internet Engineering Task Force*) [IETF] tem sido o organismo onde esta problemática tem tido maior atenção, mas existem outros (ETSI - *European Telecommunications Standards Institute*, por exemplo [ETSI]) que contribuíram e continuam a contribuir com novas soluções e alternativas. Hoje em dia existem dois modelos de controlo de QoS em redes IP, definidos no âmbito do IETF: os serviços integrados (IntServ – *Integrated Services* – Integração de Serviços) [Brad94] e os serviços diferenciados (DiffServ – *Differentiated Services* – Diferenciação de Serviços) [Carl98]. Ambos os modelos têm

vantagens e desvantagens; o IntServ oferece garantias estritas de qualidade mas é pouco escalável; e o DiffServ apesar de não poder oferecer garantias estritas, é totalmente escalável. Baseados neste modelos existem agora diversos modelos mistos e derivados que tentam usar o melhor dos dois modelos, em abordagens integradas de controlo e gestão.

2.3.5 Gestão

A gestão dos recursos, plataformas, serviços e utilizadores nem sempre foram vistas de um modo integrado. Hoje em dia procura-se concentrar num sistema de gestão de rede (que poderá ser implementado num sistema distribuído) um conjunto alargado de informação da rede, seus elementos, utilizadores e seus serviços. Esta informação pode ser obtida por diferentes vias, recorrendo quer a módulos internos de cada um dos elementos de rede, quer a plataformas de monitorização passiva e activa (em tempo real ou diferidas no tempo). Desta forma, é possível manter informação e controlar o estado da rede, dos seus elementos, dos serviços a ser prestados e de eventuais violações de contratos (SLAs e SLSs). Um sistema de gestão que tenha acesso a esta informação pode controlar novos acessos, determinar as políticas de crescimento da rede, identificar lacunas e até potenciar o aparecimento de novos serviços, tendencialmente mais atractivos. Estas plataformas de gestão de serviços podem ainda tornar mais simples a utilização dos serviços, de modo a esconder do utilizador o máximo da complexidade dos mesmos. Por exemplo, não faz sentido o utilizador saber que para fazer um telefonema de voz sobre IP, o seu terminal terá que sinalizar os recursos e o *codec* necessários e fazer marcação de pacotes de voz numa prioridade alta, entre outras coisas. Do ponto de vista de gestão, também não interessa ao operador/fornecedor ter e disponibilizar aos clientes serviços de complexidade elevada. Gerir serviços complexos envolve ter plataformas de controlo mais "pesadas" e dispendiosas. Assim, a complexidade dos serviços deverá ser reduzida a um mínimo que satisfaça as expectativas dos clientes. Caso seja necessário disponibilizar serviços mais complexos, com maior granularidade, e em que o cliente tenha um maior controlo, esses serviços devem reflectir isso no custo ao cliente. Assim, garante-se que apenas quem realmente necessita de serviços complexos os irá querer subscrever (em princípio, poucos clientes), e que o serviço também é economicamente vantajoso para o fornecedor.

A plataforma de gestão deve ser a responsável pelo conhecimento dos SLAs e SLSs. Tradicionalmente deverá implementar as funcionalidades do modelo FCAPS (*Fault,*

Configuration, Auditing, Performance and Security) [M.3400], que contempla todas as funcionalidades exigidas a um sistema de gestão.

Existem referências para a implementação de SLAs em redes *Frame Relay* [Fram98] bem como em redes IP. Em particular, a arquitectura do grupo de trabalho do DiffServ chegou a conclusões similares em relação à noção de um "acordo", que implica considerações que vão desde o preço, contrato ou outra natureza de negócio, a considerações meramente técnicas.

2.3.6 Promoção comercial

O sucesso de um serviço não depende apenas do facto de serem cumpridos ou não os contratos, e de este ser ou não fácil de utilizar. O *marketing*, a arte de criar uma necessidade onde ela pode não existir, é por vezes a única razão para o sucesso ou insucesso de um produto ou serviço. Este aspecto é algo que os operadores de telecomunicações estão particularmente cientes, e que podem condicionar as características mais técnicas.

Os serviços, por mais ou menos complexos que sejam na sua essência e em termos de exigência das plataformas de gestão e controlo de rede, devem ser mostrados ao utilizador num embrulho atractivo fazendo com que o utilizador os queira utilizar. Para se ter um serviço de sucesso do ponto de vista de um operador, é necessário que os utilizadores se viciem nesse serviço. Existem dois caminhos distintos que conduzem a este objectivo: ou o serviço surge de uma necessidade real dos utilizadores, e estes estão dispostos "a tudo" (isto é, pagar e aprender a usar) para o obter, ou, caso não surja de nenhuma necessidade premente dos utilizadores, é necessário deixar que os utilizadores se habituem a ele até não poderem passar sem ele. Em ambos os casos, é frequente os fornecedores optarem por campanhas de *marketing* onde a promoção do serviço se faz deixando os utilizadores experimentá-lo de uma forma (quase) gratuita, até não conseguirem passar sem ele. Nesta altura o fornecedor poderá começar a ter retorno pelo investimento feito. Isto tem frequentemente impacto nos aspectos técnicos, tendo os sistemas de gestão de ser configurados de acordo com "regras" impostas pelos aspectos comerciais. Estas regras devem considerar o desempenho, SLAs e controlo de acesso.

Também os planos tarifários podem fazer um serviço ou mesmo um fornecedor serem bem ou mal sucedidos. Ainda assim, mesmo em casos onde os serviços são claramente mais dispendiosos que os seus concorrentes, o *marketing* e a imagem que este

gera do operador e do serviço são determinantes. Os utilizadores estão muitas vezes dispostos a pagar mais, se estiverem convencidos que estão a usufruir de um serviço melhor. Acontece muitas vezes que, havendo dois serviços muito parecidos, os utilizadores tendem a optar pelo "mais conhecido", aquele cujo *marketing* foi mais eficaz. Este aspecto é naturalmente essencial no ambiente actual de telecomunicações e a tecnologia tem de ter a capacidade de reacção e adaptação exigida pelos aspectos comerciais.

2.4 Desafios para o fornecimento de novos serviços

Nas secções anteriores vimos que o fornecimento de um serviço de telecomunicações envolve vários aspectos e pode estar revestido da um grau de complexidade elevado. Isto é verdade mesmo nos casos em que o serviço é fornecido apenas numa infraestrutura monolítica, com poucos graus de liberdade.

Os cenários de telecomunicações de próxima geração apontam para um aumento da complexidade dos serviços. Esta complexidade adicional advém do suporte de infraestruturas flexíveis, por exemplo, mobilidade entre tecnologias de acesso distintas, QoS garantida durante a mobilidade, possibilidade de negociação e renegociação de recursos de rede, portabilidade de serviços, entre outros.

Nas redes de próxima geração o inter-relacionamento entre os diversos elementos envolvidos no fornecimento de um serviço necessita ser mais estreito. As funções de QoS e de mobilidade devem interagir para garantir o cumprimento dos SLSs e SLAs durante o movimento. O sistema de AAAC deve interagir com os anteriores para que seja feita uma correcta contabilização dos serviços de determinado utilizador e para garantir que durante o movimento, o cliente continua correctamente autenticado. A plataforma de gestão deve ter informação de todos os outros sistemas para poder controlar e gerir a rede de uma forma que garanta o máximo desempenho. Os departamentos comerciais devem conhecer as potencialidades das plataformas de serviços para que possam tirar o melhor partido nas campanhas de angariação e/ou reforço de clientes e número de serviços vendidos.

As arquitecturas de rede actuais ainda não estão preparadas para estes desafios que se aproximam. São, na sua maioria, arquitecturas "estanques", que não suportam a interacção com outras (ou têm apenas um suporte básico), especialmente se os géneros de serviços oferecidos são bastante diferentes. É então necessário definir novas arquitecturas capazes de responder afirmativamente aos desafios que a provisão de serviços em cenários de nova

geração colocam. Algum trabalho tem já vindo a ser feito e ao longo desta tese serão apresentados aspectos de uma arquitectura que visa abordar com sucesso a problemática exposta nas secções anteriores. Seguindo as tendências actuais, esta arquitectura é baseada em protocolos associados à *Internet*.

CAPÍTULO 3

SUPORTE DE SERVIÇOS EM REDES HETEROGÉNEAS

3.1 Introdução

A provisão de serviços atractivos em redes heterogéneas requer um conjunto de características das tecnologias de suporte. Algumas destas características podem ser consideradas fundamentais, no sentido em que sem elas o fornecimento do serviço não é de todo possível, enquanto que outras, embora não sendo fundamentais, permitem enriquecer o serviço, e são requeridas pelos departamentos comerciais. Este capítulo apresenta um conjunto de tecnologias e protocolos necessários para a provisão de serviço em redes de próxima geração com mobilidade e qualidade de serviço em ambientes de rede de operadores. Este apanhado de tecnologias representa uma visão pessoal, do ponto de vista do operador de telecomunicações, sobre protocolos especialmente relevantes para esta próxima geração de redes. Assim, iremos abordar o ATM como a primeira tecnologia que visou a integração total de serviços; o IP (em especial o IPv6) como protocolo integrador

de transporte e de mobilidade; o RADIUS (*Remote Access Dial In User Service*) [Rign00] e o DIAMETER [Calh01] como protocolos de AAA (*Authentication, Authorisation and Accounting*); o SIP (*Session Initiation Protocol*) [Rosb02] como protocolo de sessão; e o DMIF (*Delivery Multimedia Integration Framework*) como protocolo de abstracção entre o nível da aplicação e o de rede.

Antes porém de entrarmos nos aspectos de pormenor da tecnologia, importa fazer uma análise crítica à evolução da tecnologia digital nos últimos anos. De um modo radical, podemos dizer que muito pouco de novo aconteceu conceptualmente em termos de tecnologia digital nos últimos 8 anos. O crescimento das capacidades computacionais (velocidade de processadores e capacidades de armazenamento) tornou-se extremamente previsível. A maioria dos bens de consumo tecnológico, tais como computadores e telemóveis, já são comprados numa lógica de que serão trocados passados poucos anos (se chegam a ser anos). Algumas das "inovações" actuais não são mais do que conceitos desenvolvidos há algumas décadas atrás, mas que na altura não tiveram o objecto onde a sua aplicação fosse necessária, ou os sistemas electrónicos não possuíam desempenhos aceitáveis para os tornar viáveis. As técnicas de espalhamento de espectro de rádio são um bom exemplo. Pode reclamar-se que a *Internet* é realmente diferente e inovadora, mas o surpreendente é que pouco mudou nos últimos anos. Foi em termos de nível físico que houve algumas inovações importantes, tais como a introdução dos backbones em fibra óptica (e o WDM – *Wavelength Division Multiplex*), e as técnicas de comutação e encaminhamento de pacotes, realizadas em *hardware* dedicado. No entanto, até estes desenvolvimentos não levaram a mais do que à manutenção do ritmo de crescimento da largura de banda disponível que já se verifica há mais de 20 anos.

Quanto aos diferentes elementos das pilhas protocolares, o modo dos terminais comunicarem e a linguagem que utilizam, assistimos a uma incrível resistência à mudança. Quase todas as tentativas de mudança esbarraram em obstáculos à partida desprezados. Não foi por estas inovações serem impraticáveis ou tecnicamente deficientes, mas sim porque as normas existentes se entranharam de tal forma na sociedade que a inércia para a mudança é demasiado elevada. Aquilo de que se fala não é mais do que um fenómeno chamado de "efeitos de rede" que foi inicialmente estudado por economistas. Estes fenómenos são comuns em tecnologias como a eléctrica e de redes que dão um enorme ênfase à compatibilidade. O conceito é de fácil compreensão: quando várias companhias

eléctricas decidiram adoptar a energia eléctrica alternada em vez da energia contínua, a energia alternada passou a ser muito mais atractiva para quem a seguir teve de decidir qual deveria fornecer. Nesse momento deu-se o "trancar da tecnologia" ("*lock-in*"). Podemos aplicar o mesmo exemplo aos sistemas operativos dos computadores – veja-se como a Microsoft [Micr] conseguiu o domínio mundial (outros sistemas procuram agora o seu espaço, tais como o *Linux* [*Linux*], mas tendo uma preocupação constante de compatibilidade com as principais aplicações do mundo "Microsoft"). No mundo das redes, olhando para o IP, que define como transportar informação entre dois extremos, vemos que durante os anos 80 havia muitas normas semelhantes. Chegou mesmo a pensar-se que o IP seria substituído por um protocolo OSI (*Open Systems Interconnection*) (o CLNP – *ConnectionLess Network Protocol*), com o apoio do governo dos Estados Unidos. Mas, enquanto decorria a normalização do CLNP, o IP angariava utilizadores. Em 1992, com o aparecimento da *WWW* tinha-se chegado ao "ponto de não retorno" – o IP atingiu o "*lock-in*". Nos últimos anos, em termos de rede, não assistimos a mais do que ao crescimento do número de utilizadores e utilizações do IP. O IP, tal como um furacão tecnológico, atingiu WANs (*Wide Area Networks*) e LANs (*Local Area Networks*), deixando para trás um conjunto de equipamentos e protocolos agora obsoletos, que poucos ainda se lembrarão daqui a alguns anos.

Neste momento o IP prepara-se para assumir preponderância nos sistemas telefónicos mundiais, consubstanciado no sucesso dos novos sistemas celulares. Este será um momento crítico em que uma tecnologia que atingiu o "*lock-in*" será substituída por outra (mas até neste caso podemos considerar que o "*lock-in*" foi atingido pelo serviço de voz, e não pela tecnologia de suporte – os circuitos comutados). Enquanto que há poucos anos se dizia que o IP nunca poderia ser utilizado para o transporte de tráfego de tempo real pois nunca poderia atingir a mesma qualidade que um circuito comutado, observa-se agora o IP a avançar definitivamente nestas áreas. Facilmente se compreende que com um crescimento do tráfego da *Internet* de 400 a 700 por cento ao ano, contra uns escassos 7 por cento do tráfego de voz, os operadores prefiram unificar o tráfego sobre uma mesma infra-estrutura de dados, em vez de manter duas infra-estruturas separadas, ou tentar colocar tudo sobre a infra-estrutura de circuitos comutados (tal como acontecia até há poucos anos).

Os efeitos de rede, ao tornarem o progresso previsível, trazem um benefício claro: as firmas de alta tecnologia podem planear os seus negócios. Agora que é claro que o IP ganhou a guerra do transporte, as empresas de *software* podem investir o seu esforço em aplicações sem a preocupação de saber se quando terminarem o desenvolvimento ainda terão a tecnologia de suporte que inicialmente planearam.

O IP trouxe ainda outros desenvolvimentos importantes noutras áreas não tão visíveis, como o nível físico do *core* da rede. Há uns anos atrás, quando os operadores começaram a instalar as suas redes de *backbone* em fibra, fizeram-no baseados no crescimento de cerca de 7 por cento do tráfego telefónico, e em (à data) 21 por cento de crescimento do tráfego da *Internet*. Quando a *Internet* atingiu ritmos de crescimento entre os 400 e os 700 por cento, rapidamente as poucas fibras instaladas começaram a atingir a sua capacidade máxima de transmissão. Foi esta a motivação para as técnicas WDM (para partilha da fibra por vários feixes de luz), e mais tarde para os amplificadores ópticos (de ébrio) que tornaram o uso de WDM viável. Desta forma, do mesmo modo que a lei de Moore prevê que a capacidade dos circuitos electrónicos duplique em cada 18 meses, também os operadores podem confiar nesta rede de transporte "transparente". A isto chama-se progresso, ou melhor, chama-se "progresso automatizado", assim definido pelo economista Austríaco, Joseph Schumpeter (embora aplicado a áreas bem distintas, como é o caso do Capitalismo) [Schu34].

Em contrapartida, no nível protocolar seguinte, o nível lógico, a imposição do IP levou por arrasto ao esmagamento (quase completo) de qualquer tentativa de mudança. Nos anos 80 a *Ethernet* tornou-se a norma para as LANs. No caso das WANs, a norma foi o *Frame Relay*. Mais de uma década após, e numa escala de várias ordens de magnitude acima, a *Ethernet* reina na LAN, enquanto que se pode considerar que a WAN é partilhada pelo *Frame Relay* e pelo ATM. No entanto, não era suposto isto ter acontecido. No fim do século esperou-se pacientemente o "ano do ATM", uma norma que viria a unificar as LANs e as WANs, o tráfego de dados e o de voz. O ATM iria ser utilizado em todo o lado e em todos os equipamentos, desde o PC pessoal até aos grandes comutadores e encaminhadores do *core*. O ATM pode ter falhado parcialmente devido ao seu projecto. Inicialmente simples, rapidamente se tornou em algo tão complexo e gigantesco que poucos o podiam compreender. No entanto, o maior obstáculo ao ATM foram sem dúvida os "efeitos de rede", ou seja, a inércia à mudança. Teoricamente o ATM tinha todas as

características necessárias para garantir o seu sucesso excepto uma: não era compatível com as redes e aplicações existentes e foi esmagado pelo progressivo sucesso da *Internet*. Nessa altura, a dimensão e o crescimento das redes, o número de aplicações e de utilizadores do IP era já demasiado para este movimento poder ser parado. As LANs utilizam o protocolo *Ethernet*, não por este ser particularmente bom, mas porque toda a gente e todos os equipamento o usam. Por exemplo, a *Gigabit Ethernet* teve que ser meticulosamente adaptada para cumprir os requisitos da *Ethernet*, algo que seria impensável quando Bob Metcalfe a desenvolveu nos anos 70. Qualquer desvio significativo é impraticável. Assim, independentemente das vantagens que o ATM prometia, é hoje uma tecnologia moribunda. Em termos de WAN, a história é semelhante. O *Frame Relay* foi projectado para velocidades até aos 56 Kbps. Após isso passou para os 1,5 Mbps e quando chegou aos 45 Mbps aventou-se que seria o máximo. De alguma forma foi sendo sempre melhorado de modo a permitir velocidades cada vez maiores. A razão é simples: há demasiados equipamentos que "falam" *Frame Relay* para que este possa ser integralmente substituído. Importa no entanto dizer que neste aspecto em particular, o ATM conseguiu ganhar algum terreno e até muito recentemente conseguia ser uma tecnologia em expansão de utilização, nomeadamente nos acessos xDSL. Também ao nível do *core*, por permitir um controlo fino dos recursos, os operadores continuavam a investir parcialmente em equipamentos e redes ATM.

O nível da rede é o nível do IP por excelência. Aqui, o IPv4 dizimou a concorrência, e mesmo o IPv6, uma versão diferente de IP, está a encontrar enormes dificuldades de implementação. Apenas o esforço concertado dos países que mais sofrem com a escassez de endereços IPv4, aliado à necessidade previsível de endereçar de um modo único todos os terminais móveis futuros, poderão fazer vingar esta nova versão do IP. Neste aspecto o IPv6 beneficia do facto de, no fundo, ser baseado nos mesmos conceitos do IP (v4) e tudo estar a ser feito para garantir a compatibilidade de elementos de rede, terminais e aplicações. Este aspecto é fundamental para que a introdução do IPv6 não represente uma quebra com o passado, mas somente um "progresso previsível" do IP.

Há no entanto factores difíceis de ultrapassar. Uma tecnologia, depois de aceite, não é só difícil de mudar, mas também se vai tornando parte do *hardware*. Isto aplica-se tanto a dispositivos terminais, como por exemplo as impressoras, mas também aos elementos de comutação no *core* das redes, os *routers*, que hoje em dia fazem grande parte do seu

processamento inteiramente em *hardware* específico, projectado para desempenhar as suas funções de um modo óptimo e que "sabe" apenas IP (versão 4). O IPv6 terá que ultrapassar estes problemas.

Há mais uma camada protocolar que importa mencionar: a camada aplicacional. Há poucos anos atrás a *Internet* era povoada por muitos protocolos de transferência de informação (FTP - *File Transfer Protocol*, *Gopher*, NNTP - *Network News Transfer Protocol*, etc). No entanto, também aqui se assistiu a uma convergência para um protocolo: o HTTP (*Hyper Text Transfer Protocol*). O HTTP venceu porque foi aquele que, através da *WWW*, conseguiu o maior número de utilizadores. Em paralelo, enquanto o trafego HTTP continuava a crescer, as companhias preocupadas com novas ameaças de segurança, começaram a instalar *firewalls* na entrada das suas redes, bloqueando a maioria dos protocolos, considerados intrusivos e perigosos para a segurança interna, e deixando passar o "inofensivo" HTTP. As empresas de *software* rapidamente se aperceberam disto e de um momento para o outro, tráfego que inicialmente utilizava outros protocolos (por exemplo o FTP) passou a ser transportado via HTTP. Pode dizer-se que o HTTP se tornou o "Esperanto" da *Internet*, tornando, de certa forma, os *firewalls* irrelevantes para o controlo de vários tipos de tráfego.

O HTTP tem agora 14 anos e não se pode de modo algum afirmar que seja um "bom" protocolo. Já não é certamente inovador, mas graças aos "efeitos de rede" e às constantes evoluções que tem sofrido, adivinha-se-lhe uma longa vida.

Em resumo, do ponto de vista das telecomunicações, o caminho para a mudança passa em grande parte pela garantia da compatibilidade e continuidade. É este aspecto que torna relevante a compreensão das tecnologias actualmente existentes.

3.2 Tecnologias de rede integradoras

3.2.1 ATM

Os conteúdos multimédia, compostos essencialmente por áudio e vídeo têm características síncronas. Caso não exista sincronismo, um conteúdo composto por dois fluxos distintos, de som e imagem, poderá parecer um filme mal dobrado. A própria natureza da comunicação implica a existência de um fluxo de informação contínuo, quase uniforme, que necessita de ser transmitido a uma velocidade aproximadamente constante

para que não sejam detectados cortes e distorções. No entanto, a maioria das aplicações que envolvem transferência de dados são assíncronas, transmitindo rajadas de informação seguidas por períodos mais ou menos longos de inactividade: a diferença entre a taxa de transferência média e o pico pode ser muito grande.

O ATM (*Asynchronous Transfer Mode*) foi adoptado pelo CCITT (*Comité Consultatif International Téléphonique et Télégraphique*, actualmente designado ITU-T) [ITUT] em 1989 para ser a norma internacional de suporte à RDIS de Banda Larga (Rede Digital com Integração de Serviços de Banda Larga). O ATM é o resultado de um compromisso obtido pela conjunção de características e requisitos entre todos os tipos de tráfego (dados, voz e vídeo) de modo a encontrar um denominador comum que satisfaça todos. Por outro lado, este mesmo compromisso faz com que o ATM não seja óptimo para nenhum tipo de tráfego. Embora o ATM tenha em grande medida fracassado, lançou conceitos que influenciaram o IP e as tecnologias de adaptação (nível 2.5), razão pela qual faremos uma breve descrição desta tecnologia.

A grande inovação trazida pelo ATM foi a capacidade de suportar o fornecimento de serviços multimédia, devido aos altos débitos associados e à possibilidade de diferenciar QoS (Qualidade de Serviço). Teoricamente não existe limite para a velocidade de transmissão permitida pelo ATM. Por outro lado, o ATM introduziu os conceitos de classes de serviço de modo a dar tratamentos distintos a serviços com características distintas. Pode afirmar-se que o ATM foi o percursor da maioria dos conceitos que são aplicados agora ao IP e às redes sem fios de próxima geração (UMTS).

O ATM é uma tecnologia orientada à ligação, que se baseia na transmissão de pacotes de tamanho fixo de 53 *bytes*, denominados células. Utiliza multiplexagem de células. Um fluxo é partido em pequenos fragmentos e colocado nas células de tamanho fixo, tendo cada uma delas a informação necessária para a encaminhar até ao seu destino. O facto do tamanho dos pacotes ser fixo facilitou o desenvolvimento dos equipamentos ao nível do sincronismo e descodificação dos pacotes. Dos 53 *bytes* do pacotes, 5 são sempre utilizados para o cabeçalho, e dos restantes 48, há ainda alguns que são utilizados pela camada AAL (*ATM Adaptation Layer*). A camada AAL é a responsável pelo acomodamento e adaptação dos diferentes tipos de tráfego às características da transmissão em ATM. Existem ainda outros *bytes* que são desperdiçados com enchimento, quando não há informação suficiente para completar os 48 *bytes* do campo de dados. Por isto,

facilmente se pode deduzir que uma das grandes desvantagens do ATM é o seu enorme *overhead*.

Com tamanhos de célula pequenos, juntamente com taxas de transferência elevadas, é possível transportar tráfego sensível a atrasos, conjuntamente com tráfego com muitas rajadas. A voz e o vídeo são transportados sem cortes ou interferências, e o tráfego de dados pode utilizar largura de banda a pedido. Na qualidade de tecnologia de transporte universal, o ATM pode então ser instalado quer no *desktop*, ao nível departamental, de um *campus*, de uma rede de *core* ou mesmo nas interligações das "super auto-estradas" da informação.

O ATM introduziu também nas redes de pacotes um esquema de endereçamento hierárquico baseado no E.164 e que permite a auto-configuração dos terminais, adicionando ao prefixo de rede anunciado pelos NASs (*Network Access Servers*) o seu ESI (*End Station Identifier*) (o endereço MAC - *Medium Access Control*). Este esquema de endereçamento veio mais tarde a ser utilizado pelo IPv6, sofrendo contudo algumas adaptações.

As células só podem ser transmitidas após estarem completamente preenchidas (este é o motivo pelo qual se faz o enchimento) e nessa altura são transmitidas à velocidade máxima da ligação. Contudo, em cada ligação física podem coexistir muitas ligações lógicas distintas, tendo cada uma as suas características particulares. É possível que num dado troço, a soma das velocidades de pico de cada uma das ligações lógicas seja superior à capacidade da ligação física, isto é, pode ser utilizada multiplexagem estatística. O ATM introduz ainda o conceito de canal lógico (VCC – *Virtual Channel Connection*) dentro de uma ligação lógica (VPC – *Virtual Path Connection*). As sessões ATM tomam lugar sobre circuitos virtuais (sendo virtuais não estão limitados a uma qualquer ligação física) que após estarem estabelecidos assim se mantêm até ao final da sessão. A grande maioria dos serviços ATM são fornecidos sobre circuitos virtuais permanentes (PVC – *Permanent Virtual Channel*), que exigem uma configuração explícita por parte do operador, a menos que se tratem de redes completamente privadas. A grande promessa de disponibilidade de largura de banda a pedido, chegaria apenas no dia em que os operadores vendessem *Switched Virtual Channels* (SVCs). Os PVCs podem ser comparados a circuitos dedicados, enquanto que os SVCs podem ser comparados ao serviço de "*dial-up*". No entanto, um SVC leva apenas uma pequena fracção de segundo a ser estabelecido.

Contudo, desde cedo se percebeu que para o ATM ter alguma hipótese de chegar ao *desktop* seria necessário garantir a compatibilidade com as aplicações existentes e, após isso, esperar que as próprias aplicações (que utilizavam essencialmente o IP) fossem, progressivamente, migradas para utilizarem o ATM de uma forma nativa. Surgiram então as normas de emulação de LAN, o LANE (*LAN Emulation*), para que o ATM estivesse disponível para aplicações que funcionassem sobre *Ethernet* e *Token Ring*. Houve ainda produtos para fazer tradução entre ATM e *Frame Relay* a serem desenvolvidos. O modo de transportar IP e ARP (*Address Resolution Protocol*) sobre ATM está descrito em [Laub94].

O ATM introduziu conceitos muito importantes de gestão de recursos de rede que vieram a ser adaptados e adoptados para as redes IP. Em termos de QoS o ATM utiliza mecanismos de CAC (*Call Admission Control*) e mecanismos de policiamento de tráfego, denominados de mecanismos de UPC (*Usage Parameter Control*). Foram definidas quatro classes de serviço, correspondentes aos vários tipos de dados a transmitir. As características diferenciadoras destas classes de serviço são obtidas pela combinação de diferentes parâmetros de QoS. Os parâmetros de QoS que podem ser considerados numa rede ATM são o rácio de células com erros (CER - *Cell Error Ratio*), o rácio de células perdidas (CLR - *Cell Loss Ratio*), o ritmo de células mal inseridas (CMR - *Cell Misinsertion Rate*), o rácio de blocos de células gravemente afectados por erros (SECBR – *Severely Errored Cell Block Ratio*), o atraso de transferência (CTD – *Cell Transfer Delay*) e a variação do atraso das células (CDV – *Cell Delay Variation*).

O ATM faz a multiplexagem de vários fluxos baseada na prioridade de cada um deles. É também ao nível da camada ATM que se realiza a identificação de congestionamento, de gestão de falhas e de gestão de tráfego.

No nível físico, o ATM suporta fibra óptica multimodo e modo único, STP (*Shielded Twisted Pair*), cabo coaxial e UTP (*Unshielded Twisted Pair*) às velocidades que se deseje, que seja possível pagar, ou que sejam realizáveis fisicamente. O tráfego ATM pode ainda ser transportado em ligações SONET (*Synchronous Optical NETwork*) ou SDH (*Synchronous Digital Hierarchy*) (155Mbps), FDDI (*Fiber Distributed Data Interface*) (100Mbps), DS3 (45Mbps), E3, E1, etc. O ATM Forum [ATMF] decidiu adoptar, sempre que possível, as normas de transmissão físicas existentes. O ATM não está limitado em termos de concepção a nenhuma velocidade de transmissão em particular, nem a nenhum tipo de meio físico de transmissão.

O ATM continua a ser utilizado hoje em dia como tecnologia de nível 2 de suporte ao transporte de dados, mas já não é visto como o integrador que se esperava vir a ser e a tendência actual aponta para o seu desaparecimento em poucos anos.

3.2.2 O IP como Tecnologia Integradora

Na concepção do protocolo IP, há três décadas, ninguém imaginaria as necessidades emergentes da sociedade em relação aos sistemas de comunicação actuais, em particular a *Internet*. A sua utilização alargada tornou-se essencial para o desenvolvimento do conhecimento e, também da economia. No entanto, o rápido crescimento da *Internet* impõe que novas medidas sejam tomadas de forma a satisfazer os requisitos cada vez mais exigentes da "sociedade da informação". As tentativas de alargar o âmbito de aplicação do protocolo IP cresceram rapidamente. No entanto, tendo sido originalmente concebido para a comunicação de dados simples entre computadores, o IP veio a demonstrar limitações funcionais e de escalabilidade.

3.2.2.1 Principais deficiências do IPv4

O rápido crescimento da *Internet* levou a situações de esgotamento do número limitado de endereços IP disponíveis. Outros aspectos, essenciais hoje em dia e em cenários futuros, tais como interoperação com outras redes tradicionais, segurança, qualidade de serviço e mobilidade não foram também originalmente incluídos no protocolo. Surgiram assim diversas soluções técnicas para colmatar estas lacunas e permitir a aplicação do protocolo IP no maior número possível de aplicações.

Dos vários problemas atrás mencionados, a escassez de endereços disponíveis tem sido um dos que tem levantado maiores problemas. O IPv4 tem um espaço de apenas de 32 bits para a representação de endereços, o que permite definir um pouco mais de 4 000 milhões de endereços. Se atendermos ao actual crescimento da *Internet* e à tendência de cada vez mais dispositivos terem necessidade de possuir um endereço IP, este espaço esgotar-se-á rapidamente. Como referência tem-se utilizado a regra "um IP por habitante" para demonstrar que o espaço disponível não é suficiente. Exige-se que este endereço seja "globalmente único", ou seja, que não exista nenhum outro dispositivo que tenha o mesmo endereço. Esta exigência, embora seja em muitos casos desnecessária, advém da necessidade que certo tipo de aplicações (interactivas e bidireccionais) têm de poderem ser iniciadas por qualquer um dos extremos da ligação. Assim, para que cada terminal possa

ser contactado em qualquer altura, terá que possuir um endereço que seja conhecido universalmente.

A acrescer ao número limitado de endereços disponíveis, existe ainda o facto de que a sua distribuição, quer geográfica, quer institucional, não foi devidamente planeada de modo a atender ao crescimento verificado. A atribuição destes endereços, especialmente na fase inicial da "*Internet*" foi feita de forma pouco estruturada e sem um planeamento realista, coadjuvado com o facto de na altura não se poder prever de todo o crescimento e importância que este protocolo viria a ter uns anos mais tarde. Assim, grandes fatias do espaço total de endereçamento foram entregues a entidades que nunca os usarão mas que por outro lado não estão dispostas a abdicar deles. Tome-se por exemplo o caso de algumas empresas que têm atribuídas classes A de endereços (16777214 endereços por cada classe), ou seja, mais do que aqueles que estão disponíveis para continentes inteiros (por exemplo África). Assim, existe uma distribuição muito pouco optimizada, o que conduzirá ainda mais rapidamente ao esgotamento dos endereços disponíveis.

Também como consequência directa da péssima distribuição dos endereços as tabelas de encaminhamento estão saturadas, levando à degradação do desempenho da rede. O simples facto dos endereços não serem delegados de uma forma contínua dentro de cada continente e país conduz a situações de tabelas de encaminhamento muito grandes devido ao facto de ser necessário haver uma rota para cada bloco de endereços que não seja contíguo.

As novas aplicações e formas de exploração comercial da *Internet* levaram também ao aparecimento de novas exigências da rede. Assim, questões como a diferenciação da qualidade de serviço, ou seja, dar tratamento preferencial ao tráfego gerado por determinado tipo de aplicações ou utilizadores face a outro tráfego, é desde há alguns anos, uma questão fundamental que o protocolo IP, da forma como foi originalmente definido, não estava preparado para resolver.

Por outro lado, as tendências actuais de portabilidade de terminais (os PDAs - *Portable Digital Assistants* – e telemóveis são o exemplo mais relevante) exigem do protocolo IP uma outra característica que não lhe é natural: a mobilidade.

Nos parágrafos seguintes vamos abordar algumas soluções para esta problemática.

3.2.2.2 Soluções para o endereçamento em IPv4

Face aos problemas descritos foram propostas algumas soluções para colmatar estas deficiências do protocolo IP. Nomeadamente, em relação à má distribuição que vinha a ser feita dos endereços IP, foi adoptado o CIDR (*Classless Inter-Domain Routing*) [Full93] que consiste em delegar blocos de endereços contíguos a regiões do planeta (Europa, Ásia, etc.) que, por sua vez, são divididos em blocos menores, ainda de forma contígua, até que cada rede tenha a sua gama. Desta forma é atingida uma melhor organização do espaço de endereçamento, por forma a reduzir o tamanho das tabelas de encaminhamento. Contudo, os erros cometidos no passado não podem ser corrigidos.

Por outro lado, para minorar a falta de endereços IP, foram desenvolvidos e implementados diversos mecanismos:

- Atribuição dinâmica do endereço IP (*Dynamic Host Configuration Protocol* ou DHCP) [Drom97] – permite usar uma gama de endereços atribuídos às máquinas de acordo com as necessidades reais e, em princípio, apenas as máquinas ligadas têm um endereço atribuído. Desta forma é possível fazer uma partilha de determinado número de endereços por um número maior de máquinas, recorrendo à multiplexagem estatística. No entanto, não se consegue garantir que a mesma máquina tenha sempre o mesmo endereço atribuído, o que é um forte impedimento para serviços que necessitem que o terminal seja "bem conhecido" e seja identificado sempre pelo mesmo endereço. A filosofia "*always-on*", com os terminais permanentemente ligados, traz também problemas a esta abordagem.
- Tradução de endereços entre redes privadas e a *Internet* – este mecanismo existe sob duas formas: *Network Address Translation* (NAT) [Sris01] e *Port Address Translation* (PAT). Estes mecanismos permitem que um grupo de utilizadores ligados a uma rede privada possam aceder à *Internet* usando um bloco restrito de endereços, atendendo ao facto que nem todos acedem à *Internet* ao mesmo tempo (NAT) ou que nem todos os portos são utilizados para esse acesso (PAT). No entanto, este mecanismo limita funcionalidades de extremo-a-extremo da *Internet* e reduz o seu desempenho global. Com efeito, um utilizador que utiliza o NAT pode comunicar com servidores na *Internet* mas não tem a garantia de estar sempre acessível quando dispositivos

externos tentarem estabelecer uma ligação, ou seja, a bidireccionalidade de acesso (necessária para determinadas aplicações) não é garantida.

- *Proxies* – solução que permite o acesso de muitos terminais em simultâneo à *Internet* utilizando apenas um endereço global. Um *proxy* é uma entidade na qual os terminais delegam a responsabilidade de obter e lhes passar a informação que eles necessitam. Tal como no caso de NAT/PAT, o uso de *proxies* tem a mesma limitação de bidireccionalidade, aliada ao facto de a utilização de algumas aplicações não ser suportada.

Em relação às deficiências inerentes ao IP em termos de qualidade de serviço e de suporte de mobilidade, houve também a definição de várias estratégias que serão analisadas com mais detalhe mais à frente neste capítulo e no capítulo seguinte.

3.2.2.3 Novas exigências: o caminho para o IPv6

A esmagadora maioria dos serviços utilizados sobre a *Internet* são de acesso a servidores e iniciados pelo utilizador. Até há pouco tempo o tipo de acesso à *Internet* era dominado pela utilização de *modems* através da rede pública de voz (*dial-up*). Actualmente emerge outro tipo de aplicações que requer a comunicação directa entre utilizadores extremo-a-extremo (ausência de servidores). Existe mesmo a necessidade de ligações à *Internet* permanentes (serviços *always-on*), o que exclui a possibilidade de usar a tradução de endereços pelos mecanismos NAT/PAT e a partilha de endereços através da atribuição temporária.

O problema da falta de endereços IP e do tamanho das tabelas de encaminhamento na *Internet* foi identificado e discutido no âmbito do IETF (*Internet Engineering Task Force*) no início da década de 90. Desde então, tem sido definido e implementado um conjunto de soluções para minimizar (ou adiar) o problema, tais como, o CIDR, DHCP e NAT referidos anteriormente. No entanto, a solução encontrada para resolver o problema a longo termo foi desenvolver um novo protocolo que, além de estender o espaço de endereçamento, colmatasse algumas limitações do IPv4. Neste contexto, foi criado um grupo de trabalho [IPv6a] para definir uma nova versão do IP, o IPv6 .

O IPv6 [Deer98] apresenta um conjunto de características que o diferenciam do IPv4:

- Capacidade de Endereçamento Expandida;

- Simplificação do Formato de Cabeçalho;
- Melhor Suporte para Cabeçalhos de Extensão e Opções;
- Capacidade de identificação de fluxos;
- Capacidades de Autenticação e Privacidade;
- Capacidades de auto-configuração.

O IPv6 aumenta o espaço de endereçamento de 32 bits para 128 bits permitindo não só definir um maior número de endereços, mas também mais níveis hierárquicos de endereçamento. O objectivo é maximizar a agregação da informação de endereçamento e, consequentemente, limitar a expansão das tabelas de encaminhamento.

Um pacote IPv6 (Figura 3) pode conter uma série de cabeçalhos seguidos da informação do utilizador (campo de dados). Um pacote começa sempre por um cabeçalho de base que pode ser seguido por outros, chamados cabeçalhos de extensão. Estes cabeçalhos de extensão são opcionais e transportam informação que não é estritamente necessária a todos os dispositivos de encaminhamento (*routers*) ao longo do percurso de um pacote na rede. Deste modo, reduzem-se os campos de informação do cabeçalho de base (contém apenas os endereços origem e destino e mais 6 campos de informação), fazendo que o seu processamento nos *routers* seja mais eficiente.

Além disso, os cabeçalhos de extensão trazem uma maior flexibilidade e permitem a introdução futura de novas opções não previstas até à data.

A fragmentação, que no caso do IPv4 é permitida em qualquer ponto da rede, no IP6 apenas pode ser efectuada na origem, recorrendo para o efeito a um cabeçalho de extensão. Assim, o processo de encaminhamento é mais rápido pois evita que os *routers* desperdicem recursos neste tipo de tarefas. No IPv6 estão previstos mecanismos que permitem a descoberta do MTU (*Maximum Transmission Unit*) mínimo para um dado segmento do percurso de um pacote na rede, permitindo desta forma formatar o tamanho máximo do pacote logo à partida.

Outra diferença significativa que melhora o desempenho do IPv6 na rede é a ausência de cálculo do *Checksum* nos *routers* intermédios ao longo de um percurso de um pacote na rede. No IPv6, não existe nenhum campo no cabeçalho para o *Checksum* a fim de verificar a integridade dos dados. No IPv4, em cada salto, o *Checksum* é calculado antes do pacote ser encaminhado. A decisão da sua eliminação prendeu-se com o facto de ser

uma tarefa relativamente pesada e desnecessária já que, quer a camada inferior (ligação lógica) quer principalmente a superior (protocolos de transporte), asseguram essa mesma integridade. Evita-se desta forma trabalho que seria redundante.

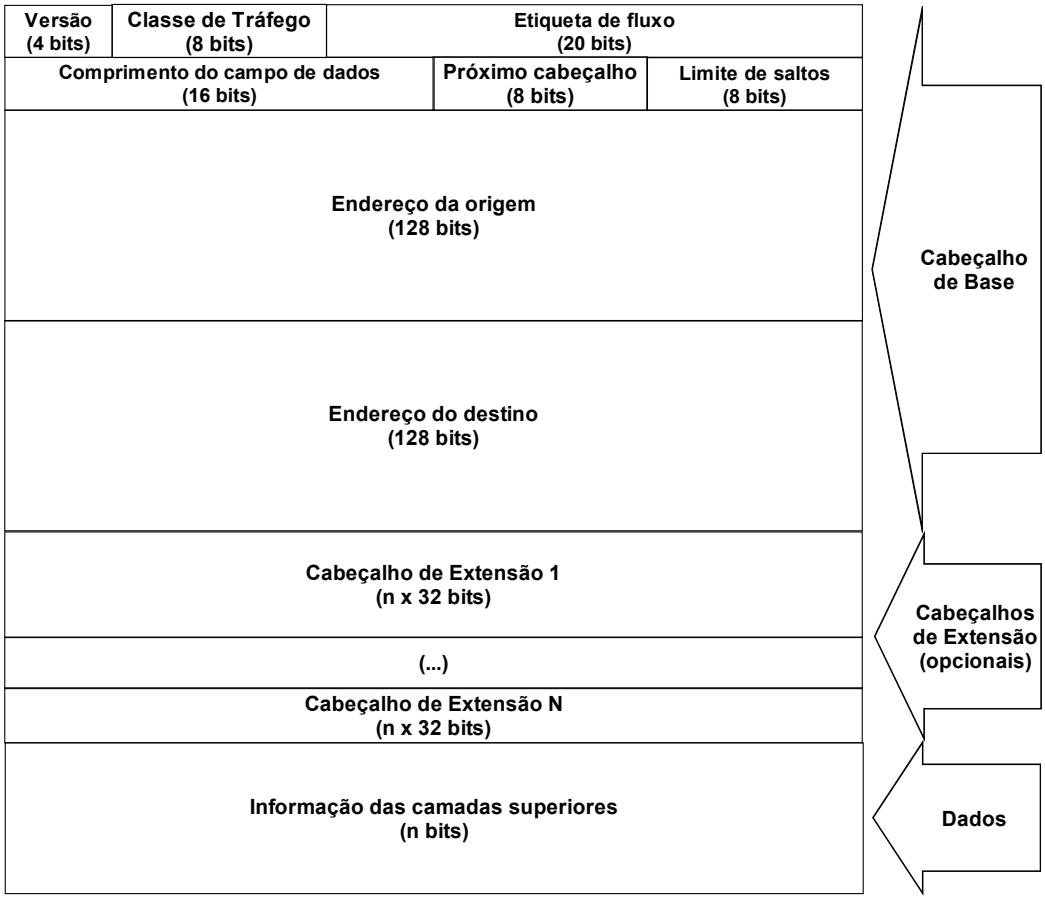


Figura 3: Pacote IPv6

Também no que respeita à segurança, o IPv4 e o IPv6 têm abordagens diferentes. Nas redes IPv4 a segurança foi deixada apenas a cargo das aplicações. Entretanto, tornou-se cada vez mais premente ter mecanismos de autenticação, integridade e confidencialidade dos dados, em particular na *Internet*. Neste sentido, foi definido e desenvolvido um protocolo de segurança, o IPSec (*IP SECurity*) [IPSE], capaz de funcionar conjuntamente com IPv4. Contrariamente ao que acontece com o IPv4, no IPv6, o suporte de IPSec é nativo ao próprio protocolo. A sua implementação é suportada por dois campos de extensão que foram definidos para esse efeito.

Em termos de suporte de qualidade de serviço, à semelhança daquilo que acontece em IPv4, o IPv6 oferece essencialmente duas abordagens para que a rede possa

disponibilizar QoS extremo-a-extremo aos seus utilizadores: Serviços Integrados (*Integrated Services* – IntServ) [Brad94] e os Serviços Diferenciados (*Differentiated Services* – DiffServ) [Carl98] (Estes e outros modelos de fornecimento de QoS serão detalhados no próximo capítulo). Em relação ao IPv4, o IPv6 apresenta um campo de 20 bits chamado Etiqueta de Fluxo (*FlowLabel*). Este campo poderá ser utilizado para efeitos de QoS, quer em termos de IntServ quer de DiffServ, para uma classificação dos fluxos mais eficiente. A forma de utilização desse campo é actualmente uma das questões mais discutidas nos grupos de IPv6, pelo que ainda nada definitivo está definido.

Resumindo, há alguns aspectos que podem contribuir para que os pacotes IPv6 tenham um tratamento mais expedito e eficiente nos nós de rede: a não fragmentação dos pacotes nos nós intermédios, a ausência de cálculo do *Checksum*, a diminuição do tamanhos das tabelas de encaminhamento e a simplificação do cabeçalho dos pacotes IPv6.

Finalmente, outra vantagem do IPv6 diz respeito às capacidades de auto-configuração. A configuração de endereços é um problema complexo em qualquer organização e a auto-configuração inerente ao IPv6 surge como um método eficaz de atribuição ou reconfiguração automática de endereços.

3.2.2.4 A transição do IPv4 para o IPv6

A transição para o IPv6 tem sido largamente discutida. Esta transição terá de ser gradual e a coexistência das duas versões do IP irá durar vários anos. Como estas duas versões são incompatíveis, foram desenvolvidas técnicas, denominadas por mecanismos de transição, que garantem a coexistência dos dois protocolos e o interfuncionamento de ambos. Estes mecanismos podem dividir-se em 3 grupos: *Dual Stack*, túneis e tradução.

O mecanismo *Dual Stack* consiste em introduzir o IPv6 na rede sem comprometer o funcionamento de toda a infra-estrutura IPv4 já existente. Dada a facilidade actual em introduzir novos protocolos em equipamentos terminais e de rede, pode ser mantido em paralelo o funcionamento do IPv4 e do IPv6. Assim, as aplicações antigas podem continuar a usar o IPv4 enquanto as novas aplicações podem fazer uso quer do IPv4 quer do IPv6. Contudo, esta estratégia não permite a interoperabilidade entre os dois protocolos ao nível da rede. Quando uma aplicação que suporta os dois protocolos pretende comunicar com outra que suporta apenas o IPv4, a comunicação é estabelecida com o IPv4. Caso contrário, usa o IPv6. Não é portanto possível que uma aplicação que suporta apenas IPv4 interaja com uma que suporte apenas IPv6.

Para a interligação de redes ou terminais IPv6 sobre uma infra-estrutura IPv4 foram desenvolvidos mecanismos baseados em túneis. Este tipo de mecanismo consiste em transmitir um pacote IPv6 como se de facto se tratasse de dados de um pacote IPv4. Deste modo, os *routers* IPv4 processam o tráfego sem o conhecimento da existência do IPv6. Na recepção, o pacote IPv6 é recuperado e encaminhado para o destino final. Existem várias implementações de mecanismos baseados em túneis (Túneis Automáticos, 6to4, *Tunnel Broker*, etc.), diferindo apenas no modo e na facilidade de configuração.

A utilização dos mecanismos de tradução permite a comunicação entre um domínio de rede IPv6 com um domínio IPv4, e vice-versa. Um desses mecanismos é o NAT-PT (*Network Address Translation – Protocol Translation*), baseado na filosofia do NAT, discutido anteriormente. Este mecanismo é instalado na fronteira entre os dois domínios de rede. Os terminais conhecem apenas um protocolo: IPv6 ou IPv4. O NAT-PT faz tradução de endereços e de cabeçalhos IPv4 e IPv6. Um dos problemas associados a este mecanismo está relacionado com o facto de existirem aplicações que não são totalmente independentes da camada de rede (por exemplo, FTP e DNS - *Domain Name Service*). A solução para este problema é o uso de ALGs (*Application Layer Gateways*) que, ao nível da aplicação, são responsáveis pela tradução dos campos necessários.

Presentemente, o protocolo IPv6 começa a ser utilizado [IPv6], embora de uma forma muito lenta devido ao facto do IPv4 responder ainda às exigências actuais. Além disso, a migração para a nova versão é complexa e envolve investimentos avultados. No entanto, este processo está já a ser acelerado de forma a não prejudicar o crescimento e desenvolvimento futuro da *Internet*. Vários países, nomeadamente os asiáticos (onde a falta de espaço de endereçamento IPv4 se faz notar com mais veemência) estão a impulsionar activamente a introdução do IPv6 com grandes incentivos financeiros. Também a Comunidade Europeia iniciou já um processo de apoio à introdução do IPv6. Do outro lado do Atlântico, também se iniciaram os esforços de migração. O Departamento de Defesa Norte Americano impõe que qualquer equipamento adquirido após Outubro de 2003 tenha suporte de IPv6, e está previsto que em 2008 toda a sua rede seja inteiramente suportada pelo novo protocolo. No entanto, nada disto garante só por si que o IPv6 será bem sucedido. O factor de sucesso do IPv6 está intimamente ligado ao sucesso das redes de próxima geração, onde se espera que a voz seja totalmente suportada sobre IP. Assim, o número de terminais a utilizarem o IP poderá crescer exponencialmente:

bastava apenas que cada um dos actuais telemóveis passasse a necessitar de um endereço IP para que o IPv4 automaticamente esgotasse os seus endereços. Mesmo neste caso, o IPv6 poderá ter ainda uma luta a travar contra o SIP (*Session Initiation Protocol*) [Rosb02] que pode, através da utilização de URLs (*Uniform Resource Locator*) para identificar determinado utilizador/terminal, colmatar a necessidade de os terminais possuírem um endereço público distinto.

3.3 Tecnologias de Suporte ao Fornecimento de Serviço

Conforme foi referido no capítulo anterior, o fornecimento de serviço impõe a existência de mecanismos que permitam o controlo quer da identificação das partes, quer da correcta contabilização e cobrança (no caso de serviços pagos) dos recursos e serviços utilizados. Nesta secção são analisados os protocolos RADIUS (*Remote Authentication Dial-In User Service*) [Rign00] e DIAMETER [Calh01].

3.3.1 RADIUS

O RADIUS é um protocolo para troca de informação de autenticação, autorização, configuração e contabilização entre um servidor (normalmente um *router*) de acesso (NAS – *Network Access Server*) e um servidor de AAA (*Authentication, Authorization and Accounting*). O seu funcionamento encontra-se sumariado na Figura 4 e Tabela 1.

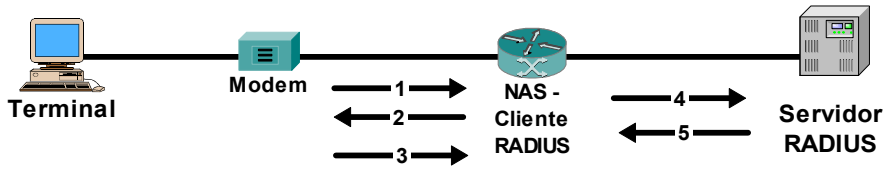


Figura 4: Exemplo do modelo RADIUS

Mensagem	Explicação
1	O utilizador/terminal inicia a sessão de PPP até ao NAS
2	O NAS responde, pedindo o *username* e a *password* (caso utilize *Password Authentication Protocol* [PAP]) ou com um desafio (caso utilize *Challenge Handshake Authentication Protocol* [CHAP]).
3	O utilizador (terminal) responde.
4	O Cliente RADIUS envia o nome do utilizador e a palavra chave cifradas para o servidor RADIUS.
5	O servidor RADIUS responde com Aceite, Rejeitado ou Desafio.

Tabela 1: Resumo da operação de autenticação do protocolo RADIUS

O servidor RADIUS gere uma base de dados onde é guardada a informação de autenticação dos utilizadores (nome e palavra chave) e os serviços permitidos pelos seus perfis (por exemplo SLIP - *Serial Line Internet Protocol*, PPP – *Point-to-Point Protocol*, telnet, rlogin, HTTP, acesso a determinados endereços IP, etc.). Desta forma o RADIUS fornece o necessário suporte administrativo de AAA. O RADIUS é bastante utilizado e existem bastantes implementações. No entanto, devido a não ter sido projectado para cenários de dimensões elevadas, sofre de degradação de desempenho e perda de informação quando é utilizado em sistemas de grande escala.

As principais características do RADIUS são as seguintes:

- Utilizado para autenticação de utilizadores de "*dial-in*": orientado para o acesso a redes privadas de grandes empresas. No entanto, como foi projectado para o acesso de utilizadores "*dial-in*", o protocolo RADIUS não permite facilmente a noção de *roaming* (necessário em cenários de próxima geração) entre redes distintas (i.e., pertencentes a operadores/companhias distintas).
- Utiliza um modelo de cliente/servidor. O NAS funciona como cliente RADIUS e é este que deve iniciar o pedido para o servidor. Desta forma não é possível os clientes receberem mensagens não solicitadas (muito úteis, em particular para o fornecimento de serviços pré-pagos).
- Os servidores RADIUS são responsáveis pela recepção e atendimento de pedidos dos clientes (NAS), autenticação dos utilizadores e devolver toda a informação necessária para que o cliente RADIUS possa fornecer o serviço ao utilizador. Assim, após a emissão de um pedido, o cliente RADIUS permanece num estado "passivo".
- Os clientes e servidores partilham uma chave (segredo), e operam segundo um modelo salto-a-salto. Independentemente do número de nós entre o cliente e o servidor "final", entre cada dois nós intermédios (em que um actua como cliente e o outro como servidor) é estabelecido um canal seguro baseado num segredo partilhado. Desta forma, cada nó deve decifrar a informação com uma chave e cifrar com outra, antes de a enviar para o nó seguinte (no sentido do servidor). O mecanismo de distribuição de chaves (de modo a fornecer o segredo partilhado a ambos os nós) não está definido pelo protocolo.

- A informação sensível é protegida por MD5. O RADIUS fornece mecanismos de autenticação, utilizando cifra MD5 sobre um conjunto de octetos, incluindo o segredo partilhado. Nos pedidos, a palavra chave do utilizador é cifrada com MD5, fornecendo a confidencialidade necessária. Não existe qualquer outro mecanismo de cifra.
- O RADIUS utiliza o UDP como protocolo da camada de transporte. Os requisitos de tempo de resposta do RADIUS são bastante distintos daqueles que o TCP fornece (em particular, a retransmissão agressiva e a entrega fiável – eventualmente esperando alguns minutos – não são particularmente úteis). Adicionalmente, o UDP simplifica a implementação do servidor por este não manter informação de estado. O RADIUS considera a camada de transporte como não segura e implementa os seus próprios mecanismos de transmissão segura. Assim, o uso de TCP é desencorajado uma vez que irá apenas replicar uma funcionalidade já existente no RADIUS, incrementando a complexidade da arquitectura.
- Utilizam-se temporizadores de "*keep-alive*" com alguma frequência para verificar se determinado servidor está ou não activo. Contudo, esta prática é desencorajada uma vez que provoca um aumento da carga na rede e principalmente no servidor, e não acrescenta nenhuma funcionalidade essencial.
- Mensagens que não contenham informação nenhuma ou mensagens com erros são descartadas silenciosamente.
- As mensagens RADIUS são baseadas em AVPs (*Attribute Value Pairs*) com alinhamento de 32 *bits*.

No entanto, o protocolo RADIUS apresenta algumas limitações [Calh02]:

- Limitação estrita do tamanho do campo de atributos: o campo do tamanho do atributo no cabeçalho do pacote RADIUS é de apenas um *byte*. Está especificado no protocolo que dados que ultrapassem as dimensões previstas possam ser repartidos por vários AVPs, mas este processo introduz novos problemas. É também permitido que múltiplos atributos de um mesmo tipo sejam enviados na mesma mensagem, dificultando a tarefa de um servidor ou cliente de determinar se de facto se tratam de múltiplos atributos independentes ou apenas um atributo fragmentado.

- Limitação estrita de número de mensagens concorrentes pendentes: o protocolo determina que o campo identificador, presente no cabeçalho, é utilizado para identificar retransmissões. O facto deste campo ter apenas um *byte* impõe uma limitação de 256 pedidos pendentes em cada instante. Além disso, o protocolo RADIUS requer que qualquer retransmissão que inclua modificações no pacote transmitido tenha um identificador distinto. Assim, foi incluído um novo valor no campo ID. Desta forma, o número total de sessões é ainda mais reduzido. As várias implementações de servidores RADIUS tiveram que ser devidamente planeadas de modo a que o campo ID seja alterado aquando das retransmissões que contêm informação actualizada.
- Inaptidão para controlar fluxos para os servidores: dada a natureza de transmissões não programadas, potencialmente repentinas e em grande número, os servidores não têm forma de gerir os seus *buffers* de recepção. Este problema deve-se em parte ao facto de o UDP não incluir nenhum mecanismo de janelas.
- Detecção de falhas limitada e descarte silencioso de pacotes: o protocolo RADIUS não inclui nenhum método para um NAS detectar se a falta de resposta de um servidor se deve a problemas de congestionamento ou falha no servidor. O descarte silencioso de mensagens faz com que o NAS assuma que o servidor local não está contactável o que provoca a retransmissão de todos os pedidos pendentes para servidores alternativos.
- O suporte de servidores alternativos é limitado e ineficiente. Como o uso de mensagens de "*keep-alive*" é desencorajado, um NAS não tem a possibilidade de saber de antemão se um determinado servidor alternativo está ou não acessível e operacional. Assim, se dois ou mais servidores consecutivos estiverem indisponíveis, poderá acontecer uma negação de serviço prolongada.
- Uso ineficiente de servidores em ambientes com *proxies*. Os NASs não têm forma de descobrir e saber se a falha de resposta de um servidor se deve a uma falha nos *proxies* no caminho ou no servidor.

- Não há possibilidade de gerar mensagens "não solicitadas": um servidor não poderá enviar mensagens não solicitadas (apenas pode responder a pedidos) para o NAS. Com o aumento da complexidade dos serviços de rede (por exemplo, serviços pré-pagos), esta limitação forçou os fabricantes a se desviarem do protocolo e encontrar soluções alternativas.
- Segurança salto-a-salto: o RADIUS não permite a existência de segurança extremo-a-extremo, entre o NAS e o servidor RADIUS de origem (no caso de utilizadores em "*roaming*" é o servidor onde os seus dados estão guardados, na sua rede de origem). Desta forma é possível que servidores no caminho da mensagem (*proxies*) possam modificar componentes críticos das mensagens.
- É obrigatória a existência de um segredo partilhado. Mesmo que exista segurança ao nível do IP a assegurar a segurança da comunicação, é sempre exigida uma chave partilhada, ou seja, apesar das camadas inferiores facultarem a necessária segurança, ao nível das camadas superiores o mecanismo de autenticação permanece inalterado.

3.3.2 Diameter

Com o surgimento de novos requisitos de AAA nos NAS, concluiu-se que o RADIUS tem graves deficiências, especialmente quando se colocam cenários inter-domínio. O protocolo DIAMETER surge assim na sequência do RADIUS, resolvendo as deficiências demonstradas pelo RADIUS.

Apesar do RADIUS e o DIAMETER não possuírem um *PDU (Packet Data Unit*) comum, para rentabilizar os investimentos realizados no passado e proporcionar uma migração suave, é possível reservar os códigos RADIUS no espaço de numeração do DIAMETER. É garantida assim a continuidade da operação das redes actuais, enquanto a migração é realizada, e sem que os utilizadores sofram qualquer perturbação de serviço.

3.3.2.1 Arquitectura

A arquitectura do protocolo DIAMETER consiste numa base protocolar, enriquecida através de várias extensões (Figura 5). As funcionalidades básicas e mais comuns são implementadas na base do protocolo enquanto que as funcionalidades específicas relacionadas com determinadas aplicações podem ser implementadas como extensões.

3.3.2.2 Protocolo Base

O protocolo base deve ser suportado por todas as aplicações DIAMETER e define o formato básico do PDU, algumas primitivas e os serviços básicos de segurança oferecidos pelo protocolo.

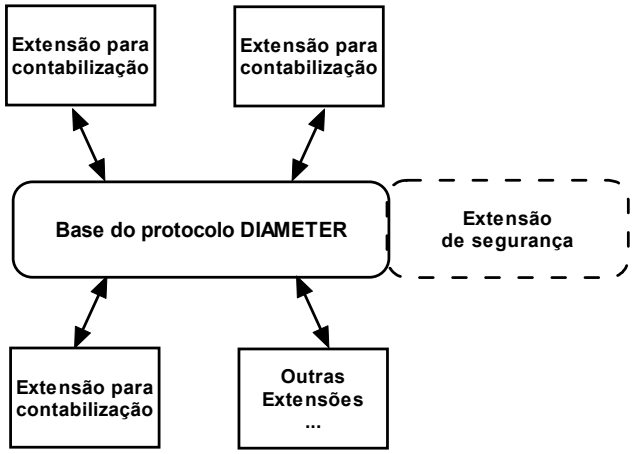


Figura 5: Arquitectura do protocolo DIAMETER

Toda a informação transportada pelo DIAMETER, à semelhança do RADIUS está na forma de AVPs (*Attribute Value Pairs*). Alguns dos valores dos AVPs são utilizados pelo próprio protocolo, enquanto que outros distribuem informação associada com aplicações específicas que utilizam o protocolo DIAMETER. Tipicamente, a informação transportada pelos AVPs é a de autenticação, autorização e de contabilização de recursos. A base do protocolo fornece as funcionalidades que respondem aos requisitos mínimos em termos de um protocolo de transporte de AAA. Ainda assim, a sua utilização não deve ser em modo isolado, mas sim com extensões de aplicações específicas. As suas características básicas são:

- Modelo de comunicação extremo-a-extremo em que a comunicações pode ser iniciada por qualquer um dos extremos. A entidade ou dispositivo que inicia um pedido é designado por cliente DIAMETER. O servidor DIAMETER é o dispositivo que responde ao pedido ou que o reenvia para outro servidor.
- É um protocolo orientado à sessão. Por cada utilizador que esteja a ser autenticado, existe uma sessão entre o cliente e o servidor. As diferentes sessões são distinguidas por um identificador, que naquele momento é globalmente único. Existe uma mensagem específica para terminar uma sessão DIAMETER. De modo a garantir que as sessões que de alguma forma

não são explicitamente terminadas são apagadas, existe um *time-out* de sessão. A dimensão do campo que especifica o tamanho dos pacotes RADIUS é 8 *bits*, enquanto que no DIAMETER esse campo tem 32 *bits*. O protocolo base do DIAMETER não inclui uma mensagem de autorização, uma vez que estas são fortemente dependentes das especificidades das aplicações e por isso, as mensagens de autorização são definidas em extensões ao protocolo. No entanto, o protocolo base define um conjunto de mensagens que são utilizadas para terminar sessões. O objectivo destas mensagens é permitir a libertação de recursos nos servidores, pois estes guardam informação de estado de cada sessão.

- O protocolo DIAMETER suporta a existência de *proxies*. Este suporte permite também o suporte de *roaming*, pois o servidor remoto actuará como *proxy* para o servidor de origem. Cada nó da rede é responsável pelas suas próprias retransmissões e o protocolo permite que um nó saiba de antemão o estado do nó com quem vai trocar mensagens DIAMETER. Desta forma é possível ter um esquema de retransmissões eficiente, reduzindo o número de nós DIAMETER e a latência associada na mudança de um nó para outro.
- O protocolo DIAMETER prevê o suporte de mensagens de redireccionamento. Um servidor de redireccionamento faz encaminhamento simples de mensagens DIAMETER: quando um servidor de redireccionamento recebe um pedido, emite uma resposta ao originador do pedido contendo a informação necessária para que o originador consiga contactar directamente o servidor no seu domínio de origem.
- O DIAMETER é baseado em SCTP (*Stream Control Transmission Protocol*). De modo a permitir descoberta rápida de falha de um nó, retransmissão agressiva e transacções rápidas, os nós DIAMETER têm de ser capazes de enviar e receber mensagens sobre SCTP. Contrariamente ao RADIUS, o protocolo DIAMETER exige que cada nó numa cadeia de *proxies* confirme ao nível da camada de transporte os pedidos ou respostas. Para tal, o DIAMETER pode utilizar o TCP uma vez que este assegura a transmissão fiável. O SCTP fornece a detecção de retransmissões (evitando a sua

duplicação a nível superior), o que simplifica de sobremaneira as implementações dos servidores.

3.3.2.3 Extensões

As extensões ao DIAMETER são desenvolvidas no âmbito de vários grupos do IETF. Actualmente existem os seguintes grupos responsáveis pela definição de extensões:

- ROAMOPS (*Roaming Operations*): Desenvolve procedimentos, mecanismos e protocolos que suportem a mobilidade de um utilizador entre vários fornecedores de serviço distintos.
- NASREQ (*Network Access Server Requirements*): Este grupo dedica-se ao alargamento das características dos NAS para suportarem mais do que apenas o serviço de *dial-up*, permitindo o suporte de VPNs, métodos de autenticação mais inteligentes e *roaming*.
- MobileIP (*IP Routing for Wireless/Mobile Hosts*): Este grupo dedica-se ao desenvolvimento de suporte de encaminhamento de modo a permitir que os nós IP possam mover-se de uma forma transparente entre várias redes e tecnologias de acesso distintas.
- AAA (*Authentication, Authorization, and Accounting*): Este grupo está a desenvolver extensões ao DIAMETER com vista a acrescentar contabilização, mecanismos de transporte, segurança e suporte de *proxies*.

As extensões DIAMETER actualmente já definidas são apresentadas de seguida.

3.3.2.3.1 Segurança

A base do protocolo DIAMETER permite que os servidores DIAMETER comuniquem de modo seguro, utilizando autenticação salto-a-salto. As extensões de segurança definem um conjunto de adições ao protocolo base que permitem autenticação, confidencialidade e não-repúdio ao nível do AVP. Através destas extensões é permitido assegurar segurança de uma forma selectiva a apenas algumas partes de uma mensagem DIAMETER. Assim, utilizando o protocolo DIAMETER, os *proxies* podem adicionar, apagar ou modificar as partes não protegidas dos AVPs.

3.3.2.3.2 IP Móvel

As extensões do protocolo DIAMETER para mobilidade permitem que um servidor autentique, autorize, e recolha informação de serviços que são fornecidos a nós móveis. Desta forma é permitido que os nós móveis possam usufruir de serviços de outros fornecedores. As extensões de contabilização são utilizadas para os agentes de mobilidade (locais e remotos) transferirem informação de utilização para os servidores DIAMETER. O protocolo RADIUS não dá suporte a este tipo de serviços.

3.3.2.3.3 Servidores de acesso

Esta extensão é utilizada para fornecimento de capacidades de AAA em ambientes de *Dial-UP* PPP/SLIP e *Terminal Server Access*. Nesta extensão são definidos um conjunto de comandos de autenticação e autorização que podem ser utilizados para CHAP (PPP *Challenge Handshake Authentication Protocol*), PAP (*PPP Password Authentication Protocol*) e EAP (PPP *Extensible Authentication Protocol*). A base do protocolo DIAMETER adicionada a estas extensões fornece as mesmas funcionalidades que o protocolo RADIUS.

3.3.2.3.4 Contabilização

A extensão de contabilização do DIAMETER foi projectada de modo a permitir que a informação de contabilização seja enviada através de vários domínios administrativos (opcionalmente, através de *brokers*). São suportadas transferências de dados em tempo real ou em datas e horas pré-programadas.

3.3.2.3.5 Gestão de recursos

Apesar das funcionalidades providenciadas pela base do protocolo e as extensões de contabilização, mobilidade IP e servidores de acesso que permitem que os nós mantenham informação de estado, poderá ser necessário que os nós DIAMETER necessitem de obter informação actual de todas as sessões activas de um terminal. Esta extensão (de gestão de recursos) define mensagens que são necessárias a um nó para pedir essa informação (que pode ser utilizada quando a informação de estado se perdeu, ou periodicamente de modo a assegurar que a informação de estado está actualizada).

3.4 Tecnologias de Mobilidade e Heterogeneidade

3.4.1 Mobilidade IP

Quando o IP foi definido, não foi considerada a possibilidade de os terminais mudarem a sua localização enquanto ligados a uma rede IP, i.e. assumiu-se que um terminal não se poderia mover de um local para outro, sem que tenha de se proceder à sua reconfiguração para se adequar à sua nova localização. Nessa altura, não se colocava ainda a possibilidade de haver terminais móveis que utilizassem IP.

Nas redes de telecomunicações tradicionais, o identificador do terminal contém informação com a qual é possível determinar a localização geográfica onde o terminal se encontra ligado, sendo acessível através do mesmo endereço independentemente da sua localização. No caso das redes IP, o identificador do terminal (endereço IP) contêm informação relativa à rede à qual se encontra ligado (o chamado prefixo de rede). Quando o terminal se move para uma outra rede, não existe uma forma intrínseca de fazer com que esse terminal continue a ter conectividade com as restantes redes IP, a não ser que o endereço IP seja alterado, ajustando-se ao novo prefixo IP existente na rede onde está. No entanto, a mudança de endereço também não é a solução perfeita, já que isso vai fazer com que o terminal não possa mais ser contactado por aqueles que sabiam da sua existência na rede original, já que ele não poderá responder pelo seu antigo endereço. Para todos os efeitos, ao mudar o endereço, o terminal muda de "identidade", desaparecendo a "identidade" antiga.

Hoje em dia, e devido às limitações no espaço de endereçamento IPv4, não é possível que cada terminal tenha um endereço global único e é bastante frequente a atribuição de endereços privados. O acesso destes terminais ao "mundo exterior" (i.e., *Internet*) é então feito, por exemplo, recorrendo a NAT/PAT (tipicamente dinâmicos). Com esta abordagem, não é possível que uma máquina na *Internet* consiga estabelecer uma ligação para estes terminais com base no seu endereço privado, ou seja, não são suportados serviços do tipo *network initiator* (iniciados na rede). O início de sessão terá sempre de partir do terminal que está a usar o endereçamento privado. Daqui se conclui que não são possíveis, na generalidade dos casos, ligações entre dois terminais pertencentes a redes distintas e ambos com endereçamento privado. Tendo em conta o crescimento de aplicações P2P (*Peer-to-Peer*), esta capacidade de estabelecer ligações de e para qualquer

terminal é cada vez mais uma obrigatoriedade. Este é também um factor importante que se coloca quando falamos na mobilidade dos terminais. Por um lado, é necessário reconfigurar o endereço do terminal para que, em termos de encaminhamento, o tráfego de e para o terminal possa ser encaminhado. Para que isto aconteça é necessário que o terminal possua um endereço com o prefixo da rede à qual está ligado. Por outro lado, é cada vez mais importante que o terminal possua um identificador universal e fixo, de forma a poder ser contactado a qualquer momento em qualquer lugar (tal como acontece com os terminais GSM, que mesmo em *roaming* estão acessíveis).

Ainda que, numa primeira abordagem, se possa pensar que as mudanças de rede não são tradicionalmente comuns, a realidade actual mostra que a mobilidade dos terminais começa a ser significativa e com tendência para um claro aumento. São cada vez mais as soluções de acesso à *Internet* via infra-estruturas *wireless*, tais como WLAN (*Wireless LAN*) empresariais, *Hotspots* públicos e redes GPRS e UMTS. Uma deslocação de um terminal ao longo de um percurso em que o ponto de acesso rádio (estação base) não é fixo, poder-se-á assemelhar na prática, e ao nível da camada L3, a uma mudança de rede (ainda que isto possa nem sempre ser assim, podendo recorrer-se apenas à mobilidade de nível 2). A Figura 6 exemplifica um possível cenário de mobilidade com redes de acesso heterogéneas.

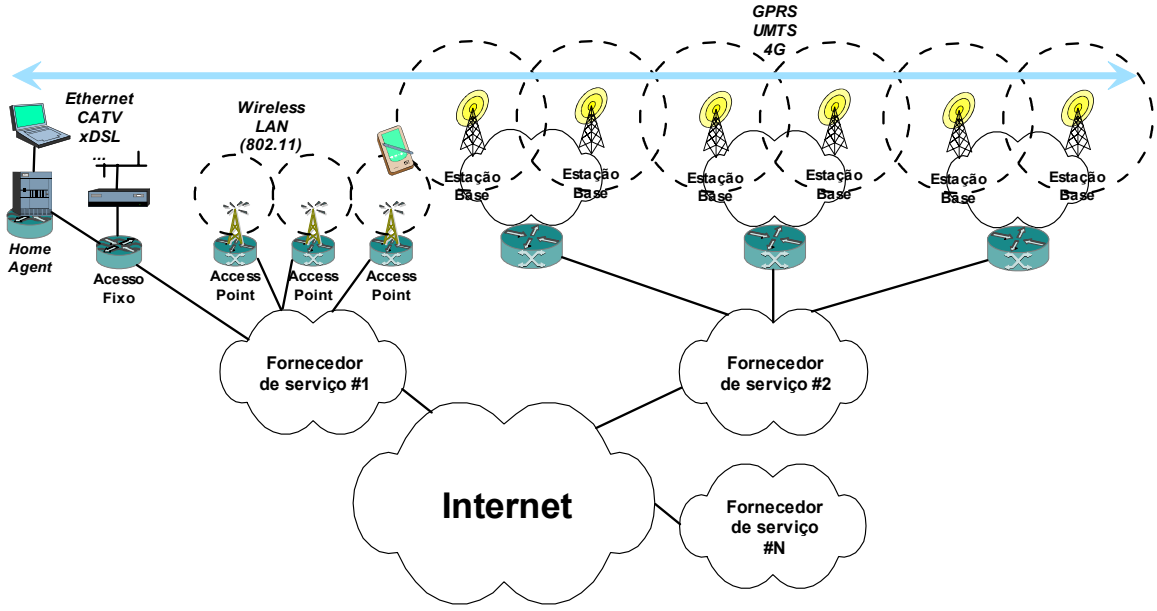


Figura 6: Cenário de Mobilidade IP entre redes heterogéneas.

Esta lacuna do IP foi identificada e estudada, tendo daí surgido um método que visa separar o endereço IP do terminal da sua localização, o chamado *mobile IP*. A mobilidade IP existe quer em IPv4 (MIPv4) [Perk02], quer em IPv6 (MIPv6) [Jonh03]. No entanto, como veremos na secção seguinte, a mobilidade IPv6 apresenta mais vantagens.

Com efeito, a Mobilidade IPv6 (MIPv6) é uma das mais importantes melhorias que a versão 6 do protocolo *Internet* traz, permitindo que os terminais se movam entre diferentes redes – mesmo entre diferentes tecnologias – mantendo a sua capacidade de contactar e ser contactado como habitualmente. E tudo isto, de uma forma transparente para as camadas superiores (TCP/UDP e aplicações) e, consequentemente, também para os utilizadores.

De uma forma simplista o funcionamento da MIPv6 baseia-se numa operação envolvendo vários endereços IPv6. Um deles, a que normalmente se chama *Home address* (Ha), é um endereço através do qual o terminal é bem conhecido, sendo utilizado por todos para o identificar univocamente a qualquer momento (este deverá constar nos servidores DNS). Este endereço têm normalmente um prefixo na rede onde o terminal se encontra habitualmente – a *Home Network* (HN); por exemplo, o prefixo da rede local num escritório ou o endereço atribuído pelo operador no âmbito de um contrato. Os outros endereços são construídos com base no prefixo de cada uma das redes (que não a sua *home network*) visitadas – chamadas *Foreign Networks* (FN). Este endereço, que se designa de *Care-of Address* (CoA), é o que possibilita que o tráfego chegue fisicamente até ao terminal no local onde ele encontra.

Resumindo, o terminal usa o Ha para estar sempre identificável, enquanto o CoA, adquirido em cada rede que visita, permite que o tráfego seja encaminhado fisicamente até ao terminal. Aqui, serão feitas as devidas alterações de forma a que a informação que é "passada" às camadas superiores (TCP, UDP, etc.), seja a equivalente àquela que receberia se o terminal se encontrasse na sua *home network* (e.g. escritório).

Mas para que a MIPv6 possa funcionar correctamente, é necessário que exista uma outra entidade neste processo - o *Home Agent* (HA). Este agente localiza-se na HN, e tem como função reencaminhar os pacotes destinados ao terminal móvel (normalmente designado por MN – *Mobile Node*), para a sua localização corrente, usando para isso o CoA. Para isso, o HA tem que saber onde é que o MN se encontra, i.e. tem que conhecer a cada momento a correspondência entre Ha e CoA. A esta correspondência chama-se *Binding*.

Assim, a primeira coisa que o MN tem de fazer quando muda de rede será informar o seu HA – esta mudança de rede denomina-se normalmente de *handover*. A partir daí, os pacotes destinados a ele ser-lhe-ão reencaminhados pelo HA e, numa fase imediatamente posterior, os pacotes passarão a ser trocados directamente entre o MN e o terminal que o contactou. A este terminal que estabelece uma comunicação com o MN chama-se *Correspondent Node* (CN), podendo este ser também móvel ou não.

3.4.1.1 MIPv4 vs. MIPv6

As diferenças entre as versões de mobilidade para os dois protocolos (IPv4 e IPv6) são bem evidentes e claramente favoráveis à MIPv6.

O primeiro problema do MIPv4 reside no endereçamento. O IPv4 foi vítima do seu próprio sucesso e a escassez de endereçamento começa logo por dificultar a implantação da mobilidade a nível comercial, já que não é possível dotar todos os utilizadores de um endereço bem conhecido e único. Em IPv6 este problema não se coloca.

Outra questão tem a ver com a necessidade de, em IPv4, existir uma entidade extra no processo de Mobilidade - o *Foreign Agent* (FA). Esta entidade é colocada nas redes visitadas com o objectivo de fazer com que os MNs adquiram um endereço IPv4, assim como informação adicional que lhe permita contactar o seu HA. Em IPv6 isso não é necessário já que através de auto-configuração os terminais adquirem um endereço, capaz de ser utilizado globalmente. A questão do FA torna-se um obstáculo bem real na hora de colocar MIPv4 nas redes dos ISPs, já que eles não controlam todo processo, necessitando a colaboração dos administradores de outras redes – as redes visitadas.

Um outro factor importante prende-se com o facto de que em IPv6 a mobilidade conta com o mecanismo simples e flexível dos cabeçalhos de extensão, o que não acontece com a MIPv4. Estes mecanismos são de grande ajuda não só para transportar as informações adicionais relativas à mobilidade, como também para implementar mecanismos de segurança que, ao contrário do IPv4, são suportados nativamente em IPv6 (e.g. IPSec).

No entanto, a maior e mais importante diferença entre o MIPv4 e o MIPv6 prende-se com o caminho percorrido pelo tráfego entre o MN e o CN. Também aqui os cabeçalhos de extensão do IPv6 são fundamentais. Enquanto que em IPv6, logo após o primeiro pacote, a comunicação é feita de forma directa entre os dois intervenientes, em IPv4 o tráfego com origem no CN e com destino ao MN, tem que passar obrigatoriamente pelo HA. No

entanto, no sentido oposto esta comunicação é feita de forma directa. A Figura 7 mostra esquematicamente essa situação, a que se dá o nome de *Triangle Routing*, devido ao triângulo formado pelos caminhos percorridos. Isto causa vários problemas:

- Ineficiência de percurso – o tráfego não utiliza o caminho mais curto entre a origem e o destino. Numa situação caricata o MN e o CN poderão estar na mesma rede, o HA do outro lado do mundo, e a comunicação deverá passar por este.
- Ineficiência de processamento – o HA e o MN têm de encapsular e descapsular os dados de modo a estes fluírem correctamente.
- Assimetria de percursos – este modo de funcionamento faz também com que os caminhos (ou *paths*) sejam assimétricos, o que nunca é boa prática quer para a detecção de falhas na rede, quer para o funcionamento de algumas aplicações/protocolos (como por exemplo a sincronização horária através do protocolo NTP – *Network Time Protocol*).
- Escalabilidade e tolerância a falhas – os HAs tendem a ficar sobrecarregados, convertendo-se também nos únicos pontos de passagem de grandes quantidades de tráfego, pelo que uma falha destes colocará muitos utilizadores fora de serviço.

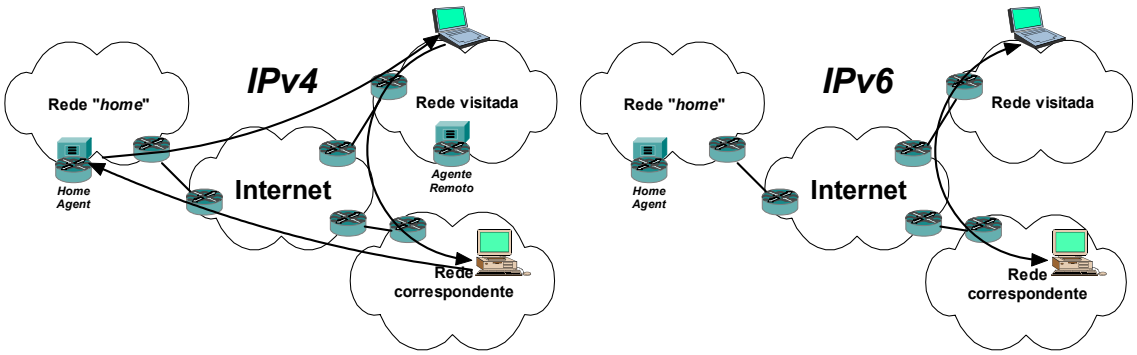


Figura 7: Comparação de caminhos em mobilidade IPv4 e IPv6

3.4.2 *Protocolos a nível da sessão: o SIP (Session Initiation Protocol)*

O SIP é um protocolo de sinalização utilizado para estabelecer sessões em redes IP. Uma sessão poderá ser uma conversação telefónica, ou qualquer sessão colaborativa multimédia. A capacidade de estabelecer este tipo de sessões abre caminho a um leque de serviços que podem variar desde o comércio electrónico enriquecido com voz, serviço

telefónico baseado em WEB, serviços de mensagens instantâneas com multimédia e serviços de localização, entre outros.

Os últimos três anos foram caracterizados pela adopção do SIP como protocolo de sinalização base para suporte de voz sobre IP. A indústria focou as suas atenções nesta norma que entretanto começou a emergir. Também dentro da comunidade 3GPP o SIP foi adoptado como protocolo para sinalização de sessões multimédia. O SIP está ainda em fase de amadurecimento e de extensão acompanhando a maturação tecnológica e integrando os novos produtos que são lançados no mercado.

O SIP segue um modelo tradicional do IETF que passa pela definição apenas do essencial em cada protocolo. Neste sentido, o SIP foi desenvolvido apenas como um mecanismo puro de início, fim e modificação de sessões, ignorando os detalhes da sessão. Desta forma, o SIP é escalável, extensível e pode ser facilmente aplicado em várias arquitecturas e cenários de evolução.

O SIP é um protocolo de pergunta-resposta, bastante semelhante na sua filosofia ao SMTP (*Simple Mail Transfer Protocol*) e ao HTTP (os protocolos base da WWW e dos correios electrónicos). Utilizando o SIP para suporte de telefonia, faz com que a telefonia IP seja vista apenas como mais uma aplicação WEB e que pode ser facilmente integrada com outros serviços *Internet*. Assim, o SIP é uma ferramenta simples que os fornecedores de serviço podem utilizar para convergirem os serviços de voz e multimédia.

3.4.2.1 A base da arquitectura SIP

Conforme foi referido atrás, a arquitectura base do SIP é cliente-servidor. As principais entidades definidas no SIP são o *User Agent*, o SIP *Proxy Server*, o SIP *Redirect Server* e o *Registrar*. A Figura 8 apresenta um cenário com a arquitectura básica SIP.

Os *User Agents*, ou terminadores SIP, são chamados clientes (UACs – *User Agent Clients*) quando enviam os pedidos de início de sessão, e são denominados servidores (UASs – *User Agent Servers*) quando desempenham a função de resposta aos pedidos. Os *User Agents* podem comunicar directamente com outros *User Agents*, ou podem fazê-lo através de um servidor intermediário. É também da responsabilidade dos *User Agents* manter e gerir o estado das sessões.

Os servidores intermediários têm a capacidade de se comportarem como *proxies* ou como *redirect servers*. Os servidores de *proxy* têm como responsabilidade o reenvio dos

pedidos dos *User Agents* para o servidor ou *User Agent* seguinte e a manutenção de informação de contabilização e de tarifação.

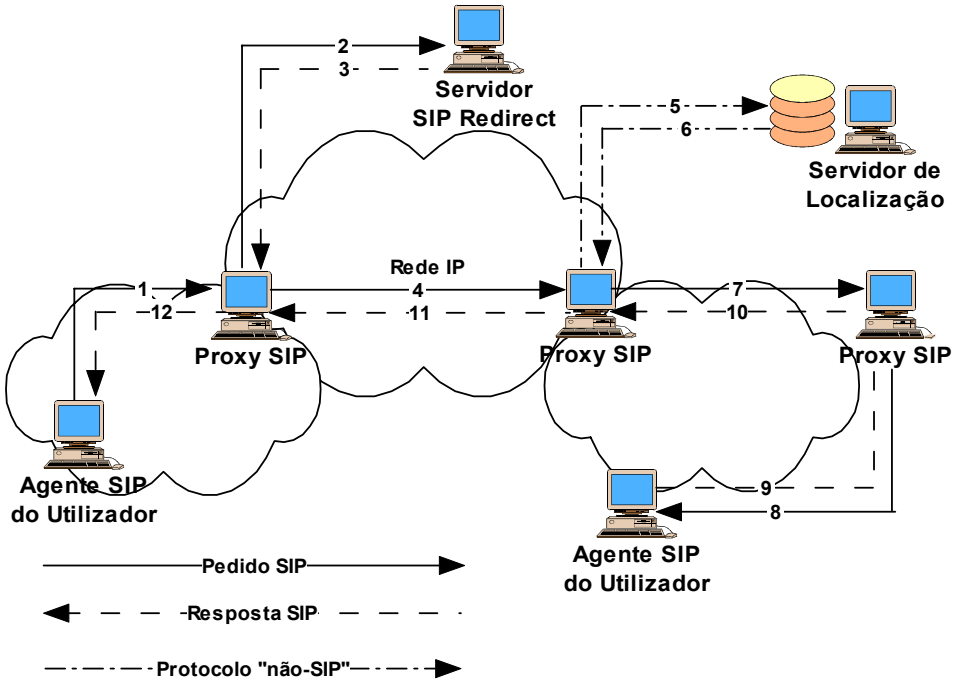


Figura 8: Arquitectura básica SIP

Os SIP *Redirect Servers* respondem a pedidos dos clientes e informam-nos acerca do endereço do servidor procurado. Até que um pedido/resposta chegue ao seu destino final pode passar por um número elevado de nós. A flexibilidade do SIP permite ainda que os servidores possam contactar servidores de localização externos (por exemplo um *home Agent* IPv6) de modo a determinar a localização de determinado utilizador ou a política de encaminhamento. Desta forma, poderão coexistir diversos mecanismos de localização de utilizadores. De modo a manter a escalabilidade, os servidores SIP podem manter informação de estado, ou podem reenviar os pedidos de uma forma directa, sem manutenção de estado.

A terceira entidade que compõe a arquitectura SIP é o SIP *Registar*. O *User Agent* envia uma mensagem de registo para o SIP *Registar* e este armazena a informação de registo num servidor de localização externo, através de um protocolo "não-SIP". Após a informação guardada, o *Registar* envia a resposta apropriada de volta ao *User Agent*.

O SIP permite desta forma que a mobilidade seja tratada a um nível superior ao IP e, de certa forma, concorrente. No entanto, a utilização do SIP com mobilidade IP poderá ser completamente transparente.

3.5 O DMIF: Tecnologia de Interligação Multimédia

Até há algum tempo, um dos factores mais limitativos no desenvolvimento de novas aplicações e serviços, em especial aqueles que lidam com multimédia, era o facto de a evolução das redes e protocolos passar por processos conturbados, sem se saber exactamente como terminariam ou qual seria a próxima etapa da evolução. Desta forma, havia receio de se fazerem investimentos no desenvolvimento de aplicações que utilizassem determinada tecnologia, e que após o seu desenvolvimento essa tecnologia estar ultrapassada e todo o investimento realizado fosse desperdiçado. O desenvolvimento das aplicações era condicionado pela infra-estrutura de suporte. Este problema existia porque a camada aplicacional não se dissociava nem da camada de transporte nem da camada de rede e por vezes, nem mesmo da camada de ligação. Para colmatar este tipo de problemas e permitir que as aplicações fossem desenvolvidas de uma forma independente das camadas inferiores, surgiu um grupo de trabalho dedicado ao desenvolvimento de um *middleware* adequado, o DMIF (*Delivery Multimedia Integration Framework*). O DMIF é na sua essência uma camada de abstracção e adaptação entre a camada da aplicação e as inferiores. Embora definido como parte integrante da norma MPEG-4, no âmbito da sua "parte 6", o DMIF é independente deste e a sua utilização não está restrita ao MPEG-4.

O DMIF está dividido em dois planos distintos (Figura 9): o plano do utilizador e o plano de controlo. Estes dois planos são separados e encapsulados pelas interfaces DAI (*DMIF Application Interface*) e DNI (*DMIF Network Interface*). Todas as funções situadas acima da DAI são consideradas pertencentes ao plano do utilizador. O utilizador DMIF controla a abertura e fecho de serviços e canais. O plano de controlo está situado entre a DAI e a DNI. Desta forma, o DMIF permite que a aplicação não tenha conhecimento do cenário operacional onde se encontra (isto é, se a comunicação é local ou remota, ou mesmo se se trata de um cenário *broadcast/multicast* ou *unicast*) assim como das tecnologias de transporte e protocolos específicos. Isto é obtido através da definição precisa da DAI, que mantém as funções da aplicação e do transporte separadas. O DMIF define um mecanismo extensível para a passagem de parâmetros e métricas de QoS através da DAI. Este mecanismo permite a passagem de parâmetros distintos para cada fluxo ou conjunto de fluxos, designados por "fluxos elementares" em termos da norma.

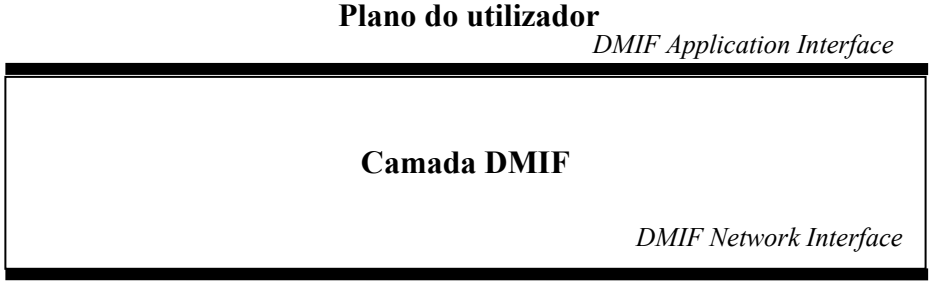


Figura 9: Modelo de camadas DMIF

As características que fazem o DMIF tão atractivo são a capacidade de agregação de fluxos em sessões de serviço, o mecanismo de entrega fiável mas transparente, a QoS ajustável e a interface bem conhecida e consistente. O DMIF especifica que a QoS ao nível do transporte deve obedecer a determinadas regras. Deve ser mantida a todo o custo a QoS ao nível especificado e, em caso de falha neste cumprimento, a aplicação deverá ser notificada.

3.5.1 Flexibilidade DMIF

A abordagem tomada pelo DMIF é bastante flexível e permite que seja explorada de várias formas. Esta flexibilidade pode ser explicada pelos motivos seguintes:

- Multiplexagem de fluxos elementares: cada fluxo elementar pode ser transmitido separadamente ou poderá ser agregado com outros numa única ligação. De modo análogo, pode ser feita qualquer combinação de fluxos de modo a permitir a agregação de fluxos baseada em algum critério específico, como por exemplo, características de QoS semelhantes.
- Escolha do método de transporte: por exemplo, no caso de transmissão sobre IP, pode-se escolher entre os protocolos TCP (*Transmission Control Protocol*) e UDP (*User Datagram Protocol*) e possivelmente RTP/RTCP (*Real-Time Control Protocol / Real-time Transport Protocol*). Pode ainda escolher-se a tecnologia de transporte, caso existam várias disponíveis (por exemplo, IP e ATM).
- Mapeamento de QoS: existência e possibilidade de definição de critérios de mapeamento da QoS de cada fluxo elementar em recursos de rede, de acordo com a tecnologia de transporte (IP com RSVP, DiffServ, ATM, etc.).

Cabe à implementação DMIF decidir quando deve agregar diferentes fluxos ou transmiti-los em separado, qual a camada de transporte que deve utilizar, e que QoS deve pedir e exigir à rede, baseada quer na QoS exigida pelos diferentes fluxos quer no conhecimento da tecnologia de transporte. É permitido que instâncias DMIF distintas sejam escolhidas de uma forma dinâmica.

O DMIF tem a capacidade de criar sessões fiáveis com um servidor remoto, que desconhece o método de transporte e entrega da informação. Isto significa que, quer a aplicação quer as camadas de *streaming* e sincronização, não têm conhecimento da arquitectura de rede de suporte. Desta forma, é possível que as redes futuras ou protocolos ainda não integrados o possam ser, sem necessidade de qualquer modificação ao nível da interface para as aplicações e das próprias aplicações.

CAPÍTULO 4

QUALIDADE DE SERVIÇO EM REDES IP

4.1 Introdução

O crescimento da utilização das redes IP – *Internet*, *intranet* e *extranets* – tem feito com que cada vez mais serviços sejam suportados sobre esta infra-estrutura. Aplicações como VoIP (*Voice over IP*) [VoIP04], de *streaming* de áudio e vídeo, ou videoconferências, são alguns exemplos. Estas aplicações não têm o comportamento das aplicações mais tradicionais em redes IP, apresentando requisitos de tempo real. Estas novas aplicações não se adaptam aos recursos existentes na rede (aplicações não adaptativas) e por isso necessitam de garantias de serviço. É necessário que existiam mecanismos que permitam controlar a qualidade de serviço (QoS) extremo-a-extremo disponibilizado entre as aplicações. Na prática é a rede (ou o operador) que terá de disponibilizar essa funcionalidade.

A *Internet*, que foi originariamente concebida para oferecer o nível mais simples de QoS, o *Best Effort* (o melhor esforço), não é vista como um meio fiável de suporte de

aplicações de tempo real. De facto, a rede faz o melhor que pode para conseguir uma entrega fiável do tráfego, mas sem efectuar qualquer diferenciação entre fluxos. Tradicionalmente, o tratamento diferenciado é entendido como a atribuição de largura de banda necessária ao funcionamento correcto das aplicações. No entanto, há outros parâmetros que importa controlar pois também são importantes para as aplicações (em especial para aquelas que têm requisitos de tempo real) tais como: o atraso, a sua variação, ou a taxa de perdas.

Com o futuro que se avizinha, em que o protocolo IP deverá ter a função de integrador de todos os serviços é extremamente importante que a diferenciação de QoS seja suportada, de forma a manter a qualidade actual desses serviços (por exemplo TV ou voz).

Este capítulo começa por apresentar um modelo de QoS estruturado em camadas, focando-se nas camadas de aplicação e de rede. O resto do capítulo centra-se na provisão de QoS ao nível da rede, sendo primeiro apresentados os mecanismos normalizados para provisão de QoS em redes IP. Após isto, apresentam-se os modelos de gestão e controlo de QoS, tendo como base os mecanismos atrás referidos. O capítulo termina com a exposição das estratégias de QoS extremo-a-extremo em ambientes de operadores.

4.2 Definição de QoS: visão estruturada em camadas

A implementação de um esquema de QoS levanta múltiplas questões, nem sempre de resolução simples. Por exemplo, existem estratégias distintas dependentes de onde a diferenciação da QoS ao nível da rede é aplicada: apenas na fronteira da rede, no *core* ou extremo-a-extremo. Estas diferentes estratégias têm diversas implicações. Se a QoS for implementada apenas na fronteira diminui os custos no *core*, mas todo o tráfego será tratado de igual forma no core. Se a QoS for implementada apenas no *core* os custos na fronteira diminuem mas, o tráfego na fronteira não será prioritizado correctamente e as garantias que podem ser oferecidas são baixas. A única implementação eficaz, capaz de realmente garantir QoS, será extremo-a-extremo, incorporando aspectos de QoS tanto na fronteira como no *core*. Para que a implementação extremo-a-extremo seja eficaz, é aconselhável a utilização de um modelo estruturado em camadas, sendo conveniente averiguar e assegurar os parâmetros relevantes da QoS em cada uma das camadas. Este modelo de camadas de QoS encontra-se esquematizado na Figura 10. Este modelo é

particularmente útil no caso de serviços multimédia, compostos por meios (áudio, vídeo, dados) distintos, que necessitam QoS para cada um deles e também para o agregado (por exemplo, é importante o sincronismo entre o áudio e o vídeo).

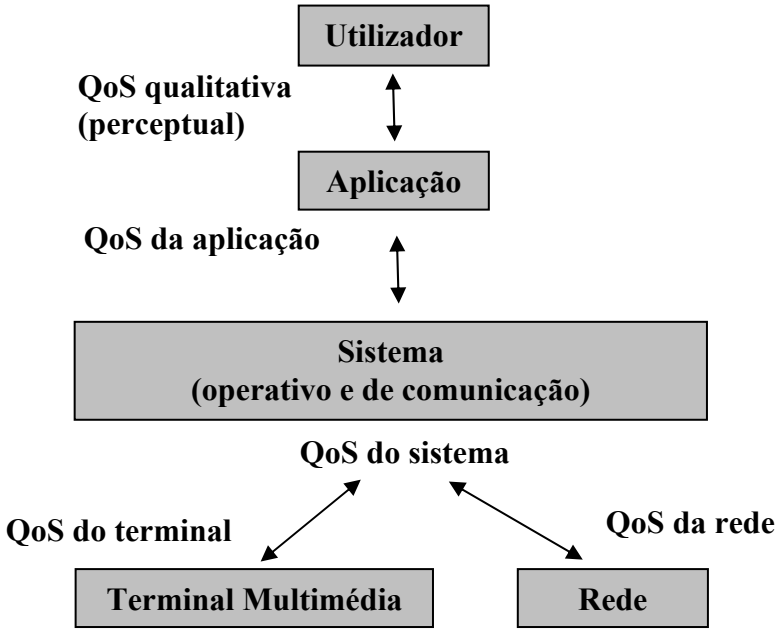


Figura 10: Modelo de camadas de QoS

Antes de um serviço ser iniciado, deve haver a admissão e negociação do serviço em cada camada, assim como entre camadas (serviço de tradução, tal como vimos no caso do DMIF). Só desta forma, se garante a condição de disponibilidade de recursos necessária para fornecer a QoS desejada. Durante a sessão devem estar activas (de preferência em cada camada) técnicas de policiamento, controlo de débito e de erros. A monitoria da QoS e o comportamento adaptativo dos recursos é também de grande relevância.

Nesta secção serão abordadas apenas as camadas relativas à aplicação e rede, o que não significa que as restantes camadas sejam menos importantes. Por exemplo, a camada de sistema permite a adaptação de redes e terminais heterogéneos (inclui especificação de *buffers* e tarefas), e a camada relativa aos terminais possui características tais como a reformatação de dados (por exemplo de acordo com o tamanho de um ecrã). A operação conjunta de diversos mecanismos nas várias camadas permite a convergência em termos de eficiência para a obtenção de QoS extremo-a-extremo. No entanto, estes aspectos têm menor importância para os objectivos deste trabalho.

4.2.1 QoS ao Nível da Aplicação

A QoS de aplicação de um fluxo multimédia consiste nas descrições das qualidades de um meio individual dentro do fluxo (áudio, vídeo ou dados), na maneira como os meios são combinados no fluxo multimédia (Qualidade do meio) e na descrição do "Interfuncionamento com o meio", de acordo com a Figura 11 [Tiph00].

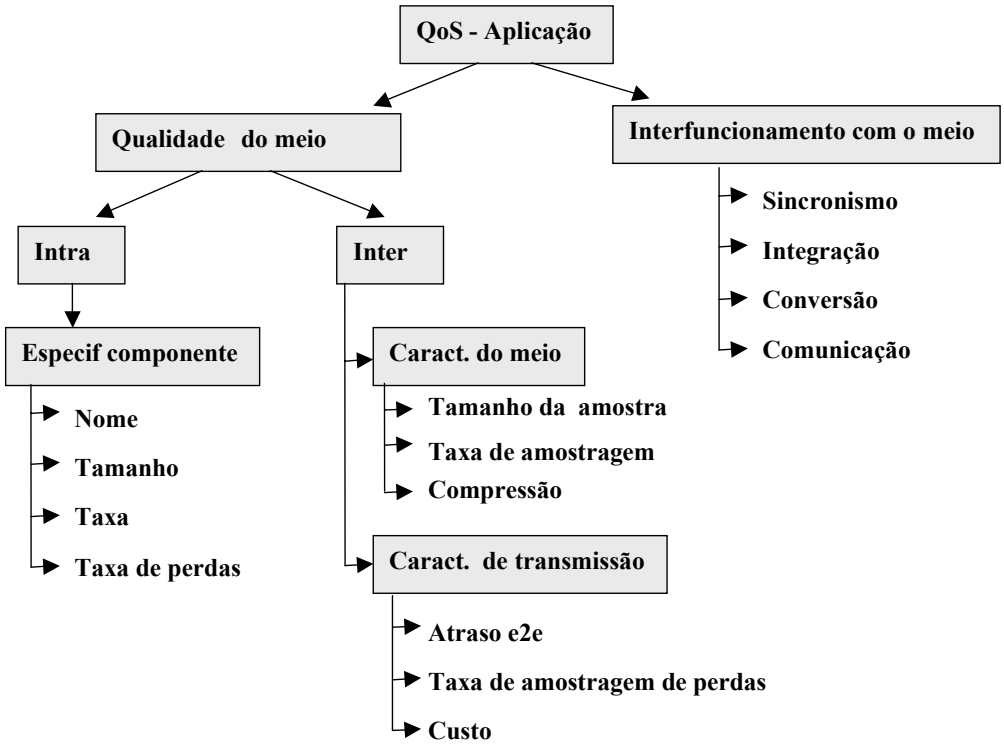


Figura 11: Parâmetros de QoS da aplicação

A qualidade do meio consiste neste caso na especificação de aspectos inter-trama e intra-trama. A especificação inter-trama apresenta as características de um fluxo num meio homogéneo, como o tamanho da amostra e a taxa de amostragem. A parametrização inclui também especificação, orientada à aplicação, dos requisitos das características de transmissão para entrega extremo-a-extremo. Se as amostras individuais no fluxo diferirem na qualidade (várias qualidades para o áudio) então deve haver uma especificação intra-trama, significando que cada sub-amostra é definida pelo utilizador/aplicação utilizando a especificação de componente.

O interfuncionamento com o meio especifica relações entre os vários fluxos componentes de uma sessão multimédia (note-se que podem existir diversos fluxos do mesmo tipo de meio, e.g., duas fontes de áudio distintas na mesma sessão). O desvio de

sincronismo representa um limite superior no *offset* de tempo entre dois fluxos numa única direcção. O parâmetro "comunicação" define a topologia de comunicação como *unicast* (um para um), *multicast* (um para vários) ou *broadcast* (um para todos). O parâmetro "conversão" especifica transformações de meio (conversão de áudio para texto em aplicações de reconhecimento de voz, por exemplo).

A comunicação multimédia utiliza assim um conjunto diverso de serviços elementares, tornando-se necessária a parametrização destes serviços de acordo com a qualidade de serviço pretendida. A tabela seguinte apresenta alguns exemplos de parâmetros de QoS ao nível da aplicação para áudio e vídeo.

Tipo de meio	**Parâmetro QoS**	**Gama**	**Caracterização da Qualidade**
Áudio	Tamanho da amostra	8 bit	Qualidade de voz PSTN
	Taxa de amostragem	8KHz	(atraso intermédio 125 □s)
	Tamanho da amostra	16 bit	Áudio - CD
	Taxa de amostragem	44.1 KHz	(atraso intermédio 22,7 □s)
Vídeo	Taxa de tramas	30 tps	Formato NTSC
		25 tps	Formato PAL
		60 tps	Formato HDTV
	Largura da trama	≤ 720 pixels	Sinal de vídeo com codificação MPEG
	Altura da trama	≤ 576 pixels	Tamanho vertical
	Resolução da cor	8 bit/pixel	Resolução escala cinzenta com 256 cores
		16 bit/pixel	65536 cores possíveis
	Tamanho da trama	Largura da trama * Altura* cor)	
	Relação - Aspecto	4:3	Formato de TV NTSC, PAL
		16:9	Formato HDTV
	Relação - compressão	2:1	Compressão HDTV sem perdas
		50:1	Compressão HDTV com perdas

Tabela 2: Exemplos de parâmetros QoS de áudio e vídeo

Dada a cada vez maior importância do serviço de voz sobre IP, apresentam-se como referência na Tabela 3 as características das normas de compressão de áudio mais utilizadas hoje em dia.

Norma	Tipo de codificação	Débito (Kbps)
G.711	PCM	64
G.729	CS-ACELP	8
G.723.1	ACELP MP-MLQ	6.3 5.3

Tabela 3: Comparação de normas de compressão de voz

No que respeita à interacção e interfuncionamento de meios distintos, a Tabela 4 apresenta exemplos dos parâmetros de QoS mais relevantes.

Tipo de meios	Parâmetro QoS	Gama	Caracterização da Qualidade
Áudio/vídeo	Desvio de sincronismo	+/- 80 ms	Sincronização dos lábios
Áudio/imagem	Desvio de sincronismo	+/- 5ms	Música com anotações
Áudio/ponteiro	Desvio de sincronismo	+ 750 ms	(+) áudio à frente do ponteiro
		- 500 ms	(-) áudio atrás do ponteiro

Tabela 4: Exemplos de parâmetros QoS com interfuncionamento de meios

Estes valores podem servir como indicações gerias quanto ao comportamento desejado por parte das camadas de rede.

As aplicações podem dispor de algoritmos que permitam comportamentos mais resistentes a alguns dos parâmetros acima mencionados. Um destes algoritmos é o de *error concealment*, que permite uma camuflagem das perdas ocorridas no sinal com base em técnicas de preservação das características espectrais do sinal emissor; outros algoritmos utilizam técnicas CELP (*Code Excited Linear Prediction*) para determinar o conteúdo de um pacote que falta através da observação do pacote anterior. Uma outra técnica para minorar o *jitter* é a existência de "*jitter buffers*", preferencialmente adaptativos, que permitem retirar as variações elevadas nos tempos de chegada de pacotes, ajustando o atraso do *jitter buffer* à taxa de perdas detectada.

Embora a utilização deste tipo de técnicas seja muito variada, a Tabela 5 presenta, em termos genéricos, as principais características de algumas aplicações.

A implementação de QoS ao nível da aplicação passa pois pela caracterização dos componentes que compõem a aplicação, e pela configuração apropriada dos parâmetros de QoS por forma a que não haja degradação da qualidade. Naturalmente que será necessário um conhecimento das condições existentes na rede ou uma previsão do tráfego para se

decidir melhor sobre os compromissos a tomar na escolha dos vários algoritmos possíveis para cada um dos fluxos.

Aplicação	**Largura de banda**	**Sensibilidade a:**		
		Atraso	***Jitter***	**Perdas**
VoIP	Baixa	Elevada	Elevada	Média
Videoconferência	Elevada	Elevada	Elevada	Média
Streaming de vídeo a pedido	Elevada	Média	Média	Média
Streaming de áudio	Baixa	Média	Média	Média
Transações Cliente/Servidor	Média	Média	Baixa	Elevada
E-mail	Baixa	Baixa	Baixa	Elevada
Transferência de ficheiros	Média	Baixa	Baixa	Elevada

Tabela 5: Principais características de algumas aplicações

4.2.2 QoS ao Nível da Rede

Na camada de rede, a definição de QoS é o resultado da conjugação de três tipos de análise e especificações: a taxa de transmissão (*throughput*), o tráfego e o desempenho. Esta divisão é ilustrada na Figura 12. A estrutura representada é respeitante à QoS sobre uma ligação de rede única para um dos serviços elementares (do ponto de vista de uma aplicação multimédia). Normalmente, a caracterização da QoS na rede é realizada tendo em consideração apenas o subconjunto de especificação de tráfego. No entanto, é cada vez mais comum (e será "obrigatório" no futuro) a caracterização da QoS na rede ter em conta também a taxa de transmissão (pertencente à especificação de *throughput*) e a prioridade (pertencente à especificação de desempenho).

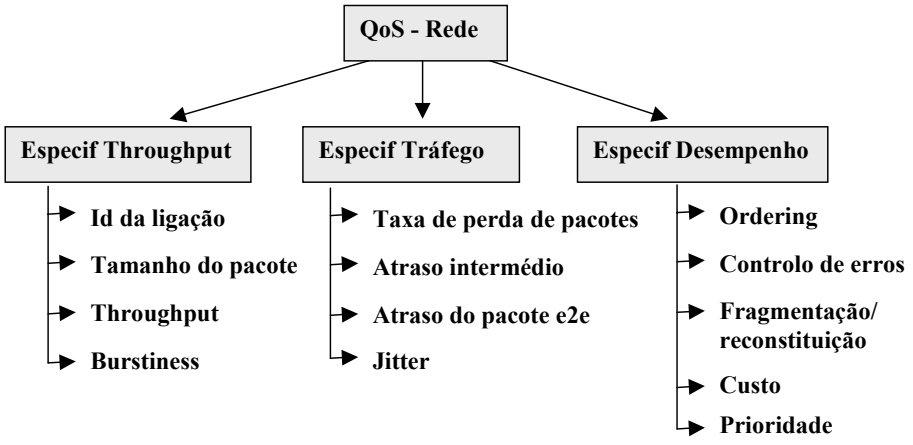


Figura 12: Parâmetros de QoS de rede

Nas redes de voz tradicionais, baseadas em TDM, o tráfego de voz sofre um atraso fixo pois tem associado um circuito dedicado e não sofre perdas devido à estrutura de comutação de circuitos da rede telefónica, resultando numa elevada qualidade de voz. Esta não é no entanto a realidade das redes IP em que o tráfego pode sofrer atrasos, variações de atraso e perdas em determinados instantes. Estes efeitos podem provocar degradações dos serviços. A voz é particularmente sensível, apresentando distorção, palavras cortadas (sinónimo de perda de pacotes) e ecos.

Os três factores que mais influenciam o desempenho e qualidade dos serviços em geral, e os do tipo VoIP ou videoconferência em particular, são: atraso, variações de atraso e perda de pacotes:

- Atrasos – Os serviços de voz e vídeo consistem em comunicações em tempo real, bidireccionais, em que o atraso extremo-a-extremo pode causar grandes constrangimentos. Atrasos inferiores a 150 *ms* são considerados bons em termos de voz. Considera-se por exemplo que o serviço de voz para atrasos superiores a 400 *ms* não é de todo aceitável para uma comunicação bidireccional. Em aplicações de transferência de dados, o atraso não introduz efeitos nocivos significativos.
- Variações de atraso – Numa rede IP nem todos os pacotes sofrem o mesmo atraso. As variações do atraso (*jitter*) dos pacotes resultam na chegada de pacotes ao destino sem um padrão constante, provocando uma degradação da voz. Tipicamente a solução para problemas de variação de atraso consiste no aumento do "*jitter buffer*" (filas de espera para acomodação de atrasos), nos equipamentos terminais. Deste modo, os pacotes são entregues à camada de aplicação a um ritmo aproximadamente constante. No entanto, a solução do "*jitter buffer*" tem o efeito nocivo de aumentar o atraso médio total, devendo ser consideradas as características do atraso na rede para um dimensionamento correcto das dimensões das filas de espera. À semelhança do atraso, para ligações de transferência de dados, a variação do atraso não introduz efeitos nocivos significativos.
- Perda de pacotes – O facto de existirem perdas de pacotes nas redes tem efeitos distintos no tipo de aplicações. Devido ao tamanho pequeno dos pacotes de voz, a perda de pacotes ocasional produz um impacto

insignificante. No entanto quando as perdas aumentam e/ou ocorrem em pacotes consecutivos, a qualidade de voz degrada-se. Em relação a serviços de transmissão de vídeo a perda de pacotes é um factor que também deve ser levado em consideração, o vídeo é dependente também do tipo de perdas (em rajada ou ocasionais) e, por exemplo, do tipo de codificação. Vários estudos foram efectuados e concluiu-se que, em termos de percepção humana, a degradação do vídeo é melhor suportada do que a degradação do áudio. No que respeita a ligações de transferência de dados, estas normalmente são executadas sobre TCP (*Transmission Control Protocol*), que força a retransmissões aquando da ocorrência de perdas de pacotes. No entanto, as perdas provocam uma redução da taxa de transmissão efectiva, o que deve ser levado em conta, em especial se levar à violação dos SLAs existentes.

Existem assim limites distintos para cada um destes parâmetros de avaliação de qualidade, de acordo com o serviço a prestar. Torna-se necessária a implementação de mecanismos de gestão e controlo de QoS nas redes IP por forma a garantir que o tráfego possua uma boa qualidade de modo a atingir as expectativas dos utilizadores/clientes.

4.3 Mecanismos para oferta de QoS ao nível da rede IP

O mecanismo mais simples de atribuição de QoS é o sobre-dimensionamento das redes de maneira a que exista bastante largura de banda disponível para todos os serviços. No entanto, em situações de congestionamento ou picos de tráfego, os serviços sensíveis a atrasos e perda de pacotes serão afectados de forma negativa.

De modo a possibilitar a existência de uma verdadeira QoS IP extremo-a-extremo, o IETF (*Internet Engineering Task Force*) definiu dois modelos distintos: os Serviços Integrados (*Integrated Services* – IntServ) e os Serviços Diferenciados (*Differentiated Services* – DiffServ). Estes dois modelos são duas formas de resolver o problema fundamental de fornecer QoS aos pacotes IP. O IntServ segue um modelo de QoS baseado em sinalização explícita, que os terminais utilizam para indicar à rede a qualidade desejada. O modelo DiffServ opera de um modo em que os recursos de várias classes de serviço (com parametrizações específicas para cada tipo de tráfego e utilização) são previamente reservados. Existem outros mecanismos que permitem a implementação de

QoS em redes IP, nomeadamente a utilização de precedências IP (o primeiro e mais simples mecanismo de QoS IP a ser definido e implementado), a classificação nos *routers* (por exemplo dando prioridades distintas a pacotes com portos distintos) e a utilização de MPLS (*MultiProtocol Label Switching*). Nesta secção o modelo IntServ será apresentado em primeiro lugar, descrevendo-se de seguida o esquema de precedência IP e mostrando-se como foram utilizados os conceitos deste esquema para a definição do modelo DiffServ. A secção termina com a apresentação do MPLS e de como este é actualmente aplicado para o fornecimento de QoS em redes IP.

Nesta tese é apresentada uma estratégia de QoS extremo-a-extremo, partindo de alguns destes conceitos do IETF. Pretende-se oferecer uma abordagem a este problema, num ambiente heterogéneo, englobando a interacção da QoS com a mobilidade e o controlo de autenticação e autorização, como veremos no Capítulo 5.

4.3.1 IntServ

O IntServ apresenta uma solução QoS extremo-a-extremo bastante rica, através da sinalização extremo-a-extremo, da manutenção de estado (por cada fluxo e reserva), e controlo de admissão realizado em cada um dos elementos de rede ao longo do percurso. O modelo IntServ utiliza geralmente o protocolo RSVP (*Resource Reservation Protocol*) [Brad97][Zhan93] de modo a sinalizar e reservar a QoS de cada fluxo na rede.

Um fluxo é definido como uma única corrente de informação, unidireccional, entre duas aplicações e é univocamente identificado pelo quinteto composto pelos endereços IP de origem e destino, os portos de origem e de destino e pelo protocolo de transmissão; alguns destes elementos podem não ser utilizados, em certos casos.

Em IntServ, um terminal pode usar dois tipos de serviço distintos. O primeiro, denominado de "Serviço Garantido" (*guaranteed service*) [Shen97] consiste num serviço com garantias estritas de atraso e largura de banda extremo-a-extremo, conforme as especificações da reserva efectuada. O segundo tipo de serviço é denominado de Serviço de Carga Controlada (*controlled load service*) [Wroc97] e oferece um tipo de serviço melhor que o tradicional Melhor Esforço (*Best Effort*) em termos de largura de banda e atraso. Em caso de congestionamento na rede, este serviço permite que os pacotes a ele pertencentes, tenham um tratamento semelhante ao obtido pelos pacotes "*best effort*" no caso da rede estar com uma carga baixa ou moderada. Quando uma aplicação necessita de determinados recursos da rede (e os consegue quantificar), constrói uma mensagem

especificando esses valores, e envia-os através da rede de forma a sinalizar não só o receptor, como também todos os nós de rede no caminho. Assim, todos os nós no caminho poderão rejeitar ou aceitar o novo fluxo e reservar os recursos especificados, dependendo da disponibilidade dos recursos existentes.

Desta forma é possível (pelo menos teoricamente) fornecer a QoS necessária a cada fluxo, desde que devidamente sinalizado com RSVP e dependente da disponibilidade de recursos.

No entanto existem várias desvantagens de ordem prática nesta abordagem:

- Cada dispositivo ao longo do percurso entre os extremos (incluindo os próprios terminais) da comunicação deve ter a capacidade de processar a sinalização RSVP e sinalizar a QoS pretendida.
- As reservas não são fixas (*soft reservations*), ou seja, é necessário que sejam refrescadas periodicamente, adicionando tráfego à rede e aumentando a probabilidade da reserva não ser mantida caso se percam os pacotes de refrescamento. Apesar de existirem mecanismos criados com o intuito de resolver este problema, eles trazem complexidade adicional ao RSVP.
- A manutenção da informação de cada fluxo nos *routers*, combinada com a necessidade de fazer controlo de admissão em cada nó da rede, aumenta a complexidade de cada um dos nós de rede, assim como a necessidade de aumentar a sua capacidade de processamento e de memória para manter informação acerca de cada um dos fluxos.
- Uma vez que é necessário que todos os nós no percurso, para cada uma das reservas, mantenham a informação de estado, apliquem policiamento individualmente de forma a não aceitar trafego para além do estabelecido e procedam ao escalonamento de pacotes por cada fluxo, havendo centenas ou milhares de fluxos activos, especialmente nos nós do *core* da rede, a solução não é escalável.

4.3.2 Precedência IP

Devido à dificuldade de conseguir obter QoS por fluxo, extremo-a-extremo, sem adicionar uma elevada complexidade, custo e problemas de escalabilidade, surgiu a ideia de classificar os fluxos em agregados (classes) e garantir os níveis de QoS aos agregados. Todo o tráfego relativo a ligações TCP pode ser agrupado numa única classe, sendo a

largura de banda atribuída à classe e não a cada fluxo. Desta forma reduz-se bastante a complexidade dos sistemas e diminui-se a quantidade de sinalização trocada entre os elementos. O IETF avançou para uma solução que, utilizando o campo TOS (*Type of Service*) dos pacotes IPv4, implementasse um sistema de prioridades e precedências. Um pacote marcado com a prioridade ou precedência IP apropriada é tratado ao longo do seu percurso de acordo com essa marcação. Pacotes mais prioritários, são encaminhados antes dos menos prioritários.

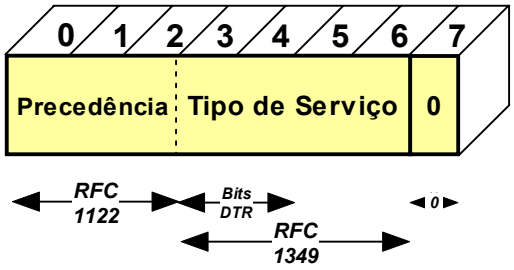


Figura 13: Campo TOS

O campo TOS dos pacotes IPv4 é apresentado na Figura 13. O três bits de precedência são utilizados nos nós fronteira de modo a classificar os pacotes em uma de 8 categorias possíveis. Os pacotes que têm menor precedência (valor mais baixo) poderão ser descartados em favor dos mais prioritários, se ocorrer uma situação de congestionamento na rede. Além da precedência, os pacotes podem também ser marcados de modo a obterem tratamentos de encaminhamento distintos em relação a atraso, taxa de transmissão e fiabilidade (os bits DTR – *Delay*, *Throughput* e *Reliability*) (de acordo com o RFC 791 - IP). No entanto, [Almq92] redefine estes três bits e utiliza ainda o sétimo bit do TOS para designar o tipo de serviço do pacote, para além da prioridade. À primeira vista pode-se pensar que este esquema tem tudo aquilo que é necessário para implementar QoS IP numa rede. No entanto, existem algumas limitações e falhas cruciais que o impedem:

- O esquema de precedências IP permite apenas especificar a prioridade relativa de um pacote, não separando prioridade de probabilidade de descarte. Não foi definido um mecanismo que permita especificar diferentes precedências de descarte para pacotes com uma determinada prioridade. Por exemplo, pacotes de aplicações distintas podem ser marcados com a mesma prioridade, mas em caso de congestionamento, o administrador poderá querer

descartar uns primeiro que os outros, pois estes podem ser mais "importantes".

- Os três bits disponíveis permitem apenas definir oito classes de tráfego distintas. Existem duas classes (Controlo de Rede e Controlo de Interfuncionamento de Redes) que estão geralmente reservadas para pacotes gerados nos elementos de rede (por exemplo, para pacotes dos protocolos de encaminhamento). O objectivo é proteger os pacotes considerados vitais para o bom funcionamento da rede, mas desta forma restringe-se a seis o número de classes disponíveis para classificar o restante tráfego.
- Nem o esquema de precedência IP, nem a utilização dos bits DTR estão implementados de forma consistente pelos vários fabricantes. Adicionalmente, [Almq92] redefine o campo TOS, eliminando o conceito de DTR.

As limitações expostas reduzem a possibilidade de implementar com sucesso QoS extremo-a-extremo, utilizando este esquema. No entanto, este esquema tem sido frequentemente utilizado para gestão interna de redes em operação.

4.3.3 DiffServ

A primeira definição do DiffServ surgiu no âmbito do IETF no final de 1998 [Carl98]. O grande objectivo do DiffServ é conseguir um método simples e básico de fornecer a capacidade de diferenciar tráfego, através do agrupamento dos fluxos num conjunto limitado e bem definido de classes de serviço (CoS), para suportar diferentes tipos de aplicação e tráfego. Este agrupamento é apenas realizado nos nós exteriores das redes (nós fronteira ou *edge*). Desta forma, não é necessário armazenar informação sobre qual o tratamento a dar a cada fluxo, mas apenas sobre cada uma das N classes de agregação.

O modelo DiffServ baseia-se num conjunto pequeno e bem definido de blocos a partir dos quais uma grande variedade de comportamentos agregados se podem obter.

De modo a obter QoS extremo-a-extremo, a arquitectura DiffServ apresenta dois componentes fundamentais: a marcação de pacotes e os "comportamentos por salto" (*Per Hop Behaviours* – PHBs).

Numa rede DiffServ, os nós fronteira têm responsabilidades diferentes das dos nós do núcleo. Os nós fronteira têm as tarefas de classificação dos pacotes e de

condicionamento do tráfego. Os nós do núcleo necessitam apenas de classificar os pacotes com base nas classes de serviço mapeadas no cabeçalho dos pacotes. Os nós fronteira são os responsáveis pela verificação do cumprimento dos limites dos contratos, procedendo a remarcações, formatação e descarte, caso exista violação dos valores contratados.

A desvantagem deste modelo reside no facto de não estar orientado ao fluxo e, portanto, só por si, não garantir que cada aplicação recebe os recursos de que necessita. Para colmatar esta deficiência, conjuntamente com este mecanismos poderão ser usados agentes de gestão e controlo de admissão (*Brokers*), como veremos mais adiante.

4.3.3.1 Marcação de pacotes

Existe no cabeçalho de cada pacote IP um campo de 8 bits, TOS no IPv4 e Classe de Tráfego no IPv6, que é utilizado para marcar um pacote de forma a que este receba um tratamento particular no encaminhamento e em cada um dos saltos nos nós de rede. Para obter uma operação consistente e comportamentos previsíveis em ambientes inter domínio e multi-vendedor, é necessário que exista um entendimento comum acerca da interpretação e utilização destes campos. Assim, o grupo de trabalho do DiffServ normalizou um formato comum para os 6 primeiros bits de ambos os octetos (TOS e Classe de Tráfego), chamados de campo DS (*Differentiated Services field*) (Figura 14) na arquitectura DiffServ. A cada valor possível de obter com esses 6 bits chama-se DSCP (*Differentiated Services Codepoint*). Os RFCs 2474 [Nich98] e 2475 [Carl98] definem a arquitectura e a utilização geral dos bits do campo DS (sobrepondo-se às definições do RFC 1349 acerca do TOS).

Utilizando a marcação por DSCP, em qualquer nó da rede é possível definir 64 classes ou agregados de serviços distintos (2^6). No modelo DiffServ, a classificação e a QoS é controlada pelo campo DSCP.

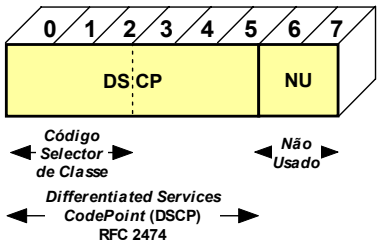


Figura 14: Campo DSCP

4.3.3.2 Per Hop Behaviours (PHBs)

A marcação dos DSCPs dos pacotes só por si não resolve a questão de atribuir tratamento diferenciado a determinado pacote IP. Definiu-se que os pacotes marcados com o mesmo código DSCP, que atravessem determinado nó de rede em determinada direcção, teriam um comportamento semelhante, denominado de Comportamento do Agregado (*Behaviour Aggregate* – BA). Desta forma, pacotes relativos a diferentes aplicações, e provenientes de terminais distintos podem pertencer ao mesmo BA. O RFC 2475 define um PHB como sendo o comportamento observável a partir do exterior, aplicado por um nó DiffServ a um agregado DiffServ. Em termos mais específicos, um PHB refere-se ao escalonamento, formatação, policiamento e *queuing* que um nó DiffServ aplica a um qualquer pacote pertencente a um BA, de acordo com as políticas e SLAs do gestor.

Até à data existem quatro PHBs normalizados, que permitem implementar redes DiffServ, de modo a conseguir definir classes de QoS que possibilitem garantir QoS extremo-a-extremo.

4.3.3.2.1 Default PHB (Definido no RFC-2474)

O *default* PHB [Nich98] especifica que um pacote marcado com um valor DSCP '000000' (valor recomendado) terá nos nós DiffServ o tratamento do serviço tradicional de melhor esforço (*Best Effort*). De igual forma, se um pacote chega a um nó DiffServ com um valor de DSCP que não pertence a nenhum outro PHB, será mapeado para o *default* PHB.

4.3.3.2.2 Class-Selector PHBs (Definido no RFC-2474)

De modo a preservar compatibilidade com o esquema de precedências IP, os valores dos DSCPs foram definidos segundo a forma 'xxx000', onde cada x pode ser 0 ou 1. Estes DSCPs são designados DSCP Selectores de Classe (*Class-Selector Codepoints*) [Nich98]. É de salientar que o DSCP do *default* PHB é também um DSCP Selector de Classe. Estes PHBs têm um comportamento semelhante àquele que é obtido utilizando um esquema baseado em precedências IP. Como exemplo, os pacotes marcados com um DSCP de '110000' (Precedência IP de 110) têm um tratamento (escalonamento e *queuing*) preferencial em relação a pacotes marcados com um DSCP de '100000' (Precedência IP de 100).

Estes PHBs asseguram desta forma que um nó DiffServ seja compatível e possa coexistir com nós que implementem um esquema de Precedência IP (com excepção dos bits DTR – estes não são compatíveis).

4.3.3.2.3 Expedited Forwarding (EF) PHB (Definido no RFC-2598)

Tal como o RSVP no modelo IntServ oferece um serviço de largura de banda garantida, o PHB EF [Jaco99] é o PHB chave na arquitectura DiffServ para oferecer serviços de perdas baixas, atraso baixo, variação de atraso baixa e largura de banda garantida, vulgarmente designados de serviços "*premium*".

O PHB EF pode ser implementado utilizando filas de espera de prioridade estrita, juntamente com limitação de taxa de transmissão. Apesar do EF fornecer um serviço *premium*, deve ser utilizado preferencialmente apenas pelas aplicações mais críticas (e.g., voz), uma vez que tem prioridade sobre todo o restante tráfego. Isto deve-se ao facto de, em caso de congestionamento, se todas as aplicações estiverem a utilizar prioridade elevada, não será possível dar um tratamento de prioridade elevada a todo o tráfego. O valor do DSCP recomendado pelo RFC 2474 para o EF é o '101110'.

4.3.3.2.4 Assured Forwarding (AFxy) PHB Group (Definido no RFC-2597)

O PHB AF [Hein99] permite implementar em ambientes DiffServ um serviço equivalente ao serviço de carga controlada da arquitectura IntServ. Este PHB define um método segundo o qual se podem dar garantias de encaminhamento distintas aos BAs. O tráfego pode, por exemplo, ser dividido em três classes, "ouro", "prata" e "bronze", sendo atribuída 50% da largura de banda à classe 'ouro', 30% à 'prata' e os restantes 20% à 'bronze'. O PHB AFxy define quatro classes, a AF1, AF2, AF3 e a AF4. A cada uma destas classes é atribuído determinado espaço de *buffer* e de largura de banda na interface, dependendo do SLA com o fornecedor de serviço. Em cada uma destas classes é possível especificar 3 valores de precedência de descarte. Assim, em caso de congestionamento num nó DiffServ numa ligação específica, se for necessário descartar pacotes pertencentes a uma classe AFx, os pacotes da AFxy serão descartados de forma a que a probabilidade de descarte de pacotes da AFx1 será menor que a probabilidade de descarte da AFx2 que por sua vez também será menor que a da AFx3. Isto quer dizer que, em média, os pacotes da AF13 serão descartados antes dos da AF12 e estes antes da AF11. A aplicação deste conceito permite que os pacotes pertencentes a fluxos de um BA que tenham excedido a

largura de banda contratada, possam ser penalizados em relação a fluxos que estão dentro do contrato. Esses pacotes podem ser remarcados para terem uma precedência de descarte maior, e desta forma, em caso de congestionamento, serão os primeiros a ser descartados.

A Tabela 6 apresenta os valores dos DSCPs para cada uma das classes e respectivas precedências de descarte. Os três primeiros bits representam a classe (001, 010, 011 e 100) enquanto que os dois seguintes representam a precedência de descarte (01, 10, 11). O último bit tem sempre o valor '0'.

Precedência de descarte	Classe #1	Classe #2	Classe #3	Classe #4
Baixa	(AF11) 001010	(AF21) 010010	(AF31) 011010	(AF41) 100010
Média	(AF12) 001100	(AF22) 010100	(AF32) 011100	(AF42) 100100
Alta	(AF13) 001110	(AF23) 010110	(AF33) 011110	(AF43) 100110

Tabela 6: Tabela dos DSCPs do PHB AF

4.3.3.3 Relação do DiffServ com os SLAs e SLSs

A arquitectura DiffServ define que o SLA pode incluir regras para condicionar o tráfego que em parte constituem o acordo de condicionamento de tráfego, TCA (*Traffic Conditioning Agreement*). O TCA é "um acordo que especifica regras de classificação e quaisquer padrões correspondentes de tráfego, contadores, regras de marcação, descarte e/ou formatação, que são para aplicar a fluxos de tráfego seleccionados pelo classificador" [QoSF][QoSF98]. Um TCA envolve todas as regras de condicionamento do tráfego explicitamente especificadas no SLA bem como as regras implícitas dos requisitos de serviço relevantes e/ou de uma política de aprovisionamento num domínio DiffServ.

Uma especificação de condicionamento de tráfego, TCS (*Traffic Conditioning Specification*) é um conjunto de parâmetros e os seus valores, que associados especificam um conjunto de regras de classificação e um padrão de tráfego. Um TCS é um elemento integral de um SLS.

4.3.4 Classificação de tráfego nos *routers*

Existem outras estratégias de garantir QoS a fluxos com determinadas garantias. Uma dessas estratégias, que não segue nenhuma norma e que portanto se designa por *ad-hoc*, consiste na implementação de regras de tratamento diferenciado de acordo com um ou

mais dos campos dos cabeçalhos IP ou TCP/UDP. Este tipo de classificação e tratamento diferenciado pode ser executado apenas nos nós fronteira ou na rede toda. No entanto, como não se trata de uma estratégia normalizada, será válida, à partida, apenas no domínio onde é implementada.

No nível 3 (IP) / nível 4 (TCP/UDP), os *routers* e comutadores de nível 3 podem classificar o tráfego através dos campos seguintes dos pacotes IP:

- Endereço IP origem/destino: O tráfego relativo a uma dada aplicação também pode ser prioritizado com base no endereço IP. Esta aproximação é ideal para dispositivos com atribuições estáticas de endereços IP. O administrador da rede pode configurar os *routers* para filtrarem (classificarem) e prioritizarem todos os pacotes originados nestes endereços IP. Para facilitar a marcação dos pacotes através do endereço IP podem seleccionar-se gamas de endereços IP para atribuição e o *router* ou comutador de nível 3 pode classificar os pacotes baseado nesta gama. De notar que este método não permite a diferenciação do tipo de fluxo (voz, vídeo ou sinalização) se não for utilizado nenhum filtro adicional para classificação dos fluxos.
- Número do porto TCP/UDP origem/destino: Pode ser atribuída uma gama de portos para determinadas aplicações e os *routers* podem classificar os pacotes baseados nessa gama. Este método tem um inconveniente: se o mesmo porto for utilizado por outros terminais para uma outra aplicação, ser-lhe-á atribuída uma QoS incorrecta.
- Identificador de protocolo: A classificação por identificador de protocolo é a menos útil para determinar o tipo de aplicação envolvida. O protocolo RTP, por exemplo, é usado em muitas aplicações multimédia distintas, tais como voz, vídeo e fax em tempo real, e por isso não pode ser utilizado com rigor para a classificação da aplicação.
- DSCP: O campo DSCP pode ser utilizado para a sinalização de QoS extremo-a-extremo e é através dele que, ao nível da WAN, os Fornecedores de Serviço *Internet* (ISPs) implementam os diversos serviços a oferecer aos clientes. Isto provoca desde já um problema de uniformidade de critérios entre os vários ISPs, já que num sistema deste tipo a coerência tem que ser mantida em todo o percurso. A solução passa por todos os ISPs utilizarem serviços

semelhantes, do conhecimento de todos. Este mecanismo é também implementado na LAN, havendo dispositivos terminais que fazem a marcação dos pacotes através do DSCP.

4.3.5 MPLS

O MPLS [Rose01] emergiu como uma solução simples para responder aos requisitos de gestão de largura de banda e de serviços das redes de *core* IP. É uma solução versátil para resolver alguns problemas das redes de comunicações de pacotes, ao nível do *core*, nomeadamente aspectos de velocidade, escalabilidade, gestão de QoS e engenharia de tráfego. O MPLS cobre os problemas de escalabilidade e encaminhamento (baseado em métricas de QoS e de serviço) e pode coexistir sobre redes ATM ou *Frame Relay*. Pode mesmo afirmar-se que o MPLS nasceu com base em muitos dos conceitos existentes no ATM (e também no *Frame Relay*), tais como, por exemplo, a comutação baseada em identificadores de ligação.

Em redes MPLS a transmissão de dados é realizada sobre caminhos denominados de *Label Switched Paths* (LSPs), baseados na comutação de etiquetas. Cada pacote de informação carrega consigo uma etiqueta que define o caminho até ao próximo nó de rede. Nos nós de rede pode haver tradução de etiquetas. Devido ao comprimento fixo das etiquetas é possível obter comutação de alta velocidade, por exemplo baseada em processamento por *hardware*. Os LSPs são uma sequência de etiquetas em todos os nós da rede desde a origem até ao destino. Há duas formas de estabelecer os LSPs: antes do início da transmissão de informação (iniciados por controlo) ou baseados na detecção de um determinado fluxo (iniciados pela informação). As etiquetas que constróem os caminhos são distribuídas utilizando um protocolo designado *Label Distribution Protocol* (LDP), ou utilizando RSVP associado a protocolos de encaminhamento (tais como BGP – *Border Gateway Protocol* – ou *OSPF – Open Shortest Path First*).

As decisões de atribuição e gestão de etiquetas podem ser tomadas baseadas em diversos critérios, tais como:

- Protocolo de encaminhamento
- Engenharia de tráfego
- VPN (*Virtual Private Network*)
- QoS

O MPLS permite portanto o isolamento de tráfego e permite a gestão de agregados de tráfego de acordo com as suas características.

As operações realizadas antes e durante a transmissão dos pacotes numa rede MPLS são as seguintes:

- Criação e distribuição das etiquetas
- Criação da tabela de encaminhamento em cada nó
- Criação do caminho
- Inserção das etiquetas
- Encaminhamento dos pacotes

Num domínio MPLS não é forçoso que toda a informação entre dois nós siga o mesmo percurso. Podem existir diversos caminhos, com LSPs distintos, dependentes, por exemplo, das características do tráfego e eventualmente das suas classes de serviço. A Figura 15 ilustra o processo de criação de LSPs e encaminhamento num domínio MPLS. O fluxo de dados é encaminhado após a criação e distribuição das etiquetas.

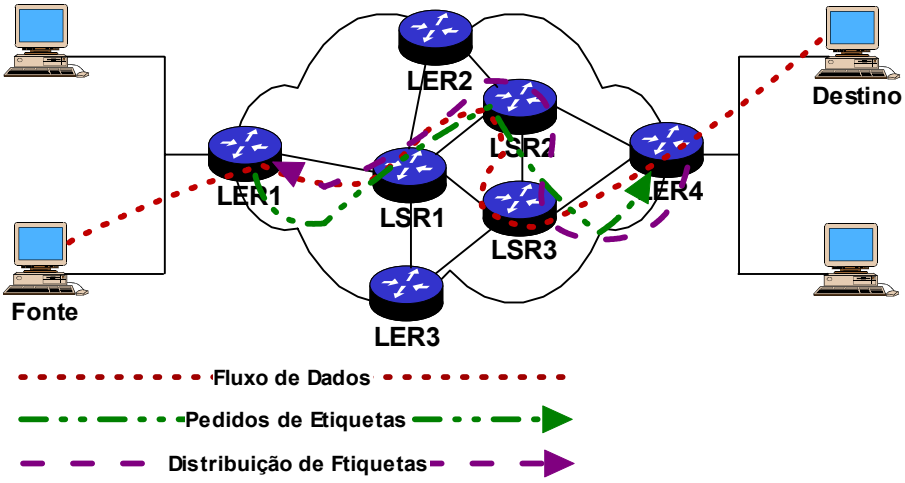


Figura 15: Criação do LSP e encaminhamento de pacotes num domínio MPLS

4.4 Gestão e controlo de QoS

Apesar de existirem diversos mecanismos de fornecer QoS em redes IP, para que seja possível controlar e garantir essa QoS de uma forma realista, é necessário aliar esses mecanismos a sistemas de gestão e controlo que tirem o máximo proveito dos recursos existentes. Os modelos actualmente mais populares, denominados de *Policy Based*

Management (PBM), recorrem à utilização de políticas de gestão, armazenadas em servidores e implementadas por elementos específicos.

4.4.1 Políticas e servidores de políticas de QoS

A norma desenvolvida pelo IETF para o controlo de políticas de QoS define pontos de imposição de políticas (PEP, *Policy Enforcement Points*) e pontos de decisão de políticas (PDP, *Policy Decision Point*) [Yava00] (Figura 16). As políticas estão armazenadas num servidor de políticas (PS - *Policy Server*) e são transmitidas aos PDPs de modo a estes as imporem em cada um dos PEPs. Os PEPs incluem *routers*, comutadores e outros equipamentos capazes de actuar como agentes para o controlo de admissão. Normalmente os PEPs trabalham em conjunto com os PDPs por forma a implementar as políticas de QoS definidas pelo administrador da rede (ou, num sentido mais lato, pelo operador). Os PDPs detêm a inteligência necessária para o processamento de políticas abstractas, e respectiva tradução.

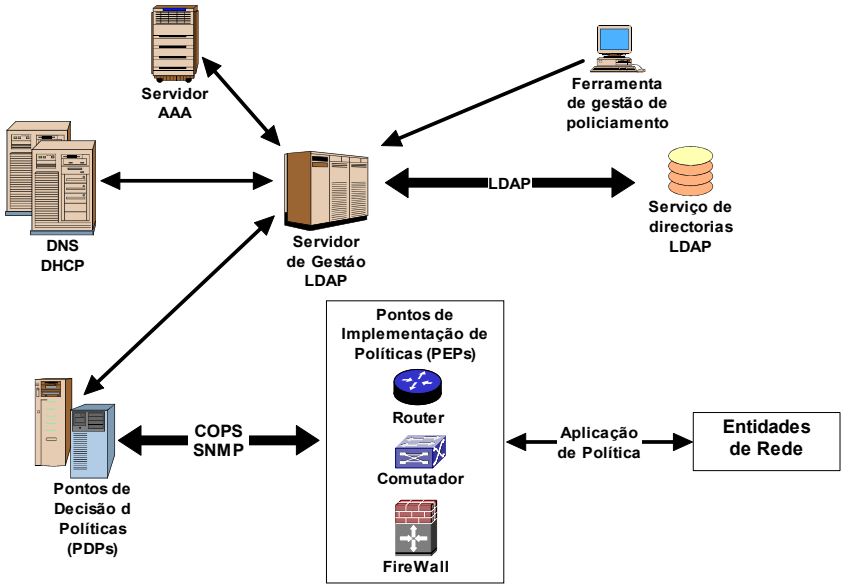


Figura 16: Arquitectura de políticas e policiamento do IETF

4.4.1.1 COPS / PEPs / PDPs

Em [Boyl00] é descrito o protocolo COPS (*Common Open-Policy Service*) que é um protocolo de pergunta e resposta simples utilizado para trocar informação de políticas entre um servidor de políticas (PDP) e os seus clientes (PEPs) conforme se observa na Figura 17.

O protocolo COPS define três tipos de mensagens distintas: pedido, decisão e relatório. O PEP efectua pedidos e o PDP responde com decisões. As mensagens de relatório são também enviadas pelo PEP ao PDP mas não necessitam de resposta por parte deste. As mensagens de pedido podem conter um ou mais elementos de política, os quais serão examinados pelo PDP para determinar a resposta. A mensagem resposta, de decisão, indicam ao PEP as acções a tomar em relação ao pedido realizado. As mensagens de relatório são utilizadas, por exemplo, para o PEP passar informação de estado ou de contabilização ao PDP.

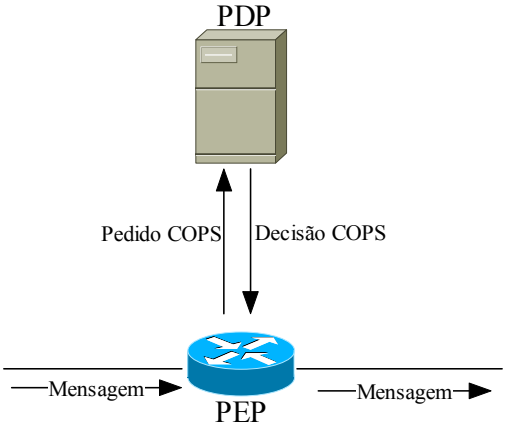


Figura 17: Interacção COPS entre PEP e PDP

Existem duas variantes do protocolo COPS: o COPS-PR e o COPS-RSVP. As diferenças entre o COPS-PR e o COPS-RSVP são as seguintes: o COPS-PR apresenta uma arquitectura de provisão. O PEP envia ao PDP um pedido com uma lista de requisitos de políticas. O PDP devolve as políticas (*queuing*, *buffering*, prioridade, etc.) numa mensagem de decisão PDP. Após isto, o funcionamento do COPS-PR resume-se quase exclusivamente à manutenção de *keep-alives*. O PDP pode enviar decisões não solicitadas e o PEP pode fazer novos pedidos ou renovar os existentes. A arquitectura COPS-RSVP é de *outsourcing* no sentido em que as políticas não estão pré-definidas. O PEP faz pedidos apenas quando surge um evento exterior que o force. No caso do COPS-RSVP, os pedidos efectuados pelo PEP podem incluir a parametrização da política pretendida. A decisão do PDP é relativamente simples, limitando-se a uma aceitação ou rejeição (ou aceitação com alterações).

O COPS pode ser utilizado para o PDP efectuar a configuração inicial dos PEPs do seu domínio e pode ser utilizado para configuração de políticas fluxo a fluxo. No primeiro

caso (COPS-PR), o PEP quando é activado, envia um pedido ao PDP, de acordo com as suas características para que seja correctamente configurado. No segundo caso (COPS-RSVP), por cada novo fluxo que pretenda ser admitido, o PEP faz um pedido ao PDP que lhe responde com a sua decisão de aceitar ou não o fluxo.

Algumas das decisões passadas pelo PDP ao PEP são a classificação dos pacotes (prioridades), as taxas de transmissão, os dimensionamentos das filas de espera, a formatação do tráfego, a definição das classes de serviço DiffServ e as estratégias de descarte e remarcação de pacotes.

4.4.2 QoS *Brokers* (Bandwidth *Brokers*)

Um *Bandwidth Broker* ou *QoS Broker* é uma entidade que realiza controlo de admissão e reconfiguração de dispositivos de rede de acordo com um conjunto de condições impostas pelos administradores de rede, com o objectivo de obter QoS extremo-a-extremo (eventualmente atravessando várias redes de vários fornecedores). Um determinado *QoS Broker* tem o seu domínio de acção bem definido. De modo a garantir QoS estrita, extremo-a-extremo, todos os *QoS Brokers* no caminho da ligação pretendida devem interagir em concordância.

Um QoS *Broker* é um PDP e pode também ser visto como um *Policy Server* (PS). Existem algumas funcionalidades exigidas a um PS que não o são aos *QoS Brokers*. Neste sentido, um *QoS Broker* pode ser uma parte de um PS, ou em alternativa, pode interagir com um. De um modo análogo, um *QoS Broker* pode ser parte integrante de um sistema de gestão de rede (NMS – *Network Management System*), não se distinguindo deste, ou pode ser uma entidade externa, implementando apenas algumas das funcionalidades do NMS.

Desta forma, o *QoS Broker*, o PS e o NMS podem ser sistemas distintos ou podem eventualmente estar concentrados numa única entidade. No caso de se tratarem de sistemas distintos, torna-se necessário a existência de interfaces e protocolos entre eles, que permitam o acesso a este sistema partilhado e distribuído. No caso de todas as funcionalidades estarem concentradas apenas numa única entidade, é natural que surjam problemas de escalabilidade à medida que a rede cresce e se torna mais complexa.

De um modo simplificado podemos dizer que um *QoS Broker* controla e monitoriza os recursos de rede tanto no seu domínio como nas extremidades. Também é sua responsabilidade monitorizar nas extremidades a existência de pedidos de reserva, quer para o interior do seu domínio, quer para o exterior. A informação recolhida é utilizada em

conjunção com a informação do PS de modo a tomar as decisões de controlo de admissão. Os SLSs são implementados pelos *QoS Brokers* dentro do seu domínio. O *QoS Broker* é também responsável por controlar, conjuntamente com os *QoS Brokers* vizinhos, as ligações inter-domínio de modo a coordenar os SLSs através de vários domínios.

A coordenação dos SLSs entre os vários domínios poderá ser realizada de um modo agregado nas fronteiras dos domínios. Isto é, todos os fluxos de determinado domínio que tenham características de QoS semelhantes podem ser tratados como um único fluxo, com características iguais à soma dos fluxos individuais. Para um fluxo ser admitido em determinado domínio deve inicialmente ser feito um pedido de alocação de recursos (RAR – *Resource Allocation Request*) que será analisado pelo *QoS Broker* de modo a este poder coordenar a alocação e fornecimento dos recursos necessários a todos os fluxos e agregados originados dentro e fora do seu domínio.

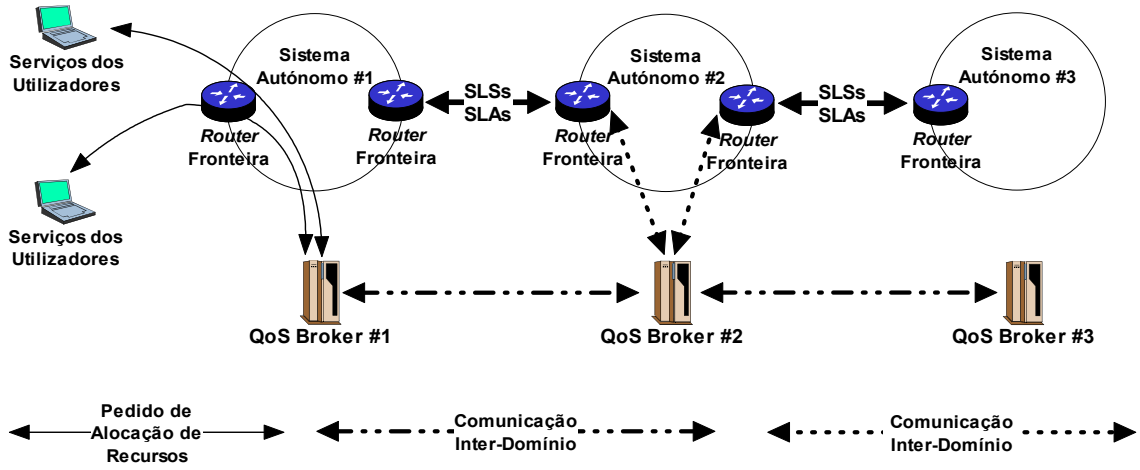


Figura 18: Exemplo de configuração de rede

Para melhor entender o conceito, a Figura 18 [Neil99] apresenta um exemplo simples de uma configuração de rede. Esta rede é composta por três domínios, os sistemas autónomos #1, #2 e #3 com os respectivos *QoS Brokers*, *QoS Broker* #1, *QoS Broker* #2 e *QoS Broker* #3. Também são mostrados os SLSs/SLAs em vigor entre os diferentes domínios assim como utilizadores individuais do sistema autónomo #1 que fazem pedidos de alocação de recursos ao *QoS Broker* #1. Neste caso, o "utilizador individual" pode ser, quer um sistema, quer uma aplicação a fazer um pedido de largura de banda. Este é o modelo base desenvolvido pela *Internet*2 [INTQ].

4.5 Fornecimento de QoS: modelo do operador

Esta secção aborda o modelo seguido pelos operadores para provisão de QoS em diversos pontos da rede, a relação inter-operador e algumas estratégias de QoS extremo-a-extremo. Primeiro é abordada a divisão da rede em segmentos distintos (cliente, acesso e *core*) sendo referidas as técnicas mais utilizadas para a obtenção de QoS em cada um desses pontos. São depois abordadas as questões dos relacionamentos inter-operadores e por fim, são referidas estratégias integradas de controlo de QoS extremo-a-extremo.

4.5.1 QoS em diferentes pontos da rede: cliente, acesso e core.

4.5.1.1 Cliente

As redes de cliente poderão ser divididas em duas classes distintas: empresariais ou domésticas. O cliente é normalmente o responsável pela garantia da QoS e do bom funcionamento nestas redes.

Neste tipo de redes (LAN) a tecnologia vulgarmente utilizada é a *Ethernet*, necessitando esta de mecanismos de QoS menos sofisticados do que os usados na WAN, devido a possuir larguras de banda mais elevadas, resultando em atrasos menores. Para serviços interactivos sensíveis a atrasos devem usar-se as agora vulgares redes *Ethernet* comutadas, associadas aos mecanismos de nível 2 suportados por estas (802.1p e q) e deve optar-se pela utilização dos *codecs* de melhor qualidade dado não existirem restrições de largura de banda. Nas LANs existe grande flexibilidade em termos de controlo de QoS por parte do gestor/administrador de rede.

4.5.1.2 Acesso

A rede de acesso é da responsabilidade do operador. Dependendo das tecnologias associadas, é utilizada quer a arquitectura DiffServ, quer a IntServ com a utilização do protocolo de reserva de recursos RSVP, quer soluções específicas com recurso a PEPs e PDPs. No entanto, a QoS a nível do IP que pode ser oferecida depende fortemente da tecnologia de suporte. Algumas tecnologias utilizadas no acesso têm suporte nativo de QoS no nível 2, enquanto que outras partilham o meio e não oferecem garantias de QoS. De entre as várias tecnologias utilizadas no acesso destacam-se o *Frame Relay*, o DOCSIS, o xDSL, a RDIS, o ATM, a TD-CDMA (*Time Division CDMA*) e a *Wireless* LAN (IEEE 802.11).

4.5.1.3 Core

A rede de *core* é também da responsabilidade do operador. Nestas redes, designadas de transporte, os mecanismos de QoS podem ser implementados em diferentes níveis. Tradicionalmente, o transporte de tráfego IP em redes de transporte era feito com base num modelo em sobreposição, geralmente "IP-sobre-ATM". Apesar de uma superior capacidade de controlo dos mecanismos de QoS oferecidos pelo ATM, este modelo implica o mapeamento estático entre classes IP e classes de tráfego ATM, uma vez que as características do ATM não são "visíveis" do ponto de vista da camada IP.

Hoje em dia existem outra soluções. A introdução do modelo DiffServ veio possibilitar a implementação de mecanismos de QoS no nível IP, de uma forma escalável e independente das características da rede de transporte, nomeadamente das tecnologias de nível 2. O MPLS é também bastante utilizado para o controlo de QoS no *core*, devido à sua grande flexibilidade e simplicidade.

Aliadas a estas estratégias de controlo de QoS no *core* das redes, poderá existir também um sistema de gestão de rede, apoiado em agentes e ferramentas de monitoria e QoS *Brokers*, que permitirá um controlo mais fino, preciso e optimizado dos recursos disponíveis.

4.5.2 Fornecimento de Rede e Serviços

Nas secções anteriores abordaram-se genericamente os mecanismos de QoS disponíveis não só ao nível da aplicação mas também ao nível dos vários segmentos de rede. Para a oferta de QoS extremo-a-extremo têm que se considerar outros aspectos como sejam a negociação de QoS e toda a gestão de recursos (policiamento, reserva, atribuição e libertação) durante a sessão. A tarefa complica-se quando se começam a relacionar todos estes aspectos com múltiplos domínios, quer ao nível de fornecedores de rede quer de serviço.

4.5.2.1 Fornecedor único de rede e de serviços

Neste caso, o fornecedor da rede extremo-a-extremo e o fornecedor de serviços é a mesma entidade. Assim para o estabelecimento de uma sessão com QoS é necessário:

1. A aplicação ou utilizador definir a QoS pretendida;
2. Os parâmetros QoS de níveis diferentes (i.e. nível 2 e nível 3) serem traduzidos, caso possuam representações diferentes;

3. Os parâmetros de QoS devem ser mapeados nos recursos pretendidos;
4. Esses recursos serem admitidos (verificação de disponibilidade dos recursos), reservados e atribuídos ao longo do percurso emissor – receptor

A Figura 19 mostra um exemplo de um protocolo genérico de reserva/atribuição de recursos com uma resposta de aceitação. Os recursos deverão ser libertados aquando do encerramento da sessão.

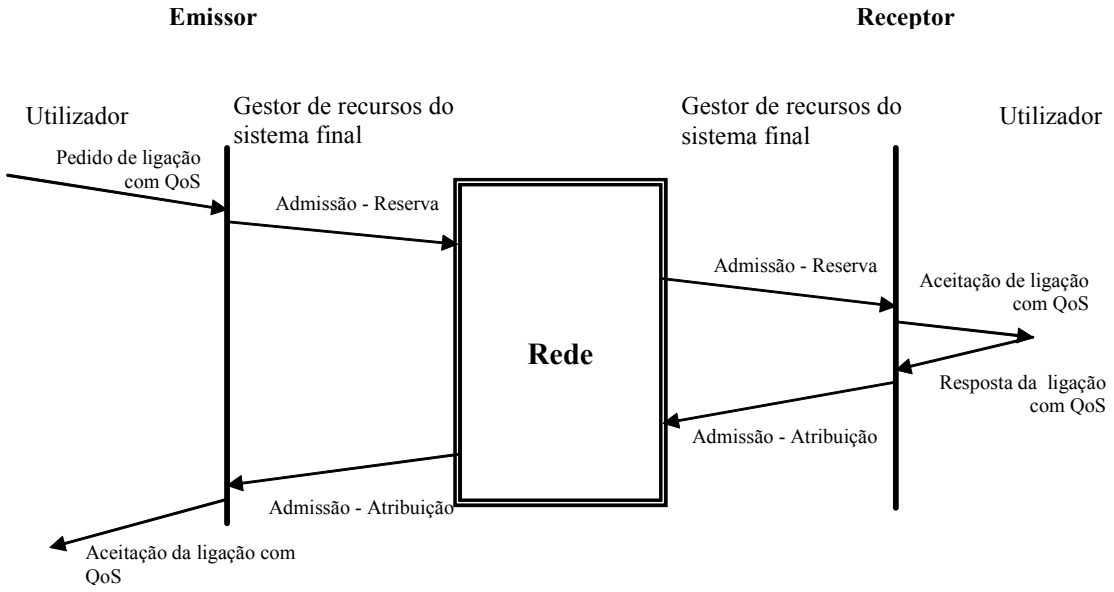


Figura 19: Modelo genérico de admissão de serviço com QoS

As garantias de QoS devem ser mantidas ao nível da aplicação, sistema e rede ao longo da sessão, principalmente se esta tiver requisitos de tempo real. Durante uma transmissão multimédia é necessário considerar as políticas de escalonamento de disciplinas ao nível da aplicação e da rede; gestão de filas de espera; controlo do débito (gestão de largura de banda, prioridades e espaço em filas de espera); controlo de erros extremo-a-extremo; funcionalidades para a monitoria de recursos e funcionalidades para a adaptação de recursos (alteração dinâmica de parâmetros de QoS).

4.5.2.2 Fornecedores múltiplos de rede e de serviços

Quando existem vários fornecedores de rede é imprescindível a existência de acordos entre eles (Figura 20). Esses acordos constituem os SLAs que se pode dizer, de uma forma simplista, que especificam a disponibilidade dos recursos afectos a um determinado domínio vizinho, bem como a definição dos valores dos parâmetros de QoS da rede.

O aspecto de monitoria da QoS na rede é fundamental e é efectuado recorrendo a medições realizadas sobretudo na camada da rede.

A medição intra-domínio, dá uma informação da forma como o trânsito de pacotes está a ser feito. Nestas medições é importante caracterizar cada tipo de tráfego (para diferentes níveis de QoS), medindo atrasos, largura de banda, ou perdas, entre outros parâmetros de interesse. As principais entidades avaliadas são os *routers* de core, os quais deverão implementar disciplinas de filas de espera que tenham em conta as necessidades de cada tipo de tráfego.

A medição inter-domínio, dá informação da forma como o diferente tráfego se comporta extremo-a-extremo (ou entre dois outros pontos). O mais importante destas medições é a garantia de que os contratos estabelecidos, quer com os clientes finais, quer com outros domínios, são cumpridos. Aqui cada domínio é visto como uma caixa negra em que apenas são importantes os valores obtidos nos seus extremos para avaliar se os SLAs estão a ser cumpridos.

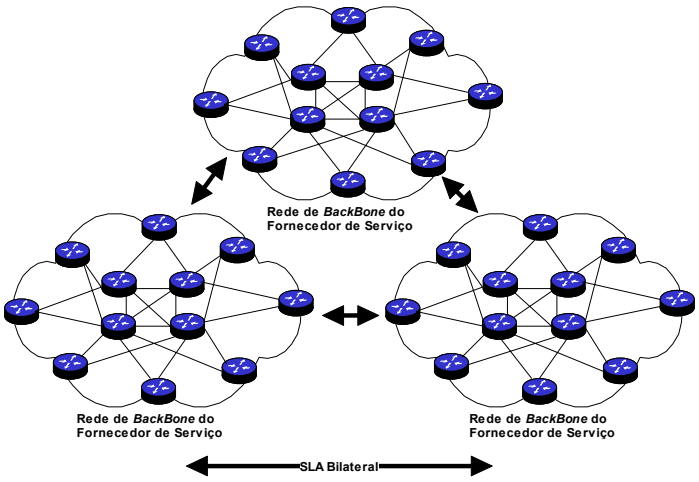


Figura 20: Rede empresarial, intra-ISP e inter-ISP e suas relações para atribuição de QoS

A utilização de um servidor para controlar a admissão de tráfego dentro de um domínio DiffServ foi considerada desde o início da discussão sobre a arquitectura DiffServ [Nich99]. O servidor que controla as admissões, tipicamente designado por *Bandwidth Broker* (*QoS Broker*), decide aceitar ou rejeitar o tráfego com base nos SLAs previamente negociados. No caso de um novo fluxo o *QoS Broker* poderá negociar um novo SLA com os domínios vizinhos dependendo dos requisitos de tráfego.

Os *QoS Brokers* permitem a reserva para fluxos entre pontos formatadores de tráfego (saída) e pontos de policiamento (entrada) nas redes vizinhas.

Os fornecedores de serviço possuem acordos com os seus fornecedores de rede e os fornecedores de rede possuem acordos entre si.

4.5.3 Estratégias de controlo de QoS Extremo-a-Extremo

As alternativas de gestão e controlo de QoS extremo-a-extremo são várias e importa ter este aspecto em consideração quando se projectam cenários de multi-operador, uma vez que poderão coexistir diversas estratégias dos diferentes operadores na provisão de um único serviço extremo-a-extremo, entre dois terminais de domínios distintos. Podem considerar-se abordagem de IntServ puras, abordagens de DiffServ puras, abordagens mistas IntServ/DiffServ/IntServ, abordagens com MPLS no *core* e abordagens com o recurso a QoS *Brokers*.

A tendência actual aponta para estratégias com IntServ nas extremidades das redes, sendo utilizado um *core* baseado em DiffServ/MPLS com agregação de fluxos de características semelhantes. Do IntServ tira-se partido da sua capacidade de oferecer garantias extremo-a-extremo a fluxos individuais, enquanto do DiffServ/MPLS se tira partido da sua boa escalabilidade no *core*.

A utilização de QoS *Brokers* para fazer o controlo de admissão na entrada dos domínios pode excluir a utilização de IntServ nesses pontos, e mesmo no *core*, a utilização de QoS *Brokers* pode ter como objectivo a gestão dos domínios DiffServ e mesmo dos caminhos MPLS.

O MPLS segue o princípio da concentração da inteligência dos serviços nos *routers* de fronteira e o transporte do tráfego pelos *routers* de núcleo usando um algoritmo muito simples baseado unicamente numa etiqueta, representada na Figura 21. Neste sentido, o MPLS pode facilmente ser associado a sistemas DiffServ, com a conversão do campo EXP no campo DSCP do DiffServ e vice-versa.

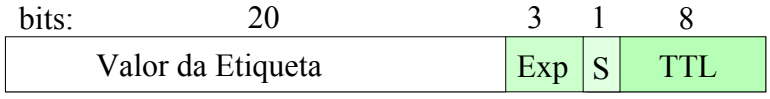


Figura 21: Etiqueta MPLS ('shim header')

O problema básico da transposição do modelo DiffServ para as redes MPLS é o facto de um *router* MPLS (LSR - *Label Switching Router*) não analisar o conteúdo do cabeçalho

IP, mas apenas a etiqueta MPLS. Por esse motivo, os mecanismos de qualidade de serviço utilizados nas redes IP (em que é utilizado o cabeçalho IP para transportar informação de QoS) não podem ser aplicados directamente no caso das redes MPLS.

Existem duas aproximações possíveis para implementar o modelo DiffServ numa rede MPLS (Figura 22), dependendo da forma como cada LSR faz a classificação dos pacotes em PHBs (isto é, a associação de cada pacote na respectiva classe DiffServ) num túnel MPLS (LSP):

- L-LSP (*Label-only-inferred-LSP*): o PHB é determinado em função do valor da etiqueta MPLS;
- E-LSP (*Exp-inferred-LSP*): o PHB é determinado a partir do campo Exp (3 bits da etiqueta MPLS).

A utilização de E-LSP segue o modelo DiffServ e permite criar um único LSP para transportar tráfego pertencente a várias classes de serviço, reduzindo assim o número de LSPs necessários numa rede MPLS. Podem ser definidas no máximo 8 classes de serviço por LSP (3 bits disponíveis para transportar informação de QoS).

A utilização de L-LSP tem a vantagem de não impor um limite no número de classes de serviço, mas exige um número de LSPs potencialmente muito elevado (deve notar-se que este é o único modo possível no caso do '*Cell-based MPLS*', em que não é utilizado o '*shim header'* MPLS, mas sim o mapeamento da etiqueta num par VP/VC).

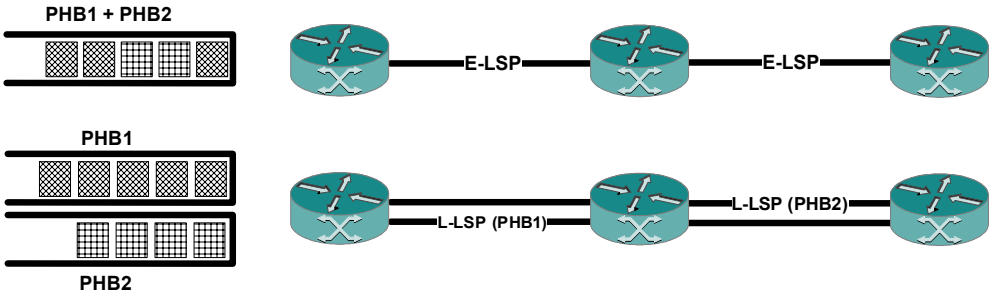


Figura 22: Abordagens DiffServ em MPLS: E-LSP versus L-LSP

CAPÍTULO 5

NOVA ARQUITECTURA DE QOS ORIENTADA AO FORNECIMENTO DE SERVIÇO

5.1 Introdução

O trabalho apresentado neste capítulo descreve uma arquitectura de rede de próxima geração, implementada utilizando tecnologias existentes. Esta arquitectura tenta responder à visão expressa no capitulo 2, tendo em atenção as componentes tecnológicas discutidas nos capítulos 3 e 4. Será apresentada uma arquitectura de rede "true-IP" (no sentido em que ao nível da rede serão utilizados somente mecanismos de controlo associados ao protocolo IP) com suporte de mobilidade, assegurando diferenciação de qualidade de serviço (QoS) por utilizador e serviço, com características de gestão suficientemente poderosas para permitir o desenvolvimento e criação de serviços avançados num ambiente de operador (i.e., suportando também AAAC). Aspectos importantes, tais como a segurança, não foram deixados de parte e estão também integrados nesta arquitectura. A arquitectura aqui apresentada foi desenvolvida no âmbito do projecto IST – Moby Dick

[Moby] de modo a suportar (1) mecanismos de *handover* transparente entre redes de acesso baseadas em tecnologias distintas (no caso, TD-CDMA, IEEE 802.11b e *Ethernet*), representativas de acessos de clientes distintos; (2) QoS extremo-a-extremo adaptada à mobilidade (aspecto central do trabalho aqui realizado); (3) sinalização de *paging* ao nível IP; (4) e suporte de mecanismos de AAAC.

Esta arquitectura não derivou das visões tradicionais "All-IP" dos grupos industriais (tais como o 3GPP) uma vez que se situa num horizonte temporal mais longínquo, mas é relativamente fácil a evolução dessas aproximações (que ainda se baseiam em protocolos de rede específicos ao nível do acesso rádio) de modo a convergirem na arquitectura aqui proposta. Nesta arquitectura é essencial que os *handovers* sejam executados o mais rápido possível, dado o tipo de serviços que se pretende suportar (interactivos e de tempo real).

O trabalho realizado neste doutoramento foi centrado no domínio específico da arquitectura de QoS, mas será inevitável uma breve descrição de outros aspectos da rede devido à interoperação de diferentes entidades e protocolos.

5.2 Elementos básicos da arquitectura de rede

A arquitectura de rede aqui apresentada [VMar03b] foi desenvolvida de modo a suportar três redes de acesso distintas (*Ethernet*, *Wireless* LAN e TD-CDMA) sendo suficientemente flexível para suportar outras tecnologias dado que todo o controlo é efectuado ao nível do IP. Os elementos que compõem esta arquitectura são os terminais móveis (os terminais dos utilizadores), os *routers* de acesso ou *gateways* (os pontos de acesso dos terminais móveis à rede), um sistema de AAAC por cada domínio de operador, um *home agent* (de modo a suportar mobilidade – IPv6), um ou mais QoS *Brokers* (por cada domínio de QoS – um domínio administrativo poderá ter vários domínios de QoS) e um ou mais agentes de *paging* por cada domínio. Estes elementos estão representados na Figura 23. Toda a sinalização entre os diversos elementos é efectuada ao nível do IP de modo a atingir a desejada convergência inter-tecnologia, independente da tecnologia de nível 2. As funções disponibilizadas por cada uma destas entidades e o seu interfuncionamento definem uma arquitectura que suporta um acesso transparente e de baixo custo neste ambiente de heterogeneidade.

De modo a permitir uma mobilidade transparente similar àquela que hoje em dia se pode experimentar nas redes de acesso celulares, esta arquitectura de rede baseia-se em IP

móvel com *fast handover* [Kood02] melhorado com características específicas de QoS e técnicas de transferência de contexto. Esta aproximação baseada em IP tem a vantagem de permitir mobilidade através de células de tecnologias de acesso distintas de uma forma completamente transparente para as aplicações.

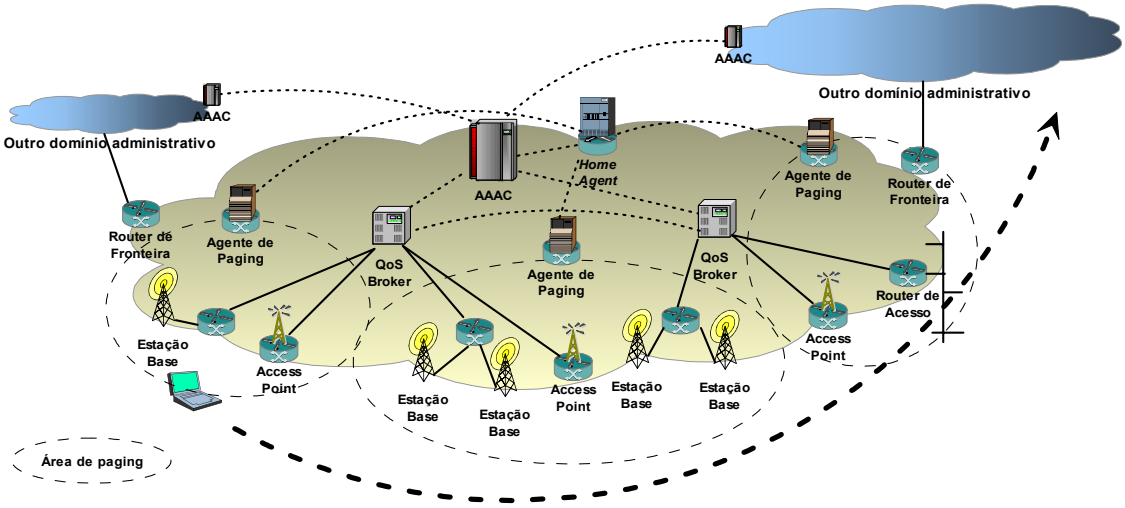


Figura 23: Representação Genérica da Arquitectura de Rede

Para obter este resultado, a infra-estrutura de TD-CDMA foi desenvolvida de modo a que o TD-CDMA seja visto apenas como mais uma tecnologia em tudo idêntica à *Ethernet*, e desta forma foram eliminados os elementos específicos da arquitectura UMTS (3GPP): o RNC, o SGSN e o GGSN [Kaar01] – embora à custa de falta de compatibilidade com os sistemas existentes. No âmbito do projecto Moby Dick foi desenvolvida uma estação base para o suporte de TD-CDMA. O acesso à rede é providenciado pelo *router* de acesso de rádio que controla as células: cada sub-rede IP é directamente mapeada numa célula rádio (este é também o modo de funcionamento das células 802.11b). Assim, mudar de uma célula de uma tecnologia para outra de outra tecnologia é um processo gerido apenas ao nível do IP uma vez que todos os *handovers* são tratados ao nível da rede.

A provisão de QoS, apresentada com detalhe mais à frente, resulta da aplicação de um modelo misto baseado em DiffServ com controlo de admissão (de serviços) realizado por QoS *Brokers*. O suporte de funcionalidades de AAAC é baseado na arquitectura de AAA definida no IETF [Laat00], enriquecida com mecanismos de tarifação de modo a ser aplicável num cenário de utilização comercial. A segurança no acesso é garantida através da utilização de IPSec, devidamente modificado de modo a suportar mobilidade. A informação sensível respeitante a cada utilizador, cuja transferência entre *routers* de acesso

é necessária em caso de mobilidade, é também transferida de um modo seguro. Adicionalmente, um novo conceito de *paging* IP foi integrado na arquitectura geral para permitir uma melhor gestão e optimização da sinalização na rede e de consumo de potência nos terminais móveis, cujas exigências de tamanho e consumo são cada vez mais importantes.

Do ponto de vista de desenvolvimento de serviços de telecomunicações, esta arquitectura tem um conjunto interessante de características:

- É baseada no protocolo IP, existindo técnicas bem conhecidas de desenvolvimento aplicacional para o mesmo, e interfaces bem conhecidas para os programadores.
- Integra diferenciação de QoS, permitindo a coexistência de serviços com diferentes requisitos de qualidade.
- Dispõe de mecanismos individualizados de AAAC, permitindo controlo por utilizador (ou grupo) a serviços, contabilização e facturação individualizada.
- Dispõe assim de mecanismos simples de implementação de subscrição de diferentes serviços
- Apresenta um suporte vertical integrado de serviços, sendo a subscrição de um dado serviço potencialmente reflectida no suporte disponibilizado pela rede (em termos de QoS, por exemplo).
- Permite uma independência do terminal, realizando uma associação do utilizador ao terminal que está a usar naquele momento. Isto permite que diferentes serviços possam ser desenvolvidos para terminais específicos, e que qualquer subscritor possam recorrer a esses serviços quando utilizam esses terminais.
- Integra aspectos de mobilidade: mobilidade de utilizador (pode mudar de terminal) e mobilidade de dispositivo (o mesmo dispositivo pode aceder à rede de diferentes formas).

Além disso, do ponto de vista conceptual, as entidades que um operador de rede tem de gerir para fornecimento de serviços são poucas, e com responsabilidades claramente identificadas: routers de acesso, *Home Agent*, sistema de AAAC, e QoS Brokers. Naturalmente, isto facilita os esforços de gestão do operador, e simplifica o NMS (*Network Management System*).

5.2.1 Suporte de Mobilidade

Nesta arquitectura o suporte de mobilidade é obtido recorrendo ao uso de funções de MIP (*Mobile* IP) e FHO (*Fast HandOver*) capazes de permitir mobilidade através de várias redes e fornecer suporte de *paging* ao nível da rede.

5.2.1.1 Paging

O sistema de *paging* [Marc04] é em tudo semelhante ao que se existe actualmente nas redes celulares e não pretende introduzir conceitos revolucionários, mas somente replicar nas redes de pacotes aquilo que é uma exigência das redes sem fios.

Um dos requisitos da mobilidade IPv6 utilizada nesta arquitectura, é que cada terminal móvel activo deve adquirir um novo endereço em cada célula que visita. Este endereço, designado por CoA (*Care-of-address*), identifica a sub-rede onde o terminal está localizado. Após a obtenção deste endereço o terminal móvel tem de o registar junto do seu *Home Agent*. Torna-se necessário manter informação da localização do terminal móvel a cada instante para permitir a entrega de alguma informação que eventualmente lhe seja dirigida. No entanto, se não existe informação para entregar ao terminal, manter esta informação actualizada junto do *Home Agent* com mensagens de *Binding Update* é uma carga de sinalização algo supérflua e que poderá representar um *overhead* demasiado pesado num ambiente em que centenas ou milhares de terminais se estão a movimentar. Este efeito é tão mais nocivo quanto maior for a velocidade dos terminais e mais pequenas forem as células rádio. Assim, caso não haja informação para ser entregue ao terminal móvel, este deverá entrar num estado *idle* (ou adormecido) e deverá ser reduzida a sinalização de localização. Desta forma é possível reduzir a carga de sinalização na rede, poupar os recursos escassos de rádio e poupar energia do terminal móvel. Nesta arquitectura, o *paging agent* é a entidade responsável pela descoberta da localização do terminal móvel, caso exista informação para lhe encaminhar quando este está adormecido. Os únicos elementos da rede que saberão a cada instante qual o estado do terminal ("adormecido" ou "acordado") são o próprio terminal e o agente de *paging*. Também neste caso, a motivação de tratar do *paging* ao nível da rede (IP) é manter o protocolo de *paging* independente da tecnologia de acesso.

5.2.1.2 Fast-Handover

O *Mobile IPv6* é usado como base para o suporte de mobilidade nesta rede. No entanto, a norma foi definida com vista a uma aplicação de gestão global de mobilidade em tudo semelhante ao *roaming* das redes celulares actuais, sem requisitos de suporte de mobilidade transparente e de serviços em tempo real. Para atingir estes objectivos é necessário adicionar algumas funcionalidades extra que permitam a transparência na mobilidade de modo a tornar imperceptível a mudança de rede de acesso para o utilizador.

Através de análise de resultados de simulação comparativa [Hart02] concluiu-se que uma proposta de mobilidade não-hierárquica, denominada de *Fast-Handover* (FHO) [Kood02], apresenta melhor desempenho para o caso concreto da nossa arquitectura de rede, pois tem o melhor compromisso entre complexidade e transparência do *handover*.

Um *handover* pode ser dividido em dois tipos distintos: *handover* de baixo-nível (que poderá impor constrangimentos rígidos dependendo da tecnologia) e *handovers* de alto-nível ou *handovers* IP. A técnica de FHO utilizada nesta arquitectura é independente da tecnologia de nível 2 e aplica o conceito de *make-before-brake* que consiste em preparar o *handover* de nível 3 através da ligação de nível 2 actual, antes da mudança de célula propriamente dita ser realizada [Marc02]. Desta forma garante-se que os recursos necessários estão já preparados naquela que será a nova célula e o atraso de *handover* será única e exclusivamente dependente da tecnologia de nível 2. Sendo assim, o *handover* de nível 3 (e superiores) não irá introduzir atrasos adicionais pois é preparado antes de a ligação com a célula antiga ser desligada. Deste modo, o atraso de *handover* é reduzido ao mínimo necessário para uma eventual reconfiguração ou mudança de interface de acesso. Durante a fase de preparação o *router* de acesso actual é mantido informado acerca do *handover* pretendido pelo terminal e, durante esta fase, todos os pacotes dirigidos ao terminal são transmitidos também àquele que será o novo *router* de acesso, de modo a que na altura da mudança efectiva as perdas de pacotes sejam mínimas ou mesmo inexistentes.

Os detalhes do fluxo de mensagens trocadas entre as diversas entidades bem como a integração com o sistema de QoS e de AAAC é apresentado mais à frente na secção 5.4, com a descrição das modificações realizadas para o suporte de QoS.

5.2.2 Mecanismos para a provisão de serviço num ambiente heterogéneo com mobilidade (AAAC)

5.2.2.1 Arquitectura geral

A arquitectura do sistema de AAAC apresentada na Figura 24 é baseada na arquitectura de AAA do IETF, enriquecida com auditoria, medição e taxação, e optimizada para um ambiente IPv6. O facto desta arquitectura ter como objectivo ser utilizada num ambiente heterogéneo de mobilidade com provisão de QoS foi considerado no seu desenvolvimento, de modo a ter características que permitam novas funcionalidades assim como uma optimização do desempenho de todo o sistema.

A função de auditoria permite ainda a existência de novas funcionalidades do sistema de AAAC, nomeadamente por permitir a avaliação de amostras de registos de AAAC gerados pelo próprio sistema de AAAC e por outras entidades. Neste contexto, um repositório de políticas foi considerado como parte integrante de um sistema de AAAC *policy-based* e pode assim ser incluído.

A Figura 24 apresenta a arquitectura do sistema AAAC. Este sistema suporta múltiplas interfaces. Um atendedor de AAAC lida com a interface para o terminal móvel: toda a comunicação é realizada usando um protocolo específico, o URP (*User Registration Protocol*). Um módulo ASM (*Application Specific Module*) é utilizado para as comunicações com o QoS *Broker*. A vantagem de desenvolver um ASM é a flexibilidade adicional que se ganha: uma grande variedade de equipamento de serviço poder ser coberta facilmente com um método uniforme do ponto de vista do sistema de AAAC. A comunicação entre o sistema de AAAC e o módulo ASM é realizada recorrendo ao protocolo DIAMETER de AAAC. A comunicação entre o módulo ASM e o equipamento de serviço pode ser efectuada por protocolos específicos do equipamento em causa (neste caso concreto por COPS ou URP).

Os vários sistemas de AAAC comunicam entre si utilizando como base o protocolo DIAMETER, devidamente enriquecido com extensões adequadas.

Nesta arquitectura há uma clara distinção entre os serviços oferecidos ao utilizador/cliente, tais como os serviços de rede com QoS (definidos na próxima secção), e os serviços necessários para a operação do sistema de AAAC, tal como a taxação. Assim, os primeiros são geralmente acedidos e fornecidos através do ASM e extensões do

protocolo de AAAC, enquanto que os últimos requerem uma comunicação directa com o sistema de AAAC, utilizando ligações dedicadas se tal for necessário.

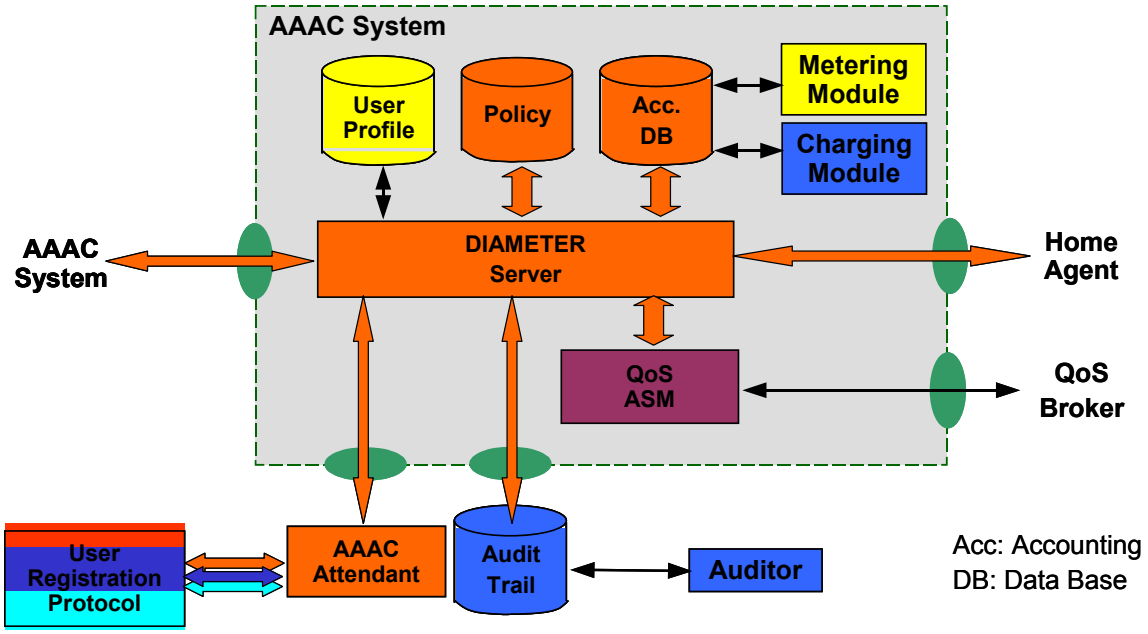


Figura 24: Arquitectura AAAC

5.2.2.2 Metering

O NeTraMet é utilizado para monitorização e medição de rede. O principal componente desta implementação de referência é o medidor de tráfego (agente NeTraMet) que captura informação a partir dos cabeçalhos dos pacotes IP, segundo os filtros e regras previamente configurados, e associa esta informação aos diferentes fluxos IP. A definição dos fluxos IP é bastante flexível e é feita de acordo com as necessidades do administrador da rede. No caso da arquitectura IPv6 apresentada, o medidor (melhorado de modo a suportar IPv6) é configurado de acordo com a descrição dos serviços subscritos pelos utilizadores e comunicados ao QoS *Broker*.

Um leitor/gestor de medição (NeMaC) comunica com o medidor de tráfego utilizando SNMP (*Simple Network Management Protocol*). Nesta arquitectura (Figura 25), o leitor de medição armazena numa base de dados de medição normalizada toda a informação recolhida. Através de uma interface específica, esta informação é disponibilizada ao cliente de AAAC localizado no *router* de acesso. Este, por seu turno, reenvia esta informação para a parte do sistema dedicada à taxação. Esta informação é também enviada ao QoS *Broker* que a usa para determinar o estado da rede.

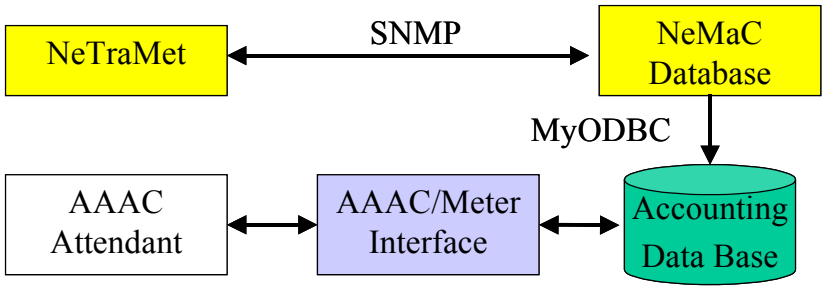


Figura 25: Arquitectura de medição

5.3 Arquitectura de QoS para provisão de serviços com suporte de mobilidade.

O sistema de QoS escolhido para suporte desta arquitectura de 4ª Geração deve suportar QoS extremo-a-extremo e deverá possuir características de simplicidade de gestão adequado a um ambiente de operador. Para atingir este objectivo, as entidades e os métodos desenvolvidos tiveram de ser definidos de modo a controlar e atribuir recursos nas redes de acesso, capazes de oferecer garantias de QoS extremo-a-extremo, mantendo a conectividade e QoS de cada utilizador enquanto o seu terminal comuta entre diferentes redes de acesso.

5.3.1 Arquitectura de QoS Global

Conforme já foi referido, um dos maiores requisitos e constrangimentos desta arquitectura é o suporte da mobilidade e de QoS em simultâneo.

Como vimos, a arquitectura de serviços integrados (IntServ) controla a reserva de recursos com a granularidade do fluxo, utilizando sinalização RSVP mas, não está preparada para suportar mobilidade e tem problemas bem conhecidos de complexidade e escalabilidade. Por seu turno, a arquitectura baseada no conceito de serviços diferenciados (DiffServ) resulta de uma aproximação que utiliza agregação de fluxos de serviços de acordo com as suas diferentes classes e prioridades. Apesar de esta última aproximação ser bastante escalável, não oferece garantias estritas de QoS e também não está dotada com mecanismos específicos de suporte de mobilidade. Outras técnicas mais complexas, como os cenários mistos de IntServ (no acesso) e DiffServ (no *core*), ou as aproximações NSIS (*Next Steps in Signalling*) [NSIS], também não parecem satisfazer simultaneamente o

suporte de mobilidade e garantir reservas extremo-a-extremo, especialmente num cenário envolvendo questões de AAAC.

Neste sentido foi desenvolvida uma utilização inovadora de QoS *Brokers* que realizam o controlo de admissão de serviços por utilizador, incorporando uma aproximação tradicional de DiffServ. A escolha de QoS *Brokers* para realizar o controlo de admissão deve-se à necessidade de obter uma grande escalabilidade e uma carga de sinalização reduzida. Esta aproximação permite o suporte de QoS em redes de larga escala e simultaneamente permite a optimização da utilização dos recursos de acesso (normalmente escassos e dispendiosos).

Esta arquitectura baseia-se no conceito de que os serviços serão disponibilizados de acordo com um contrato prévio entre o cliente e o fornecedor de serviço. Os QoS *Brokers* estão encarregues de gerir e atribuir recursos, por utilizador e por serviço, na rede de acesso. O fundamento básico no suporte de QoS extremo-a-extremo baseia-se no facto de que a negociação de QoS será feita ao nível da aplicação. A negociação de QoS poderá ser feita, por exemplo, recorrendo a SIP. Se o utilizador não está autorizado a usar um serviço de rede específico (ou porque o seu SLA não o permite ou porque o seu crédito acabou, nos serviços pré-pagos) ou se a rede não tem recursos disponíveis nesse instante, o utilizador receberá informação e poderá actuar em concordância.

Em termos de QoS, o contrato estabelecido entre o utilizador e o fornecedor de serviço baseia-se nos seguintes pressupostos:

1. Existe um perfil de QoS que é um subconjunto do SLA armazenado no sistema de AAAC.
2. Não são suportadas características de QoS relativas a atrasos, pelo menos de uma forma absoluta.
3. Não existe garantia determinística de parâmetros de QoS na rede. O fornecedor de serviço dimensionará e configurará a rede de modo apropriado, uma vez que este modelo não suporta reserva de recursos através de um caminho específico. Desta forma, os serviços são sinalizados através da utilização de um DSCP distinto por serviço, ou seja, cada DSCP corresponde a uma classe de serviço.
4. O terminal pedirá diferentes serviços utilizando DSCPs diferentes para cada um – um esquema de sinalização implícita.

5. De modo a usufruir de um serviço, o terminal marcará os pacotes de acordo com o serviço pretendido e o primeiro elemento da rede do operador implementará políticas baseadas no SLS desse utilizador específico. Como alternativa, o fornecedor de serviço poderá permitir a marcação pelo utilizador, baseado, por exemplo, no tipo de aplicação (porto a ser utilizado, por exemplo).

5.3.1.1 Definição do ambiente de fornecimento de QoS

Como foi referido anteriormente, esta arquitectura de rede baseia-se na utilização de mecanismos DiffServ associados a algoritmos de CAC (Controlo de Admissão de Chamadas) realizados pelos QoS *Brokers*. Antes de um utilizador poder usufruir dos recursos de rede, ele deverá passar por um processo de autenticação (só utilizadores com contrato válido estão autorizados a permanecer na rede) e para cada serviço pretendido deverá ainda passar por um outro processo de autorização de serviço. Nesta fase é efectuado o controlo de admissão de serviço que se baseia não só no perfil do utilizador, mas também num outro conjunto de factores dependentes do estado actual da rede. Se todos os requisitos forem cumpridos, isto é, se o utilizador tem um perfil válido para o tipo de serviço pretendido e se existem recursos suficientes na rede, o QoS *Broker* atribuirá os recursos correspondentes ao serviço em causa, permitindo ao utilizador usufruir do serviço pretendido. Após este processo concluído, os recursos serão mantidos pela infra-estrutura de gestão enquanto o utilizador se move entre diversas redes de acesso.

Esta arquitectura de QoS não suporta a noção de QoS extremo-a-extremo por fluxo, semelhante à noção de IntServ, principalmente devido aos problemas de escalabilidade que daí podem resultar. Contudo, com uma configuração e gestão de recursos adequadas, esta arquitectura poderá resultar numa aproximação extremo-a-extremo suficientemente genérica. Em termos de parâmetros de serviço, esta arquitectura será apenas limitada pelos constrangimentos associados à tecnologia de transporte (IP); portanto, características rígidas de parâmetros de tempo (atraso e variação de atraso) não são ainda suportados.

Este ambiente suporta a existência de vários domínios de QoS por cada domínio administrativo. Cada um dos domínios de QoS é administrado por um QoS *Broker*. Um QoS *Broker* é conceptualmente uma única entidade, no entanto poderá ser constituído por várias unidades distribuídas, actuando como uma só, gerindo e controlando os recursos associados ao seu domínio de QoS. A gestão e controlo de recursos será feita de acordo

com o modelo DiffServ, com PHBs (*Per Hop Behaviour*) diferentes, suportados através de provisão e controlo de acesso convenientes. Os diversos conceitos do DiffServ foram usados para definir as diferentes classes de serviço disponíveis e todos os pacotes de serviços oferecidos aos utilizadores são baseados em combinações destas classes de serviço.

Existe um SLS definido estaticamente para cada classe de serviço entre os diferentes domínios administrativos (e portanto diferentes domínios de QoS).

Cada utilizador tem um perfil associado onde está guardada a informação relativa a esse utilizador, incluindo o SLA e informação de AAAC. Do ponto de vista da rede, apenas uma pequena fracção dessa informação contém informação relevante, nomeadamente os serviços subscritos pelo utilizador e sua parametrização em termos de QoS. A essa fracção da informação chamamos *Network View of the User Profile*, ou NVUP. Na altura da autenticação, a NVUP é enviada pelo sistema de AAAC para o QoS *Broker* responsável pelo domínio de QoS onde o terminal se encontra no momento. Se o utilizador estiver num domínio que não o do fornecedor de serviço com que tem contrato (isto é, se o utilizador estiver em *roaming*), o sistema de AAAC externo, onde o utilizador está, fará um pedido directamente ao sistema de AAAC do domínio de origem para que a informação relativa àquele utilizador lhe seja fornecida. Após a recepção dessa informação, tudo funciona como se o utilizador estivesse no seu domínio local.

A reserva de recursos nesta arquitectura de QoS é efectuada de acordo com o serviço e este é sinalizado pelo DSCP. Um determinado DSCP poderá sinalizar um serviço unidireccional ou bidireccional que poderá ser simétrico ou assimétrico.

Da forma como é apresentada, esta arquitectura apresenta uma limitação óbvia: a quantidade de DSCPs está limitada pelo campo de 8 *bits* a ele destinado no pacote IPv6. Desta forma, o número de serviços distintos que podem ser disponibilizados são também limitados e com relativamente pouca granularidade, o que poderá levar a argumentar que não será uma arquitectura suficientemente flexível para ser adoptada como modelo futuro. No entanto esta limitação pode ser facilmente ultrapassada de duas formas distintas. Os pacotes IPv6 têm um campo chamado de "*flow label*" cuja aplicação não está ainda definida, tendo sido prevista para questões de QoS. Sendo assim, utilizando este campo como campo adicional para identificação de serviços, podem definir-se muitos milhares de serviços distintos. Adicionalmente, e para que a flexibilidade disponível para o operador

seja máxima, existe um DSCP específico, cujo objectivo é despoletar uma reserva de recursos "a pedido". O processo de obtenção de recursos utilizando este DSCP será mais demorado e, dada a flexibilidade disponibilizada, este serviço será mais dispendioso, logo apenas alguns utilizadores estarão dispostos a subscrevê-lo. No entanto constitui um mecanismo adequado para fornecimento de QoS completamente configurável.

5.3.1.2 Parâmetros de QoS

Esta arquitectura genérica não suporta a especificação estrita de todos os parâmetros de QoS, dadas as limitações inerentes à utilização do IP como tecnologia de transporte. Desta forma, apenas serão dadas garantias estritas na especificação da largura de banda e prioridade de cada serviço. Garantias estritas de atraso, variação de atraso e mesmo de perdas não são possíveis de dar em redes IP, onde não existe uma clara separação do tráfego de cada ligação e de acordo com a sua prioridade. Este tipo de garantias tem de ser imposto por métodos de engenharia de tráfego [Sarg03], e a sua implementação estará relacionada com os interesses comerciais do fornecedor e do utilizador (níveis de serviço, mercado, infraestruturas de rede disponíveis), que poderão ser aplicados sobre esta arquitectura de rede na altura do seu projecto comercial.

5.3.1.3 Classes de Serviço

Com o objectivo de demonstrar esta arquitectura, foram definidas várias classes de serviço distintas, que tentam cobrir uma grande variedade de serviços que podem ser disponibilizados neste tipo de redes, e sobre as tecnologias de acesso que esta rede visa cobrir [VMar02]. Uma das preocupações que se tiveram em conta na definição destas classes de serviço foi a de as tentar mapear directamente nas quatro classes de serviço definidas pelo ITU-T [ITUT] e adoptadas pelo UMTS Forum [UMTSF]: a *Conversational*, de *Streaming*, *Interactive* e de *Background* [UQoS00]. Estas classes são vistas como as classes de serviço tipo das futuras redes de telecomunicações.

Para cada classe de serviço, um ou mais serviços de rede foram definidos e são oferecidos aos utilizadores. De notar que os serviços aqui apresentados, e definidos na rede de demonstração referida, se tratam de serviços meramente exemplificativos e que pretendem demonstrar o conceito. A arquitectura é suficientemente genérica de modo a poder definir virtualmente qualquer tipo de serviço de rede que se pretenda oferecer.

A Tabela 7 apresenta os serviços de rede utilizados para demonstração da arquitectura e explora a sua relação com os parâmetros típicos do DiffServ. É também apresentado um possível mapeamento entre os serviços definidos e as classes de serviço definidas para o UMTS. Na tabela, o nome do serviço refere-se à referência sinalizada pelo utilizador, enquanto que a classe se refere ao PHB do DiffServ de suporte a este serviço. O campo prioridade relativa indica precisamente qual a prioridade que será dada aos pacotes do serviço em causa: EF (*Expedite Forwarding*) terá uma prioridade superior àquela oferecida às classes de serviços AF (*Assured Forwarding*). As classes AF1x, AF2 e AF4 são independentes entre elas e dentro da AF1x, a AF1 é a que tem maior prioridade, oposta à AF13 que tem a prioridade mais baixa. Na cauda da lista de prioridades vêm os serviços baseados nas classes de serviço BE (*Best Effort*). Os parâmetros de serviço indicam os parâmetros típicos utilizados na configuração dos *routers* de acesso para o serviço em causa, enquanto que a descrição do serviço indica o tipo de aplicação típica a que o serviço se destina. O serviço SIG, tal como o nome indica, é um serviço de sinalização, utilizado entre as diversas entidades na rede, não sendo por isso sujeito a controlo de admissão e faz parte do perfil de todos os utilizadores.

<table>
<tr><th colspan="2">Serviço</th><th rowspan="2">Prioridade Relativa</th><th rowspan="2">Parâmetros do Serviço</th><th rowspan="2">Descrição do serviço</th><th rowspan="2">Mapeamento para as classes de serviço UMTS</th></tr>
<tr><th>Nome</th><th>Classe</th></tr>
<tr><td>S1</td><td>EF</td><td>1</td><td>Largura de banda de pico: 32 Kbps</td><td>Serviços de tempo real</td><td>Conversational</td></tr>
<tr><td>SIG</td><td>AF41</td><td>2a</td><td>Não especificado</td><td>Sinalização</td><td></td></tr>
<tr><td>S2</td><td>AF21</td><td>2b</td><td>CIR: 256 Kbps</td><td>Transferência de informação prioritária (urgente)</td><td rowspan="2">Streaming and Interactive</td></tr>
<tr><td>S3</td><td>AF1*</td><td>2c</td><td>Três precedências de descarte (Kbps):
AF11 – 64
AF12 – 128
AF13 – 256</td><td>Serviço olímpico (melhor que “melhor esforço: streaming, ftp, etc)</td></tr>
<tr><td>S4</td><td>BE</td><td>3</td><td>Taxa de pico: 32 Kbps</td><td>Melhor esforço – Best Effort (BE)</td><td rowspan="3">Background</td></tr>
<tr><td>S5</td><td>BE</td><td>3</td><td>Taxa de pico: 64 Kbps</td><td>Melhor esforço</td></tr>
<tr><td>S6</td><td>BE</td><td>3</td><td>Taxa de pico: 256 Kbps</td><td>Melhor esforço</td></tr>
<tr><td>S7</td><td colspan="4">Serviço especial que despoleta uma negociação de parâmetros com a rede</td><td></td></tr>
</table>

Tabela 7: Serviços alvo para efeitos de demonstração

Cada um dos serviços constantes da tabela pode ser subscrito por qualquer utilizador e poderão ser todos utilizados em simultâneo por aplicações diferentes. Por outras palavras, a NVUP de cada utilizador será composta por uma combinação destes serviços e cada pacote enviado pelo utilizador sinalizará qual a classe de serviço em que deve ser colocado. O fornecedor de serviço pode fazer conjuntos de serviços distintos e comercializá-los como pacotes de serviços. Como exemplo podemos ter o pacote "*extra-light*", composto pelos serviços S1 e S4 e o pacote "exclusive", composto pelos serviços S1, S2, S3 e S6. O SLA de um utilizador pode ser também do tipo pré- ou pós pago, podendo alguns detalhes específicos ser resolvidos a um nível administrativo.

5.3.1.4 Gestão e controlo de QoS: QoS Brokers

O QoS *Broker* é a entidade chave de toda a gestão e controlo de recursos nesta arquitectura de QoS [VMar03a]. Esta entidade é a responsável pela alocação e libertação de recursos a cada um dos utilizadores/serviço. A libertação de recursos é efectuada após existir um *time-out* de não utilização do serviço pelo utilizador, ou de forma explícita pelo utilizador. O utilizador terá que ser validado para aceder à rede antes que o QoS *Broker* possa dar permissões de acesso aos recursos. Neste sentido, a primeira entidade com que o utilizador dialoga é o sistema de AAAC. Após uma autenticação correcta, o QoS *Broker* receberá do sistema de AAAC a informação relevante relativa ao utilizador em causa (NVUP).

O QoS *Broker* tem informação acerca de todas as ligações e todos os elementos de rede do seu domínio de competência. Este domínio é-lhe conferido de acordo com as políticas de administração da rede. É ainda responsabilidade do QoS *Broker* a distribuição dos recursos pelas diferentes classes de serviço. O QoS *Broker* faz a gestão de recursos baseado na informação de alocação e libertação de recursos, assim como de informação recolhida da rede em tempo real. Desta forma, o QoS *Broker* mantém bases de dados com informação do estado da rede e das reservas efectuadas. Um novo utilizador/serviço será admitido baseado no seu perfil e no estado real da rede, em especial nos recursos existentes da classe de serviço que está a ser pedida.

O tipo de SLA que cada utilizador possui pode também ser um factor determinante para decidir se esse utilizador pode ou não entrar numa célula que, em princípio, não tem recursos suficientes para acolher os seus serviços activos. Isto é, um utilizador com uma prioridade mais elevada pode forçar o QoS *Broker* a degradar (ou em casos extremos

cancelar e rejeitar) os serviços de outros utilizadores com menos prioridade que já estejam nessa célula. Este tipo de controlo e gestão detalhada é possível apenas devido à natureza de gestão localmente centralizada imposta pelo QoS *Broker*. De notar que os QoS *Brokers* podem também fazer a administração dos recursos, por agregado, no *core* das redes. Desta forma, e havendo uma coordenação centralizada num ambiente com vários QoS *Brokers* distribuídos (cada um gerindo um pequena parte da rede), é possível, dentro de cada domínio administrativo, garantir a QoS de cada fluxo.

5.3.1.5 Interacções com a mobilidade dos terminais

A QoS em ambientes móveis é um aspecto crítico endereçado nesta arquitectura. Para lidar com esta problemática e minimizar os problemas, foram utilizadas técnicas inovadoras envolvendo *Fast Handovers* e transferência de contexto.

O QoS *Broker* é também a entidade que do ponto de vista da QoS é responsável por manter informação sobre a localização do terminal e assegurar a existência de recursos nas células para onde ele se move. Isto é conseguido recorrendo a técnicas de transferência de contexto rápidas, no instante em que o terminal indica a intenção de mudar de célula.

Como já foi referido atrás, esta arquitectura não está neste momento preparada para permitir a renegociação de parâmetros de QoS durante uma sessão, mas é uma funcionalidade que será facilmente introduzida, dada a flexibilidade da arquitectura. O resultado prático desta limitação actual é o facto de que durante um *handover* pode acontecer não existirem recursos para suportar todos os serviços activos. Nesta situação, uma de três situações poderão ocorrer: (i) o utilizador/terminal é informado do facto, decide terminar o *handover* e tenta fazer o *handover* para outra célula, (ii) a sessão é terminada, ou (iii) a prioridade do utilizador é superior à de alguns dos utilizadores já presentes na célula, e estes vêem o seu serviço degradado para acolher o novo terminal e respectivos serviços.

Quando existem várias células para onde o terminal se pode mover, o custo dos recursos das diferentes tecnologias poderá ser um factor de decisão, isto é, caso o terminal se possa mover quer para uma célula de TD-CDMA ou para uma célula de *Wireless* LAN, se os recursos da célula de *Wireless* LAN forem menos dispendiosos, o terminal fará o *handover* para essa célula. Ou seja, o terminal poderá executar um *handover* quer por degradação da ligação existente, quer por motivos económicos. Desta forma, a célula

destino poderá ser escolhida dependendo quer da disponibilidade de recursos, quer do seu custo. Este tipo de considerações não foi implementado no demonstrador da arquitectura.

5.3.1.6 Interacções com o sistema de AAAC

O sistema de gestão e controlo de QoS (QoS *Broker*) e o sistema de AAAC estão intimamente ligados [Cuev03]. O sistema de AAAC armazena a informação acerca do SLA do utilizador e é o responsável pelo envio da informação de QoS relevante, a NVUP, para o sistema de QoS para que este possa agir em concordância com ela e fornecer recursos ao utilizador de acordo com o que o utilizador paga.

Do ponto de vista do sistema de AAAC, a informação relativa a QoS que ele armazena, é tratada exactamente da mesma forma que toda a outra informação.

O sistema de AAAC não tem a noção do que é QoS nem do que são serviços IP. Estas funções são exclusivas do sistema de QoS.

Em termos de topologia de rede, a primeira entidade que é contactada pelo terminal móvel é sempre o *router* de acesso que realizará funções específicas relacionadas com o sistema de AAAC (através do atendedor de AAAC), como a autenticação, e funções específicas de QoS (através do atendedor de QoS) como a autorização de serviço e atribuição de recursos (actuando como PEP, sendo o QoS *Broker* o PDP).

5.3.2 Blocos funcionais da arquitectura de QoS

A Figura 26, a Figura 27 e a Figura 30 apresentam os blocos constituintes dos elementos mais relevantes da arquitectura de QoS: o terminal móvel, o *router* de acesso, o QoS *Broker* e as principais funções relacionadas com QoS presentes em cada um [Beau03].

O funcionamento destes blocos será descrito mais detalhadamente na secção 5.4.

5.3.2.1 Terminal Móvel

Ao nível mais elevado das camadas protocolares do terminal móvel, encontra-se o *Networking Control Panel* (NCP) cujas funções são as de registo e de-registo na rede (relacionado com as funções de autenticação). Ao nível da rede existe uma camada de IPv6 melhorada que inclui funcionalidades extra em relação ao módulo base de mobilidade para permitir *handovers* rápidos e com suporte de QoS. Esta camada contém também o módulo de marcação de DSCP, que tem como funções marcar o tráfego de saída com o código

correcto de DSCP para o serviço de QoS pretendido. Este software pode também ser configurado de modo a enviar pacotes sem informação, apenas para despoletar a reserva de recursos na rede. Ainda incluído neste *software* está o suporte de filtros para IPv6, ICMPv6 e para os cabeçalhos de transporte (TCP e UDP) e que pode ser configurado de acordo com as políticas definidas pelo utilizador/administrador.

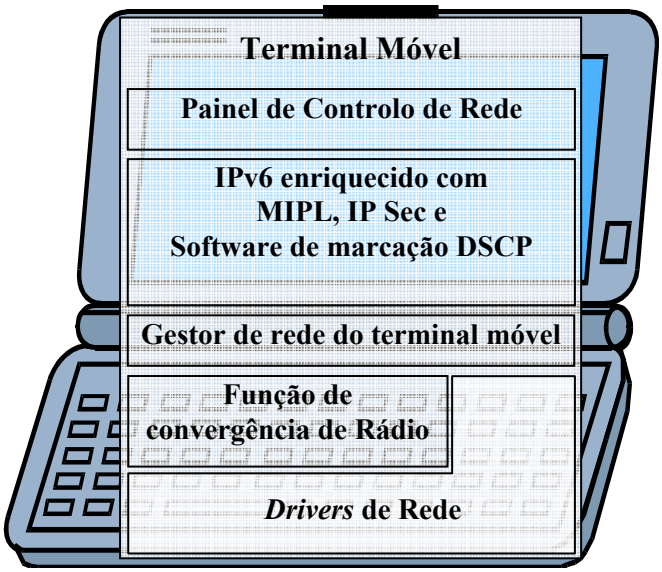


Figura 26: Componentes do terminal móvel

Entre a camada IP e o nível 2, existe o Gestor de Rede do Terminal Móvel (*Mobile Terminal Network Manager* – MTNM) que toma as decisões de execução de *handover* de acordo com as preferências do utilizador (incluindo a QoS pretendida) e com a informação recebida dos elementos de rede. Finalmente, a Função de Convergência de Rádio (*Radio Convergence Function* – RCF) actua como um *driver* providenciando as *interfaces* de nível mais baixo para o controlo do TD-CDMA de modo a abrir e fechar ligações, gerir os canais de dados e enviar e receber pacotes de acordo com os requisitos de QoS do nível IP.

5.3.2.2 Router de acesso

Nesta arquitectura, o *router* de acesso executa funções típicas de um nó fronteira (*edge*) de uma topologia DiffServ: policiamento, formatação de tráfego e escalonamento e descarte de pacotes implementados de acordo com os DSCPs (PHBs – *Per Hop Behaviours*). O atendedor de QoS permite ao *router* delegar no QoS *Broker* as decisões de controlo de admissão, seguindo o modelo COPS (*Common Open Policy Service*). Neste cenário, o *router* de acesso actua como ponto de aplicação de políticas (PEP) e o QoS

Broker como ponto de decisão de políticas (PDP). O *router* de acesso monitoriza o tráfego que passa das interfaces de acesso para as do *core*, recolhe os parâmetros relevantes e constrói o pedido para o QoS *Broker* de acordo com eles. Na implementação da arquitectura foi desenvolvida uma biblioteca para dar acesso a funções COPS tanto ao atendedor de QoS como ao QoS *Broker*.

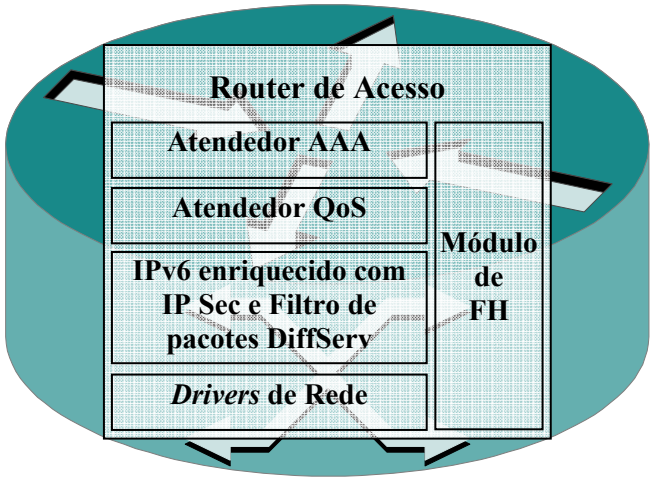


Figura 27: Componentes do *router* de acesso

Adicionalmente, o *router* de acesso também executa funções relativas ao processo de *handover* rápido (FHO). Através do módulo de *Fast Handover* (FHm), o *router* de acesso onde o terminal está presentemente (*router* antigo), notifica o QoS *Broker* e também aquele que será o próximo *router* de acesso (*router* novo), da preparação do FHO. Se o QoS *Broker* decidir que o *handover* pode ser realizado, configurará o *router* novo e este, através do FHm notificará o *router* antigo para que este por sua vez avise o terminal de que pode efectuar o *handover*.

Os PHBs executados pelo *router* de acesso estão disponíveis nas interfaces de acesso para todos os pacotes que viajam do *core* para o acesso, que podem ser ligações sem fios de recursos escassos. A gestão de filas e os mecanismos de escalonamento tais como o RIO *(Random Early Detection – RED – with In and Out*) ou o WFQ (*Weighted Fair Queuing*) do DiffServ são utilizados para a implementação dos PHBs. A API TC (*Traffic Control*) [TCAPI] da IBM foi utilizada para integrar estas técnicas de aprovisionamento de QoS. As garantias de QoS são conseguidas pela cooperação entre a imposição de PHBs no *router* de acesso e a gestão de recursos efectuada pelo QoS *Broker*, seguindo o modelo COPS. Entre as diversas funções do *router* de acesso, está também o dever de informar o QoS *Broker* da carga das suas diferentes filas, de modo a que este possa fazer alterações na configuração e

adaptá-las às necessidades reais. Por exemplo, numa noite de 24 de Dezembro, o tráfego será maioritariamente de voz (que corresponde a uma classe EF em termos de DiffServ) e haverá muito pouco tráfego de outras classes. Nessa situação, poderão ser reconfiguradas as filas de espera de modo a haver uma melhor adaptação ao tráfego existente.

5.3.2.3 Suporte de QoS sobre TD-CDMA

Conforme já foi referido, esta arquitectura tem potencialidades para o suporte de QoS independente da tecnologia de nível 2 subjacente. É no entanto, como facilmente se compreende, necessário preparar e adaptar o nível 2 de cada uma dessas tecnologias ao IP. A Figura 28 apresenta, como exemplo, os componentes dos terminais móveis e *routers* de acesso envolvidos na gestão e controlo dos recursos rádio TD-CDMA, uma das tecnologias onde esta tarefa é mais complexa. Todas as funções de *Radio Resource Management* (RRM) localizadas no *core* das redes UMTS foram excluídas, uma vez que esta arquitectura apresenta uma aproximação de sinalização "*True-IP*".

Os procedimentos de controlo de QoS IP, baseados nos DSCPs DiffServ são directamente mapeados no nível físico. Isto é, de modo a ter QoS extemo-a-extremo torna-se necessário mapear as classes de serviço de QoS IP, assinaladas pelos DSCPs, em classes de serviço rádio do UMTS, definidas nas normas do 3GPP. Assim, estes parâmetros devem ser mapeados num conjunto de parâmetros rádio de modo a assegurar a operação correcta da *interface* TD-CDMA [Figura 29].

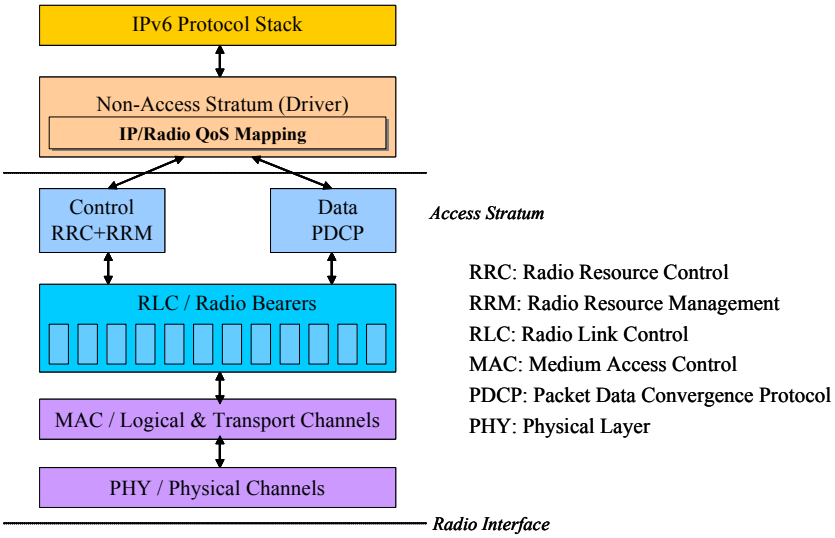


Figura 28: Arquitectura de QoS em TD-CDMA

Os parâmetros, tais como a largura de banda, atraso e taxa de perda de pacotes devem ser convertidos num determinado número de *slots* de tempo, formatos de transporte válidos, códigos de convolução, valores e TTI (*Transmission Time Interval*), etc. Esta conversão é realizada em dois passos. O primeiro consiste na definição das classes de QoS e dos seus parâmetros de modo a que os utilizadores tenham acesso a uma QoS similar àquela que o IP lhes dá. Por exemplo, um serviço de tempo real (S1 na Tabela 7, por exemplo) pode ser convertido numa classe de serviço Conversacional no mundo rádio. O segundo passo é executado quando o serviço começa, com a computação dos parâmetros rádio finais, baseados nas classes QoS rádio, na configuração da célula rádio e nos recursos previamente atribuídos.

Com esta técnica, estes recursos são também controlados pelos QoS *Brokers*. Quando o utilizador inicia um novo serviço, o QoS *Broker* mapeia os níveis e parâmetros de serviço em classes de serviço rádio e envia um pedido à componente de mapeamento de QoS, de modo a que esta atribua os recursos correspondentes. Este módulo guarda a informação de mapeamento e reenvia o pedido para o módulo *Radio Resource Control* (RRC) dos protocolos de *interface* rádio constituintes do *Access Stratum*.

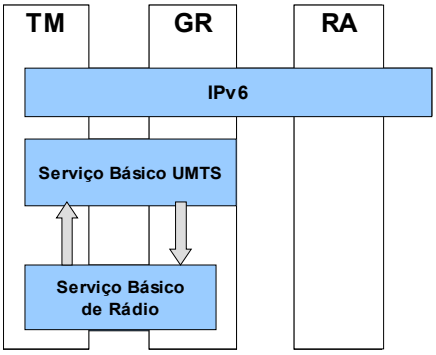


Figura 29: Mapeamento entre os serviços do mundo IPv6 e o serviços do TD-CDMA

Com a ajuda de um novo motor RRM, o RRC determina as alterações necessárias dos parâmetros de rádio necessários quer no *router* de acesso quer no terminal móvel de modo a assegurar um funcionamento correcto dos outros protocolos da interface de rádio apresentados na Figura 28: o PDCP (*Packet Data Convergence Protocol*), o RLC (*Radio Link Control*), o MAC e o PHY (*Radio Link Control*). O cálculo destes parâmetros rádio é baseado simultaneamente na QoS requerida (mapeada dos códigos DSCP), nas configurações existentes, e em recursos previamente atribuídos e em utilização. A *gateway* rádio guarda a informação de mapeamento e abre uma nova *bearer* rádio ligando-se ao

terminal móvel através dos vários módulos constituintes da *interface* rádio. A activação destes parâmetros abre um novo canal rádio, usualmente mapeado num canal físico e de transporte lógico dedicado.

Tanto no terminal móvel como no *router* de acesso existe uma função de QoS de nível 2 responsável pelo mapeamento dos parâmetros de QoS IP em rádio *bearers*. Logo que este rádio *bearer* esteja disponível, o *Non Access Stratum* pode utilizá-lo para transferir a informação através do PDCP e camadas inferiores, usando o mapeamento previamente guardado como chave para o SAP (*Service Access Point*). Quando o serviço deve ser terminado, o QoS *Broker* fecha os canais de forma similar.

A abordagem feita relativamente à tecnologia TD-CDMA e à simplificação da arquitectura UMTS trouxe a vantagem (já mencionada) de possibilitar a gestão desta tecnologia recorrendo a mecanismos IP. A desvantagem associada a esta abordagem é claramente a disrupção em relação às redes e aos serviços tradicionais (*legacy*) das redes de gerações anteriores, como é o caso dos serviços de comutação de circuitos, e a necessidade de melhoramentos na tecnologia IP (nomeadamente as modificações associadas a mobilidade rápida integrada com QoS). Contudo, esta arquitectura foi projectada desde início para ser uma arquitectura de 4G, sem qualquer compromisso de compatibilidade com as arquitecturas tradicionais.

5.3.2.4 QoS Broker

Esta entidade complexa mas flexível concentra sobre si a maioria das decisões de controlo neste ambiente heterogéneo. O QoS *Broker* monitoriza as extremidades da rede detectando e atendendo pedidos de reserva de entrada e de saída. A arquitectura interna do QoS *Broker* está representada na Figura 30 e será descrita seguidamente.

O ponto fulcral do QoS *Broker* é o motor que inclui todos os algoritmos de decisão para a gestão de QoS da rede. Este motor opera numa camada de abstracção dos *routers* de acesso. Um módulo de *router* virtual fornece uma *interface* única entre os *routers* e o motor, e mapeia as decisões de controlo nos comandos específicos de cada *router*, independente da sua *interface* específica e mesmo do seu modelo. O motor faz uso de uma base de dados, chamada NetworkDB, onde se encontra a informação de todos os *routers* do seu domínio, incluindo o formato de comandos que se deve aplicar a cada um deles. É também na base de dados NetworkDB que estão guardadas as diferentes acções de configuração que devem ser utilizadas para atingir determinado fim.

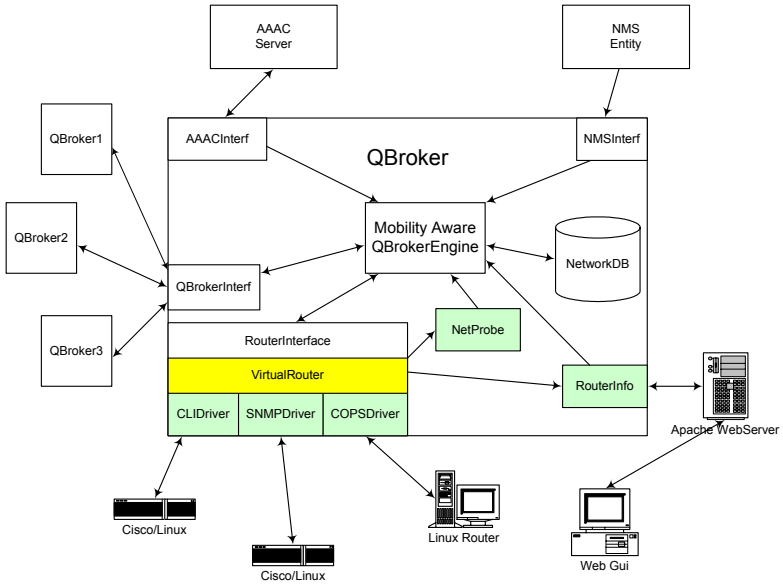


Figura 30: Componentes do QoS *Broker* e respectivas interfaces

O QoS *Broker* pode comunicar com os *routers* de acesso utilizando vários protocolos, cada um implementado por um *driver* distinto: (i) CLIDriver é uma API de *login* remoto desenvolvido para permitir a configuração remota dos *routers* de acesso; (ii) COPSDriver é o *driver* utilizado para configurar os *routers* recorrendo ao protocolo COPS; (iii) SNMPDriver é o *driver* que permite o acesso do QoS *Broker* às funções disponibilizadas pelas MIBs SNMP de cada *router*. Assim, o motor do QoS *Broker* lê as instruções acerca dos comandos que serão utilizados, assim como outra informação adicional (como por exemplo a MIB ou o *login* e *password* de gestão remota).

O QoS *Broker* incorpora outras *interfaces*. Uma *interface* para o sistema de AAAC é usada para enviar e receber informação específica de cada utilizador (a NVUP) durante a fase de registo e autenticação. Esta interface é também usada para que as políticas de domínio administrativo sejam passadas a todos os QoS *Brokers* (tais como os códigos DSCP utilizados no domínio para cada serviço, e as características dos serviços propriamente ditos). Uma *interface* para outros QoS *Brokers* está também disponível e é utilizada para troca de informação entre os vários QoS *Brokers* do mesmo domínio, para fornecer QoS extremo-a-extremo. Para a gestão de mobilidade existe uma interface de que serve para fazer a transferência de contexto dos utilizadores em movimento entre células administrados por QoS *Brokers* distintos. Na realidade esta é uma interface entre QoS *Brokers*, semelhante à que foi descrita anteriormente, mas conceptualmente com funções distintas.

Os QoS *Brokers* têm ainda três módulos distintos dedicados a aspectos de controlo e monitorização da rede:

1. NetProbe: monitoriza o estado da rede e armazena-o na base de dados NetStatus (que será descrita nesta secção).
2. *Router*Info: módulo responsável pela obtenção de informação relacionada com as capacidades dos *routers*. Esta informação pode ser obtida de uma forma automática ou manual.
3. NMSInterface: permite a um sistema de gestão global definir quais os recursos de rede que serão administrados por este QoS *Broker*.

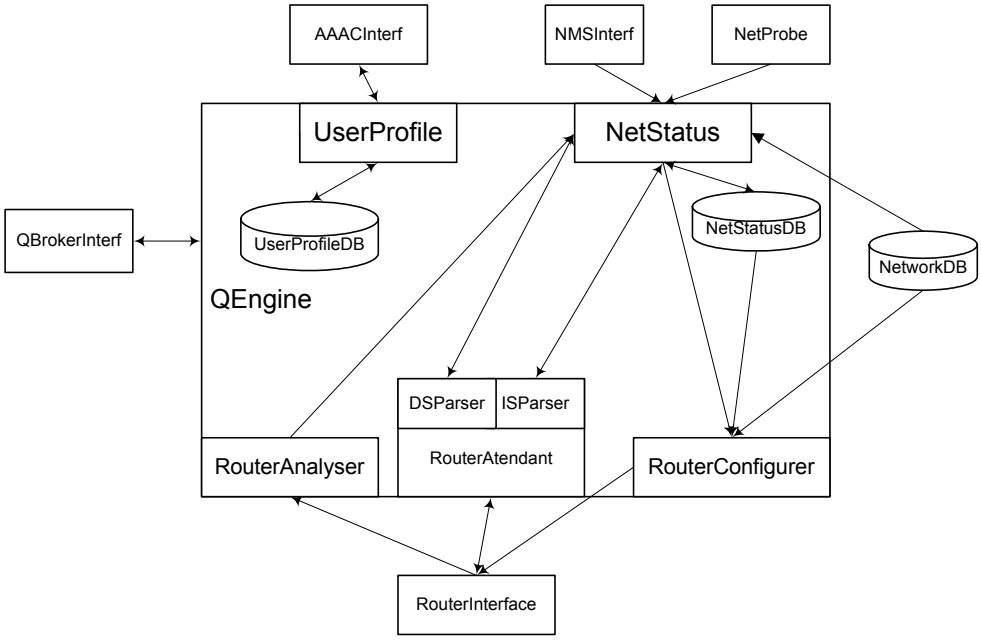


Figura 31: Motor do QoS Broker

O motor do QoS *Broker* (Figura 31), cujas funções e competências foram anteriormente descritas, é composto pelos seguintes blocos funcionais:

- UserProfile: gere toda a informação relacionada com os perfis de QoS definidos pelo sistema de AAAC assim como a informação relativa a cada utilizador. A base de dados UserProfileDB (que será descrita nesta secção) é gerida por este módulo.
- NetStatus: (i) gere toda a informação relacionada com a autorização de serviço e com a carga dos dispositivos de rede; (ii) toma decisões de controlo de admissão de serviço, recebendo e atendendo os pedidos de serviço; (iii)

implementa uma base de dados interna, descrita posteriormente nesta secção, denominada NetStatusDB.

- *Router*Attendant: recebe todos os pedidos provenientes do *router* de acesso e filtra-os antes de os enviar para o módulo NetStatus. De entre as diversas funcionalidades, destaca-se a capacidade de interpretar pedidos COPS.
- *Router*Analyser: é a entidade utilizada pelo módulo NetStatus que verifica a carga dos dispositivos de rede.
- *Router*Configurer: é a entidade que executa as acções de configuração dos elementos de rede. Utiliza a informação dos dispositivos de rede mantida na NetworkDB para determinar qual o protocolo (CLI - *Command Line Interface*, COPS, SNMP) e os detalhes de configuração (de acordo com o modelo e fabricante) que deverá utilizar para cada elemento e para a configuração pretendida.

Em termos de bases de dados, o QoS *Broker* inclui as seguintes:

- NetworkDB: conforme já foi descrito anteriormente, esta base de dados (i) mantém a informação que descreve a topologia do domínio de QoS do QoS *Broker*, (ii) inclui a informação de cada um dos elementos de rede, as suas características e *interfaces*, (iii) e a informação necessária para configurar cada um deles.
- NetStatusDB: esta é uma base de dados pequena contendo apenas uma pequena parte da informação da NetworkDB. É utilizada para que as decisões de controlo de admissão de serviço sejam tomadas mais rapidamente.
- UserProfileDB: mantém toda a informação que descreve os serviços e os seus parâmetros de QoS. É também aqui que são guardadas as NVUPs de todos os utilizadores presentes no domínio de QoS, assim como a informação dos serviços actualmente subscritos pelos utilizadores.

5.3.3 Interfaces e mensagens trocadas entre o QoS Broker e os restantes elementos

5.3.3.1 Interface entre o QoS Broker e o sistema de AAAC

É através desta interface, baseada em COPS, que as políticas de serviço do operador são difundidas para os diversos QoS *Brokers* e desta forma implementadas na rede. É também através desta interface que o QoS *Broker* é informado acerca do perfil dos utilizadores (NVUP) que estão no seu domínio e pelos quais ele é responsável ao nível do controlo de admissão.

Existem cinco mensagens distintas trocadas entre os QoS *Brokers* e o sistema de AAAC:

1. Descrição de serviços: sempre que um novo QoS *Broker* seja instalado ou reiniciado ou sempre que exista uma alteração da política de serviços do operador, existirá uma mensagem enviada pelo sistema de AAAC para o(s) QoS *Broker*(s) em questão, com a descrição dos serviços (*Services Description*) oferecidos pelo operador. Esta mensagem identifica as características dos serviços oferecidos em termos de largura de banda e de prioridade e quais os códigos DSCP utilizados para sinalizar cada um deles. Esta mensagem pode ser vista como uma mensagem de configuração inicial do QoS *Broker*.
2. Autorização de perfil: sempre que um novo utilizador se regista num domínio, o sistema de AAAC informa o QoS *Broker* responsável pelo área onde o utilizador se encontra, enviando-lhe a NVUP desse utilizador. Desta forma, o QoS *Broker* fica habilitado a realizar o controlo de admissão dos serviços desse utilizador.
3. Anulação de perfil: sempre que o utilizador abandone o domínio, ou que por qualquer outro motivo o seu perfil deixe de ser válido (por exemplo falta de pagamento), o sistema de AAAC notifica o QoS *Broker* para que este cancele serviços que estejam a decorrer e, eventualmente, pare de servir esse utilizador.
4. Pedido de validação de NVUP: o QoS *Broker* pode, em qualquer instante, fazer um pedido ao sistema de AAAC para a validação da NVUP de um

determinado utilizador. Isto é especialmente útil em casos de serviços pré-pagos, em que o QoS *Broker* necessita saber se o utilizador ainda tem crédito.

5. Validação de NVUP: esta é a resposta do sistema de AAAC para o QoS *Broker*, indicando se determinado utilizador pode ou não continuar a ser servido.

O conjunto das mensagens anteriores utiliza as seguintes estruturas de dados:

- Descrição de Serviços:
  - *DSCP code*: código DSCP que identifica o serviço;
  - *Bandwidth*: largura de banda do serviço;
  - *Priority*: prioridade dos pacotes do serviço;
  - *Delay*: atraso máximo que os pacotes deste serviço poderão sofrer (não está ainda implementada a forma de garantir este parâmetro);
  - *Destination Address*: endereço destino dos pacotes (utilizado em especial nos casos de chamadas para números "bem-conhecidos", como por exemplo chamadas de emergência, serviços do tipo "800", ou similares).
- NVUP:
  - *User ID*: identificação do utilizador;
  - *Care-of-address*: corresponde ao endereço de origem em cada momento;
  - *N services*: número de serviços que em cada momento o utilizador está autorizado a utilizar;
  - Lista de N elementos com *NetService* - que descreve um serviço de rede em termos de:
    - Endereço destino: usado em casos especiais tais como número de emergência, "800", etc.;
    - Validade da autorização: tempo de validade do serviço ao fim do qual o QoS *Broker* deverá fazer um pedido de validação;
    - Código DSCP: identificador do serviço. O QoS *Broker* utiliza este valor para procurar os parâmetros associados ao serviço na descrição dos serviços recebida do sistema de AAAC.

5.3.3.2 Interface entre o QoS Broker e os routers de acesso

Esta interface, também baseada em COPS, é utilizada para configuração e reconfiguração dos *routers* de acesso, assim como para fazer o controlo de admissão de serviços de cada utilizador.

Existem cinco mensagens distintas trocadas através desta interface:

1. Pedido de configuração: esta mensagem é enviada ao QoS *Broker* pelos *routers* de acesso na altura em que são (re)colocados ao serviço. Com base neste pedido, o QoS *Broker* irá dar indicação (via COPS) de como os *routers* de acesso devem configurar as filas de espera, classes de serviço e filtros.
2. Alteração de configuração: esta mensagem é a mensagem de resposta do QoS *Broker*, com a indicação do modo como o *router* de acesso deve ser configurado a partir daquele instante.
3. Pedido de recursos: sempre que um utilizador tenta utilizar um novo serviço, o *router* de acesso utiliza esta mensagem para fazer um pedido de autorização ao QoS *Broker*.
4. Libertação de recursos: logo que o utilizador deixa de utilizar determinado serviço, o *router* de acesso informa o QoS *Broker* que os recursos previamente alocados estão de novo disponíveis.
5. Pedido de *handover*: o *router* de acesso sinaliza desta forma a intenção que determinado terminal mostrou em fazer um *handover*. O QoS *Broker* processará esse pedido com base na informação recebida.

O conjunto das mensagens anteriores utiliza as seguintes estruturas de dados:

- Configuração dos *routers*:
  - *DSCP:* código DSCP da classe de serviço associado a cada fila;
  - *BW*: largura de banda do fluxo de pacotes reservado à aquela fila;
  - *Borrow_flag*: flag indicativa se fila deve ou não ceder tempo de serviço a outras filas;
- Pedido de recurso:
  - Endereço origem: endereço origem do fluxo;
  - Endereço destino: endereço destino do fluxo;

- Código DSCP de sinalização: código DSCP que define o serviço requisitado.

- Pedido de *handover*:
  - oCoA: antigo CoA do MT;
  - nCoA: novo CoA do MT;
  - nAR: Endereço do novo RA (*Router* de Acesso) para qual o MT pretende mudar.

5.3.3.3 Interface entre QoS Brokers

Esta interface, actualmente é utilizada apenas para a realização de *handovers* de utilizadores de um domínio de QoS para outro. No futuro, esta interface poderá ser também utilizada para efectuar uma gestão optimizada de recursos. Neste momento apenas uma mensagem é trocada nesta interface, a mensagem de pedido de *handover*. A sua estrutura de dados é a seguinte:

- Pedido de *handover*:
  - nCoA: é o novo CoA construído pelo MT através da informação recebida do *router* de acesso para onde o terminal deseja mover-se;
  - NVUP: o perfil do utilizador;
  - Lista de serviços: lista de serviços actualmente em uso pelo MT. O QoS *Broker* vai usar esta informação para configurar o *router* de acesso para onde o terminal deseja mover-se, de modo a que todos os recursos que o terminal necessita estejam já assegurados na altura da mudança.

Esta interface não é utilizada caso o *handover* se realize entre dois *routers* de acesso controlados pelo menos QoS *Broker*.

5.3.3.4 Interface entre o QoS Broker e a Radio Gateway

Na arquitectura proposta, a cada *radio gateway* está associado um *router* de acesso. Ou seja, a *radio gateway* não representa mais do que o acesso físico TD-CDMA. É necessário que o QoS *Broker* tenha controlo sobre os recursos rádio deste elemento. Esta interface é então utilizada para que possam ser abertos e mantidos canais de e para terminais móveis, situados na área de cobertura de uma *radio gateway*. O fecho dos canais dá-se sempre que não exista uma renovação dos mesmos. Assim, a única mensagem

enviada do QoS *Broker* para a *radio gateway* é a mensagem de abertura de canal, cuja estrutura de dados é apresentada de seguida:

- Abertura de canal:
    - nb_entries: define o número de elementos qosb_radio_ bearer_info_t (definidos a seguir) enviados;
    - home_Addr: endereço do MT ligado à RG;
    - nb_entries x qosb_radio_ bearer_info_t: lista de elementos que definem os parâmetros do canal a ser aberto.

Em que cada qosb_radio_ bearer_info_t é composto pelos dados seguintes:

- dscp: código dscp do fluxo;
- radio_QdS_class: código da classe rádio do canal;
- status: indica se a mensagem é enviada para abertura de um novo canal por um fluxo já existente, ou se se trata de um pedido de abertura de canal para um terminal que irá fazer um *handover* para a rede da RG.

5.3.3.5 Interface entre o QoS Broker e o sistema de gestão (NMS)

Apesar desta ser uma interface que no futuro incluirá diversas funcionalidades de gestão avançada de recursos e de controlo de QoS extremo-a-extremo, neste momento é apenas utilizada para que os QoS *Brokers* sejam informados dos recursos existentes entre os seus domínios de QoS e o *core* da rede. Existe portanto apenas uma mensagem definida, a mensagem de envio de configuração de recursos, cuja estrutura de dados é apresentada de seguida:

- Configuração de recursos:
    - *UpstreamBW*: largura de banda disponível na entrada da rede de *core*;
    - *DownstreamBW*: largura de banda máxima que poderá ser recebida a partir do *core*;
    - *Delay / Jitter*: Atraso / Variação do atraso que os pacotes podem experimentar no *core*;
    - *Errors*: taxa de erros a que os fluxo podem ser sujeitos no *core*;

5.4 Cenários Chave de integração de mobilidade, QoS e AAAC

Nesta arquitectura, existem três fases de operação e controlo que podem ser facilmente identificáveis e justificáveis como sendo as mais importantes: i) registo – nesta arquitectura, um utilizador por via do seu terminal móvel, apenas poderá usufruir dos recursos de rede após uma identificação correcta e válida, à semelhança das redes celulares; ii) autorização – o utilizador tem de estar autorizado antes de poder usar um serviço com determinadas características de rede que conste do seu perfil (serviços que não estejam incluídos não podem ser utilizados); e iii) *handover* – o utilizador necessita que os recursos que lhe estão atribuídos num *router* de acesso lhe sejam transferidos para o *router* de acesso para onde ele se irá mover.

5.4.1 Registo

A fase de registo é iniciada após o terminal móvel obter um *care-of-address* (CoA) através de auto-configuração *stateless*, utilizando os identificadores de *hardware* (globalmente únicos) de modo a evitar situações de duplicação de endereços. No entanto, a obtenção de um endereço não é só por si o suficiente para permitir ao utilizador o acesso aos recursos da rede. Nesta fase, antes de o utilizador se autenticar, apenas poderá utilizar o seu terminal para a realização de chamadas de emergência (o equivalente a uma chamada para o 112). Para aceder a qualquer outro serviço de rede, o utilizador terá de apresentar uma autenticação válida (mensagem 1 nas Figura 32 e Figura 33 – ver também a Tabela 8) junto do *router* de acesso, o que é em tudo idêntico ao processo nas redes celulares actuais. Este processo é conduzido entre o terminal e o sistema de AAAC, com o módulo atendedor de AAAC do *router* de acesso a funcionar como *proxy* e reenviando a mensagem para o servidor local de AAAC (mensagem 2). Note-se que quem é autenticado é o utilizador e não o terminal (que é o que se passa hoje em dia nas redes celulares GSM, por exemplo). Desta forma, o utilizador poderá aceder aos serviços presentes no seu perfil de um qualquer terminal que os suporte.

No caso de um utilizador em *roaming*, que é um cenário mais complexo, o servidor de AAAC local fará um pedido de autenticação e autorização ao sistema de AAAC de origem do contrato (mensagem x), ou seja, o utilizador encontra-se a ser servido por um servidor que não é pertença do operador/fornecedor com quem tem contrato, e este terá que fazer a validação do utilizador junto do seu fornecedor. Tudo isto, claro, se houver acordo

de *roaming* entre os dois fornecedores (o que implica também a existência de um SLA entre ambos). Assim, o sistema de AAAC que actualmente serve o terminal, terá de previamente verificar a existência de um contrato de *roaming* com o operador responsável pelo cliente em causa. Então, em caso de resposta afirmativa, o sistema de AAAC responsável pelo cliente verificará as credenciais do utilizador e, em caso de existência de um contrato válido, fará ainda a verificação do endereço (*home address*) fornecido junto do *home agent* (mensagem x), e este valida-o (mensagem y). O sistema de AAAC de origem informará então o sistema de AAAC local (mensagem z), enviando o perfil desse cliente contendo, entre outras coisas, a NVUP. Poderá haver tradução de perfil nas fronteiras dos domínios, sendo isso dependente quer das características dos serviços prestados em cada um dos domínios, quer do acordo existente entre os fornecedores. Este processo é facilmente implementado sobre DIAMETER.

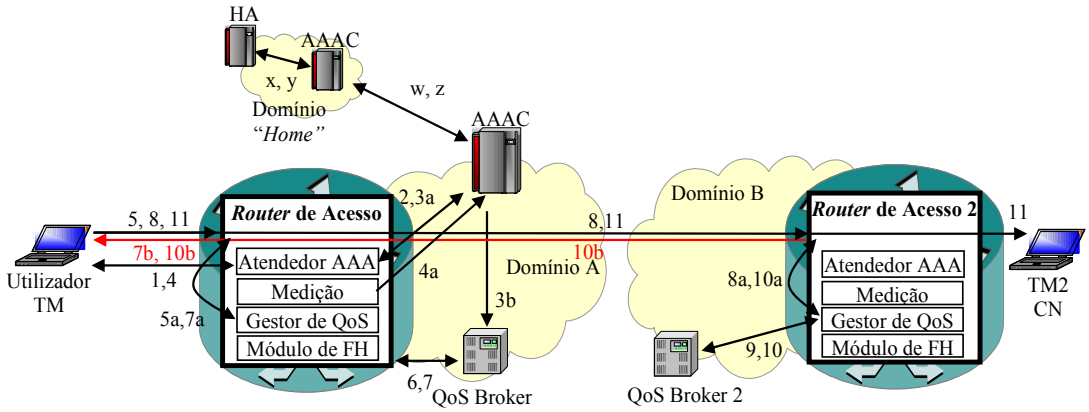


Figura 32: Processo de Autenticação e Autorização com suporte de QoS

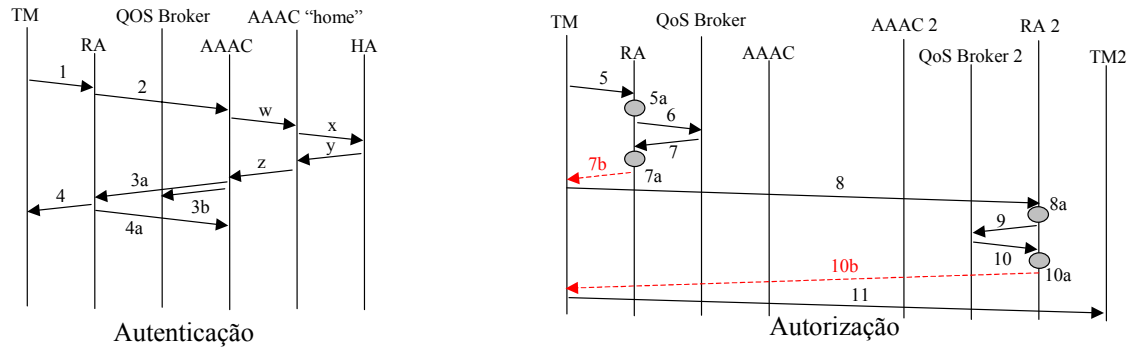


Figura 33: Diagramas de sinalização de autenticação e autorização com suporte de QoS

N.º	Mensagem/acção	Conteúdo / Parâmetros	Comentários
1	Pedido de autenticação	Endereço de identificação de rede (NAI), credenciais e CoA	NAI – *Network Address Identifier*
2	Pedido de autenticação	NAI, credenciais, CoA	O RA actua como *proxy* do utilizador/terminal
w	Pedido de autenticação	NAI, credenciais, CoA	No caso de o utilizador estar fora do seu domínio, o pedido de autenticação terá que ser enviado ao servidor de AAAC no seu domínio
x	Pedido de autenticação	CoA	Pedido de mapeamento do CoA no endereço "*home*"
y	Resposta de autenticação	Endereço "*home*"	Resposta com o mapeamento entre CoA e endereço "*home*"
z	Resposta de autenticação	Chaves para o TM e RA estabelecerem relação de confiança, NVUP para o TM, e valor do temporizador de limite de sessão	Informação para o sistema AAAC remoto
3a	Resposta de autenticação	Chaves para o TM e RA estabelecerem relação de confiança, NVUP para o TM, e valor do temporizador de limite de sessão	Informação para o RA e para o TM. NVUP: *Network View of the User Profile*
3b	NVUP do utilizador para o QoS *Broker*	Informação do TM (identificação) e NVUP	
4	Resposta de autenticação	Chave secreta para o TM dialogar com o RA, NVUP para o TM, e valor do temporizador de limite de sessão	Com esta informação o terminal saberá como criar ligações de confiança com o RA, quais os serviços do perfil do utilizador, quais os códigos para os pedir e o tempo ao fim do qual a sessão termina por inactividade
4a	Início de contabilização		
5	Pedido de serviço	DSCP do serviço, CoA e endereço destino	
5a	Mensagem interna de pedido de autorização		Pedido interno ao gestor de QoS
6	Pedido de serviço	DSCP do serviço, CoA e endereço destino	O RA actua como *proxy* do TM
7	Confirmação ou negação de serviço	Política para instalação do serviço ou negação do serviço	Informação para configuração do RA ou negação do serviço
7a	Configuração do serviço		Configuração do *router* de acesso, caso o serviço tenha sido autorizado
7b	Negação de serviço	Informação de negação	O serviço é negado caso não existam recursos disponíveis ou o perfil do utilizador não inclua o serviço pretendido
8	Pedido de serviço	DSCP do serviço, CoA e endereço destino	
8a	Mensagem interna de pedido de configuração		Pedido interno ao gestor de QoS
9	Pedido de serviço	DSCP do serviço, CoA e endereço destino	
10	Confirmação ou negação de serviço	Política para instalação do serviço ou negação do serviço	Informação para configuração do RA ou negação do serviço
10a	Configuração do serviço		Configuração do *router* de acesso, caso o serviço tenha sido autorizado
10b	Negação de serviço		O serviço é negado caso não existam recursos disponíveis
11	Ligação estabelecida	Informação das aplicações correntes	

Tabela 8: Explicação do processo de autenticação e autorização com suporte de QoS

O perfil do utilizador é gerido de uma forma centralizada, pelo seu fornecedor, e contém toda a informação específica para o aprovisionamento dos serviços contratados. A NVUP, uma vez chegada ao sistema de AAAC (e eventualmente traduzida em termos de especificidades de serviços no domínio corrente) será então fornecida ao QoS *Broker* (mensagem 3b) e, através do *router* de acesso, também ao terminal móvel (mensagem 3a e 4). Desta forma, o terminal ficará instruído acerca dos serviços presentes no perfil do utilizador e de como os utilizar (quais os DSCPs que correspondem a cada serviço). Por

outro lado, também o QoS *Broker* fica com o conhecimento da presença daquele utilizador no seu domínio de QoS, do seu perfil de serviços, e poderá a partir dessa altura realizar controlo de admissão de serviços para aquele utilizador. A mensagem 3a transporta também informação relacionada com a medição e contabilização que será utilizada pelo agente de AAAC no *router* de acesso. Nesta fase, o *router* de acesso inicia o processo de contabilização de uso para o utilizador em causa (mensagem 4a), terminando assim o processo de autenticação.

5.4.2 Autorização / Início de sessão

A Figura 32 e a Figura 33 mostram como se processa a autorização de serviço (mensagens 5 a 11). O processo é despoletado pelo terminal móvel, enviando um pacote marcado com o DSCP referente ao serviço pretendido (mensagem 5) (e.g., um serviço de voz, prioritário, a 64kbps). Este pacote poderá ser um pacote já com informação que o terminal pretende transmitir, ou em alternativa, poderá também ser um pacote destinado apenas a fazer a activação do serviço (opção configurada no terminal, no MTMN). No *router* de acesso (RA), o agente de QoS (mensagem 5a) verifica se o pacote em causa é referente a algum serviço já a decorrer, ou se se trata de um novo serviço (DSCP diferente dos serviços já activados). Caso seja um novo serviço a que o utilizador pretenda ter acesso, o *router* de acesso, irá consultar o QoS *Broker* sobre como proceder com todos os pacotes semelhantes. Para tal, o agente de QoS no *router* de acesso, envia um pedido COPS ao QoS *Broker* (mensagem 6), indicando o endereço de origem do terminal móvel, o endereço de destino (a quem o pacote se dirige) e o código DSCP (outra informação tal como os portos UDP/TCP, *flowlabels*, etc poderá ser também enviada). O QoS *Broker* analisa o pedido e, de acordo com o perfil do utilizador e com os recursos disponíveis, nega ou autoriza o serviço (mensagem 7). Caso a sessão seja autorizada, o gestor de QoS no *router* de acesso configura uma política de acordo com o serviço em causa (mensagem 7a) e todos os pacotes subsequentes (mensagem 8 e 11), dentro do perfil do serviço, poderão passar para o *core* da rede, isto é, são autorizados. Caso a resposta do QoS *Broker* seja negativa, o terminal será informado do facto (mensagem 7b), podendo optar por um serviço com menores requisitos, mudar de célula (e possivelmente tecnologia) de acesso, ou esperar e voltar a tentar mais tarde (tal como hoje em dia acontece nas redes telefónicas fixas e celulares).

Na rede de destino, quando os pacotes chegam ao *router* de acesso, o gestor de QoS, mais uma vez, pedirá informações ao QoS *Broker* acerca do modo como actuar perante estes pacotes (mensagem 8a e 9). No entanto, neste caso, o QoS *Broker* não fará controlo de admissão do serviço do utilizador (se o pacote vem do *core* da rede, já terá sido autorizado na entrada). Assim, o QoS *Broker* apenas irá analisar qual o serviço sinalizado e qual o destino (qual o terminal móvel destinatário) e, em caso de haver recursos disponíveis, irá instruir o gestor de QoS no *router* de acesso das características do serviço (mensagem 10). Este, (mensagem 10a) irá configurar a política no *router* que permitirá o tráfego atingir o seu destino (mensagem 11). No caso de não haver recursos disponíveis, o *router* de acesso envia uma mensagem indicando o facto (mensagem 10b). Com esta estratégia, ambas as redes de acesso envolvidas garantem a existência e disponibilidade dos recursos, enquanto no *core* há uma monitorização constante dos níveis de QoS e de disponibilidade de recursos.

Os pacotes com um DSCP diferente irão desencadear o pedido de autorização de um outro serviço e todo o processo se repetirá. Os serviços, identificados pelos DSCPs podem dar origem a reservas unidireccionais ou bidireccionais, simétricas ou assimétricas, dependendo apenas da política de gestão de serviços e recursos do domínio administrativo em causa. Há uma excepção no caso de comunicações inter-domínio, em que terá de haver acordo entre os domínios envolvidos e eventual adaptação e tradução de serviços nas fronteiras dos domínios.

5.4.3 *Handover* com suporte de QoS

O aspecto mais difícil de lidar em redes IP com mobilidade é assegurar um nível constante de QoS. Nesta arquitectura, a mobilidade é suportada através da utilização de técnicas de *Fast Handover* em combinação com a troca de mensagens entre os QoS *Brokers* durante o *Handover*, conforme ilustrado nas Figura 34, Figura 35 e Tabela 9.

Quando o terminal móvel, durante o seu movimento, começa a perder sinal de ligação para o *router* de acesso onde está ligado (*router* "antigo") (mensagem 1), inicia o procedimento de *handover* para um outro *router* (*router* "novo") do qual está a receber um sinal de *beacon*, contendo o novo prefixo de rede (mensagem 2). Utilizando o prefixo de rede recebido, o terminal constrói um CoA e inicia o procedimento de *handover*, enviando um pedido de *handover* IP para o *router* "novo", utilizando a ligação existente, através do *router* "antigo" (mensagem 3). O módulo de FHO no *router* "antigo" reenvia o pedido para

o gestor de QoS do *router* "antigo" (mensagem 3a) e para o módulo do *router* "novo", descoberto pelo prefixo de rede (mensagem 4a). O gestor de QoS do *router* "antigo" envia também o pedido para o seu QoS *Broker* (mensagem 4b) (agora conhecido por QoS *Broker* "antigo"). Este, envia o pedido de *handover* para aquele que será o potencial QoS *Broker* "novo" (mensagem 5), responsável pelo domínio do *router* "novo", indicando a NVUP do utilizador em causa, assim como a lista de serviços que o terminal tem correntemente activos. Este processo consiste numa transferência de contexto entre os QoS *Brokers* "antigo" e "novo". Na posse desta informação, o QoS *Broker* "novo", verificará a existência de recursos na rede de acesso do *router* "novo" para acolher o terminal com todos os serviços presentemente activos. Nesta fase, o QoS *Broker* "novo" fará também a detecção de duplicação de endereços, comparando o CoA recém construído pelo terminal com os endereços de todos os terminais já presentes naquela rede de acesso (por o prefixo ser único por cada rede, está assim garantida a unicidade do endereço). O QoS *Broker* "novo" indicará ao gestor de QoS do *router* "novo" se o resultado da verificação de existência de recursos e duplicação de endereço é ou não favorável ao *handover* (mensagem 6). Desta forma, o QoS *Broker* tem poder para abortar o processo de *handover* em caso de constrangimentos de QoS, isto é, caso não existam recursos suficientes na nova célula para acolher os serviços de que o terminal móvel está a fazer uso no momento. Caso o *handover* seja possível de realizar, o gestor de QoS do *router* "novo" configura as políticas necessárias no *router* "novo" (recebidas na mensagem 6) e informa o módulo gestor de *fast handover* (mensagem 6a), que entretanto está à espera desta informação (desde a mensagem 4a) para responder ao *router* "antigo" (mensagem 7). Quando o *router* "antigo" recebe esta mensagem reenvia-a desde logo para o terminal móvel (mensagem 8). Quando a resposta ao pedido de *handover* é favorável, o terminal móvel envia uma indicação de execução de *handover* ao *router* "antigo" (mensagem 9) que reage iniciando um temporizador e o *bicasting* (mensagem 9a) dos pacotes dirigidos ao terminal também para o *router* "novo". Durante o *bicasting*, cada pacote dirigido ao terminal será duplicado, de modo a que antes de se realizar o *handover* o terminal recebe os pacotes através do *router* "antigo", e após a realização do *handover* os pacotes são já recebidos através do *router* "novo". Seguidamente, o *router* "antigo" envia uma confirmação ao terminal (mensagem 10). Neste ponto, o terminal pode executar o *handover* (mensagem 10a) com a garantia de que será bem sucedido e sem falhas de QoS uma vez que todo o processo de

handover de nível 3 foi preparado de antemão. O tempo necessário para executar o *handover* de nível 2 será o único constrangimento em termos de QoS (em tudo semelhante ao que se passa hoje em dia nas redes celulares GSM). Uma vez efectuado o *handover* de nível 2, o terminal envia um neighbour advertisement ao *router* "novo" (mensagem 12) fazendo com que este inicie um processo de contabilização no sistema de AAAC (mensagem 13) para o utilizador em causa. Para concluir todo o procedimento de *handover* terá de se completar o procedimento de mobilidade IPv6. Sendo assim, o terminal terá de enviar um biding *update* ao seu home agent (mensagem 14) que responderá com um *binding* acknowledge (mensagem 15). Entretanto, no *router* "antigo", o temporizador iniciado para o processo de *bicasting* expira (mensagem 10b) e este *router* enviará para o sistema de AAAC (mensagem 11) todos os dados de utilização de serviços deste utilizador.

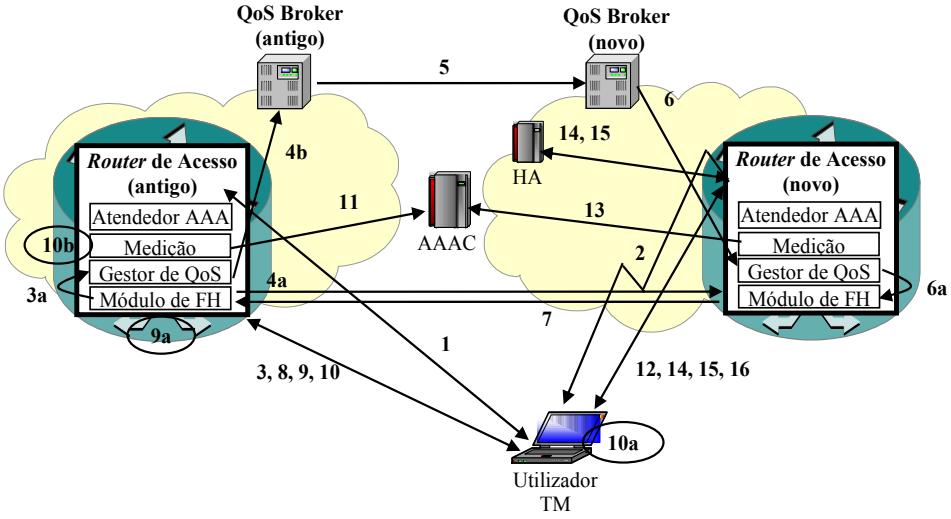


Figura 34: *Handover* com suporte de QoS

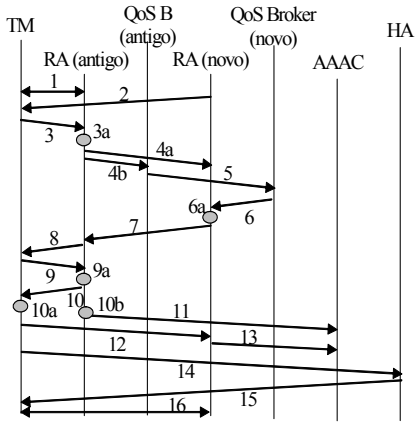


Figura 35: Diagrama de sinalização de *handover* com suporte de QoS

N.º	Mensagem/acção	Conteúdo / Parâmetros	Comentários
1	Ligação actual	Informação das aplicações correntes	
2	Anúncio de *Router*	Prefixo de rede	Novo *router* ao alcance
3	Solicitação ao *Router para* actuação como *proxy*	Endereço do *router* novo, CoA novo (construído com base no prefixo)	Pedido de *handover* rápido
3a	Mensagem interna do *router* de acesso para o gestor de QoS	Identificação do endereço do terminal	Indicação do módulo de FH ao gestor de QoS de início de *handover*
4a	Início de processo de *Handover*	Sub-perfil do utilizador, chave de segurança, e CoA novo	
4b	Indicação de pedido de *Handover*	Endereço do *router* novo, e o CoA actual (antigo)	Com base nesta informação, o QoS *Broker* sabe qual o terminal que se pretende mover e sabe o endereço do QoS *Broker* responsável pelo *router* novo
5	Indicação de pedido de *handover*	Endereço *home* do terminal, o CoA novo, a NVUP e a lista de serviços presentemente activos	Com base nesta informação, o QoS *Broker* novo verificará se já existe algum terminal na sua rede com um endereço igual e saberá se poderá aceitar o *handover* de todos os serviços activos
6	Indicação de decisão de *handover*	Resultado da detecção de duplicação de endereço, informação de decisão de aceitação, e em caso positivo, informação para configuração do *router* novo	Transporta informação para o RA novo ou indicação de falha, quer por falta de recursos, quer por detecção de duplicação de endereço.
6a	Mensagem interna do *router* de acesso para o módulo de FH	Indicação positiva/negativa de aceitação de *handover*	
7	Resposta ao pedido de *Handover*	Identificação do terminal	Confirmação ou rejeição do *handover*
8	Resposta de *handover* (actuação como *proxy*)		Confirmação ou rejeição do *handover*
9	Execução de *Handover*		Início de *handover*
9a	Início de replicação (*bicasting*) de pacotes e de temporizador		Todos os pacotes direccionados para o terminal são também enviados para o *router* novo, até o temporizador expirar
10	Confirmação de execução de *handover*		Processo de preparação de *handover* rápido concluído
10a	Abandono da ligação actual		*Handover* físico
10b	Temporizador de replicação de pacotes expirado		Termina a replicação de pacotes
11	Final da recolha de informação de contabilização e correspondente envio	Informação de contabilização dos diversos serviços e identificação do utilizador	Toda a informação de contabilização é enviada ao sistema de AAAC
12	Anúncio de vizinho	CoA novo	Ligação IP ao *router* novo
13	Início de contabilização	Identificação do utilizador	O *router* novo inicia o processo de contabilização
14	Actualização de informação de endereço (e localização)	CoA novo	O agente "*home*" é informado da nova localização (e endereço) do utilizador/terminal
15	Confirmação de actualização de endereço (e localização)		Com esta mensagem todo o processo de *handover* fica terminado
16	Continuação das sessões activas	Informação das aplicações correntes	Transferência de informação

Tabela 9: Explicação do processo de *handover* rápido com suporte de QoS

O processo de *fast handover* com suporte de QoS fica então concluído, com o terminal na rede de acesso do *router* "novo" (mensagem 16). Para o caso particular de um *handover* executado dentro do domínio de QoS de um QoS *Broker* apenas, a mensagem 5 passará a ser uma mensagem interna ao QoS *Broker* e tudo o resto decorrerá do modo explicado anteriormente. De referir que os atendedores de AAAC presentes nos *routers* são também informados do processo de *handover* e os parâmetros correntes de AAAC do utilizador (por exemplo para medição de consumos) são também trocados directamente durante o processo de iniciação de *handover* (mensagem 4a).

5.5 Considerações sobre a arquitectura proposta

5.5.1 Flexibilidade de gestão

A arquitectura apresentada engloba um conjunto de entidades que cooperam no controlo e gestão de recursos. A possível existência de um NMS (*Network Management System*) interligado aos QoS Brokers e aos sistemas de AAAC permite o controlo de todos os recursos e situações anómalas em tempo real. Este conhecimento do estado de cada elemento e dos recursos permite impor medidas correctivas em situações de previsível degradação, ou apenas uma adaptação dos elementos para uma optimização da utilização dos recursos. O facto de os recursos serem geridos ao nível do IP, garante uma uniformidade de processos de controlo e gestão em toda a rede com vantagens claras em termos da flexibilidade obtida, face a arquitecturas actuais onde serviços orientados à ligação coexistem com serviços de pacotes.

5.5.2 Serviços definidos estaticamente vs serviços dinâmicos

Na arquitectura proposta, os serviços estão na sua maioria pré-definidos em termos das suas características de recursos de rede. Esta abordagem poderá à primeira vista parecer limitativa, mas tem na realidade grandes vantagens para os operadores. Atendendo a que esta arquitectura visa suportar, entre outros, o actual serviço de voz, não faz sentido que o "típico" serviço de voz possa ter parametrizações variáveis, dependendo de quem inicia a ligação. Este serviço recorre a *codecs* bem conhecidos e parametrizados, e desta forma não faz sentido envolver um protocolo de sinalização, despendendo mais tempo, para negociar parâmetros que são bem conhecidos à partida. O facto de quer a rede quer o utilizador já conhecerem o serviço à partida, traz vantagens quer do ponto de vista de gestão da rede, quer ao utilizador em termos de tempo de estabelecimento de ligação. O exemplo das chamadas de voz é apenas um de entre muitos. Para uma utilização vulgar do serviço de navegação na *Internet*, também não é necessário estar a negociar parâmetros. O utilizador poderá incluir no seu contrato um serviço que seja o reflexo da utilização de recursos típica da navegação na *Internet*. Mais uma vez, ganha-se tempo e poupa-se em complexidade. Há no entanto, situações e utilizadores que exigem negociação dinâmica de recursos. A arquitectura proposta está também preparada para ter este tipo de serviço. No entanto, este tipo de serviço levará necessariamente mais tempo na autorização e no

estabelecimento da ligação (torna-se necessário negociar os parâmetros e verificar a sua disponibilidade em todo o percurso). Do ponto de vista comercial, por ser um serviço "à medida", mais complexo e mais exigente (em termos de recursos de controlo) será, em princípio, mais dispendioso para o cliente.

Um outro argumento a favor dos serviços pré-definidos é a capacidade de entendimento do serviço por parte dos clientes. Esta arquitectura destina-se a todos os tipos de clientes, e não apenas a pessoas que tenham percepção de telecomunicações, e de redes de pacotes em particular. Será certamente mais fácil explicar a qualquer leigo que o pacote de serviços que está a adquirir lhe permite fazer chamadas de voz e navegação na *Internet* e que por isso ele paga "x" (dependendo por exemplo apenas do tempo de utilização), do que tentar explicar à mesma pessoa, que irá pagar de acordo com as condições (dificilmente explicáveis) que usufruir em determinado instante, e que são eventualmente negociadas pelo seu terminal, sem a sua intervenção. Portanto, até neste ponto, esta arquitectura de serviços é bastante flexível, permitindo que se construam pacotes de serviços pré-definidos, que reflectirão as utilizações mais típicas, abrangendo a maioria dos utilizadores, havendo para uma minoria, que necessita de recorrer a serviços mais específicos, a possibilidade de incluírem no seu pacote de serviços, um serviço que lhe oferece a possibilidade de aceder aos recursos de rede de uma forma costumizada.

No entanto esta abordagem pode trazer um problema. Um utilizador que, por exemplo, tenha no seu perfil um serviço de 60 kbps que lhe permite realizar uma chamada de voz de boa qualidade com o *codec* "x", se esse utilizador tentar utilizar o mesmo serviço para estabelecer uma segunda comunicação de voz, verá a qualidade das suas chamadas degradada. Poderá este utilizador reclamar junto do seu fornecedor por violação do SLA? De facto, não, pois o serviço estará definido de forma a que seja possível realizar uma chamada com a qualidade pré-definida. Caso o utilizador pretenda um serviço que lhe permita realizar várias chamadas em simultâneo, terá que procurar dentro dos serviços disponibilizados pelo operador, um que lhe permita realizar esse tipo de operações. Caso não exista, então terá que incluir no seu pacote de serviços, o serviço flexível de negociação no momento.

5.5.3 Mapeamento de serviços entre domínios administrativos distintos

A arquitectura proposta visa poder ser aplicada numa escala global, envolvendo diversos operadores. Assim, torna-se essencial garantir que os serviços extremo-a-extremo

possam atravessar diversos ambientes de operadores distintos, mantendo as suas características essenciais. É então forçoso assumir que os diversos serviços oferecidos num domínio administrativo tenham um serviço igual ou equivalente no domínio contíguo. Haverá desta forma lugar à definição de SLAs entre os domínios vizinhos, permitindo a tradução dos serviços nas fronteiras. No caso mais simples, em que um serviço tem exactamente as mesmas características nos dois domínios, o mapeamento consistirá, neste caso, numa tradução do valor do DSCP dos pacotes que atravessam os domínios (caso o próprio DSCP não seja o mesmo). No entanto, conceptualmente, a equivalência do serviço deverá ser verificada em termos de parâmetros do SLA por via de testes de concordância, aplicando métodos e procedimentos de medição. Os SLAs nas fronteiras dos domínios são também os garantes do tratamento do tráfego que não esteja conforme, aplicando formatação, descarte e remarcação.

No entanto, haverá muitos casos em que não haverá serviços equivalentes. Estes casos, mais complexos, exigirão um tratamento também mais elaborado. Estes serviços dependerão de um acordo fronteira, no qual o SLA entre os operadores define as regras da sua tradução. No caso mais simples, duas situações distintas podem acontecer: o serviço de um domínio é traduzido num serviço com melhores características no outro domínio, ou o inverso, isto é, um serviço poderá sofrer degradação na sua tradução para o domínio seguinte. Estas duas situações serão conjugadas com o tipo de serviço (simétrico ou assimétrico) e a sua direccionalidade (uni ou bidireccional) resultando em sete situações distintas possíveis:

- serviço unidireccional degradado de um domínio para o seguinte
- serviço unidireccional recebendo melhor tratamento no domínio seguinte
- serviço bidireccional simétrico em que num dos sentidos será degradado
- serviço bidireccional simétrico com melhores condições no domínio destino
- serviço bidireccional assimétrico em que ambos os sentidos são melhor tratados no domínio vizinho
- serviço bidireccional assimétrico em que um dos sentidos é degradado e o outro melhorado
- serviço bidireccional simétrico degradado em ambos os sentidos

Os casos em que exista degradação das características do serviço na sua tradução, serão os de tratamento mais delicado. O SLA entre os domínios terá que definir as regras de tradução, assim como as margens de degradação máximas. Um exemplo concreto existente nas actuais redes de telecomunicações é o que acontece quando se estabelece uma ligação RDIS da Europa (com um ritmo de transmissão por canal é de 64kbps) para os Estados Unidos (em que o ritmo de transmissão é de 56 kbps por canal). A mesma problemática se aplica a um utilizador que esteja em *roaming* num domínio em que os seus serviços não tenham um mapeamento directo.

Uma possibilidade é haver lugar a uma adaptação dinâmica dos serviços, recorrendo neste caso a sinalização entre os QoS *Brokers* de cada domínio, dentro de cada domínio e ainda com os terminais. Desta forma, garante-se à partida que os terminais sabem exactamente quais os recursos disponíveis em determinada ligação e podem, desta forma, adaptar-se a eles. Esta aproximação traz consigo a necessidade da existência de um protocolo de sinalização explícita, que não foi definido no âmbito desta arquitectura, embora esta tenha sido desenvolvida de forma a suportar este tipo de funcionalidade. Esse protocolo terá de disponibilizar funcionalidades que permitam a identificação dos recursos de QoS necessários e a sua negociação, eventualmente dinâmica. Contudo, não é obrigatório que sejam os QoS *Brokers* a fazerem esta negociação se outra entidade de gestão a fizer, havendo no entanto necessidade de envolver os QoS *Brokers* numa primeira fase de identificação de recursos e, numa fase final, de reserva dos mesmos.

A identificação dos requisitos de QoS pode ser baseada no processamento dos pedidos de serviço, extraindo-os do pacote que desencadeia o pedido de autorização de serviço, e que através da rede de gestão dos fornecedores envolvidos sejam encaminhados e processados, até ao destino, nomeadamente o QoS *Broker* e o terminal destino. Seguir-se-á a fase de negociação que culminará com o estabelecimento da ligação (uni ou bidireccional).

5.5.4 Serviços unidireccionais vs bidireccionais, simétricos vs assimétricos

Existem serviços que são de carácter claramente bidireccionais e simétricos, tais como a comunicação de voz, havendo outros que embora bidireccionais têm um caracter essencialmente assimétrico, e outros cuja essência é unidireccional. A arquitectura aqui proposta está preparada para lidar com todos estes tipos de serviço.

O modo como os serviços são sinalizados nesta arquitectura é através da marcação dos pacotes IP com um valor de DSCP que identifica univocamente um serviço. Existem várias estratégias possíveis para permitir a definição de serviços bidireccionais, mas aquela que aqui é apresentada é inteiramente baseada num conceito semelhante ao da identificação de serviços unidireccionais. Assim, um utilizador que inicie um fluxo marcando os seus pacotes com o DSCP=5, por exemplo, poderá estar a identificar um serviço com características bem definidas em termos de prioridade, largura de banda, direccionalidade e simetria. Como regra geral (com eventuais excepções), podemos assumir que num serviço unidireccional, cada um dos intervenientes na ligação será responsável pelo pagamento do tráfego que gerar ou tempo da ligação. No caso de um serviço bidireccional, uma vez que a autorização de serviço será desencadeada por apenas um dos intervenientes, o chamador, é natural que este seja o responsável pelos encargos da ligação. Todos estes aspectos serão função das políticas do operador.

5.5.4.1 Serviço unidireccional (emissão apenas)

Por exemplo, suponhamos que um DSCP=x representa um serviço unidireccional, de sentido ascendente (com possibilidade de enviar tráfego apenas do terminal para a rede). Neste caso, o procedimento é aquele que foi indicado nas secções anteriores, ou seja, o utilizador sinaliza o serviço enviando pacotes com o DSCP=x e o QoS *Broker* configurará o *router* de acesso de modo a permitir que todos os pacotes originados no terminal do utilizador, marcados com o DSCP=x e dentro da largura de banda prevista para o serviço em causa, possam fluir para o *core* da rede. Todo o restante tráfego estará dependente da existência de outros serviços activos, quer iniciados pelo utilizador, quer iniciados por outro utilizador, mas com destino a este terminal.

5.5.4.2 Serviço unidireccional (recepção apenas)

No caso de um serviço unidireccional de sentido descendente, o terminal do utilizador, ao enviar um pacote marcado com o DSCP=y, irá despoletar no RA um pedido ao QoS *Broker* que, de acordo com o perfil do utilizador, configurará no RA uma regra permitindo que o terminal possa receber tráfego vindo do *core* da rede, marcado com um DSCP que poderá não ser o "y" (o DSCP com que o tráfego vem marcado será aquele que constará no perfil do utilizador, associado ao serviço sinalizado com o DSCP=y).

5.5.4.3 Serviço bidireccional simétrico

No caso de um serviço bidireccional, há a necessidade de garantir que os recursos são reservados em ambas as direcções no RA do terminal que inicia o processo, mas também no RA do terminal destino. Para o caso de um serviço simétrico, o procedimento é relativamente simples. O terminal ao enviar pacotes com o DSCP=w despoleta no RA e no QoS *Broker* locais um processo de configuração de uma política de envio e outra de recepção de pacotes, que podem ser encaminhados com uma prioridade e até uma taxa de transferência pré-definidas no perfil do serviço, desde que marcados com o DSCP=w. No lado da recepção, o mesmo pacote irá despoletar uma acção semelhante por parte do conjunto RA e QoS *Broker*, mas neste caso não é importante saber se o terminal destino subscreveu ou não este serviço (se existe no seu perfil), pois o processo está a ser suportado pelo perfil e contrato do chamador.

5.5.4.4 Serviço bidireccional assimétrico

Um serviço assimétrico poderá também ser sinalizado apenas por um único DSCP, mas, por uma questão de simplicidade de implementação, é conveniente que a comunicação entre os dois extremos se faça utilizando dois DSCPs distintos, um para cada sentido. Assim, o terminal origem, ao marcar o tráfego como DSCP=z, irá despoletar no RA e QoS *Broker* um processo de configuração de duas políticas distintas, uma para envio e outra para recepção, de acordo com a especificação do serviço identificado pelo DSCP=z. No lado da recepção, o QoS *Broker*, ao verificar que a proveniência do pacote é o *core* da rede, sabe também que deve configurar duas políticas para o utilizador destino, de acordo com o perfil do serviço. Também neste caso, o perfil do utilizador destino não é importante, pois o chamador será o responsável pela ligação. Após isto, no sentido do chamador para o chamado os pacotes deverão ser marcados com o DSCP=z, enquanto que no sentido inverso (chamado – chamador) deverá ser utilizado um outro valor de DSCP (z'), também associado ao serviço sinalizado com o DSCP=z.

Uma questão importante a ter em consideração no caso de serviços bidireccionais é o pagamento do serviço. Atendendo que o iniciador do serviço poderá na maioria dos casos ser o responsável pelo pagamento, que o serviço é sinalizado por um DSCP e que o serviço por omissão só será terminado após um *timeout* na transmissão de pacotes relativos a esse serviço, terá que haver um processo de impedir que o terminal destino aproveite o facto de

alguém lhe ter ligado para utilizar essa ligação para se ligar a outros destinos, mesmo que o chamador tenha terminado a ligação. Uma forma de garantir isto, é as políticas instaladas nos *routers* de acesso no destino incluírem o endereço origem, endereço de destino e código DSCP. Assim, se o terminal destino decidir enviar pacotes para outro endereço que não o do chamado, mesmo que utilizando o mesmo DSCP, terá que ser realizada a função de controlo de admissão e autorização de serviço, havendo em caso de sucesso, a inclusão de mais uma política no *router* de acesso relativa a esse terminal.

5.5.5 Vantagens desta arquitectura

Esta arquitectura foi definida de forma a que possa facilmente adaptar-se e integrar diferentes políticas de gestão de recursos. Devido às bases de dados e interfaces incorporadas, o QoS *Broker* definido consegue optimizar o controlo e utilização dos recursos.

A utilização de um QoS *Broker* nesta arquitectura poderá ser posta em causa, alegando-se que as funções desempenhadas por este podem ser delegadas nos *routers* de acesso. No entanto, as vantagens da utilização do QoS *Broker* podem ser facilmente explicadas.

O controlo dos recursos semi-centralizado associado aos QoS *Brokers* permite a delineação de diversas estratégias de gestão de mobilidade. Por exemplo, o QoS *Broker* pode indicar células alternativas a um terminal que pretenda mover-se para uma célula com recursos insuficientes.

Uma outra vantagem desta abordagem é o facto de permitir que se possa optar por um controlo de decisão de *handover* partilhado entre a rede e o terminal associado inclusivamente à mobilidade dos utilizadores. Assim, facilmente se pode implementar um modo de funcionamento em que a rede sugere ao terminal que este faça o *handover* para uma outra célula, quer por uma questão de melhoria de serviço para o próprio terminal, quer por uma questão de distribuição de utilização de recursos. Este cenário poderá ter interesse no caso em que a alocação e transferência de utilizadores de umas células para outras permita gerir os recursos de modo a poder acolher mais utilizadores, que noutra circunstâncias seriam rejeitados. Este é um cenário comum em redes rádio.

Vejamos um exemplo. Suponhamos o caso de duas células contíguas com um total de 10Mbits/s de largura de banda cada ilustrado na Figura 36. Suponhamos que num determinado instante, a célula do centro (“azul”) está com 90% dos seus recursos

ocupados, enquanto que a célula da direita ("verde"), tem apenas 20% dos seus recursos ocupados. Suponhamos ainda, que existem utilizadores em zonas cobertas por ambas as células, e que destes, os que estão a utilizar a célula "azul" são responsáveis pelo consumo de 50% dos recursos dessa célula. Imaginemos um autocarro, com alguns utilizadores a consumirem recursos de rede que se aproxima, pelo lado esquerdo (célula "amarela") da célula "azul". Os recursos que os utilizadores do autocarro estão a utilizar, correspondem a 3,5 Mbps (cerca de 35% da capacidade das células). É obvio que, numa situação normal, quando estes utilizadores tentassem entrar na célula "azul", apenas alguns conseguiriam manter o seu serviço. No entanto, se ambas as células forem controladas por um QoS *Broker* (ou dois QoS *Brokers* que comuniquem entre si), seria possível que o QoS *Broker* indicasse aos utilizadores "azuis" que estão na zona de cobertura da célula "verde", que fizessem *handover* para esta, libertando desta forma os recursos necessários para acolher os novos utilizadores.

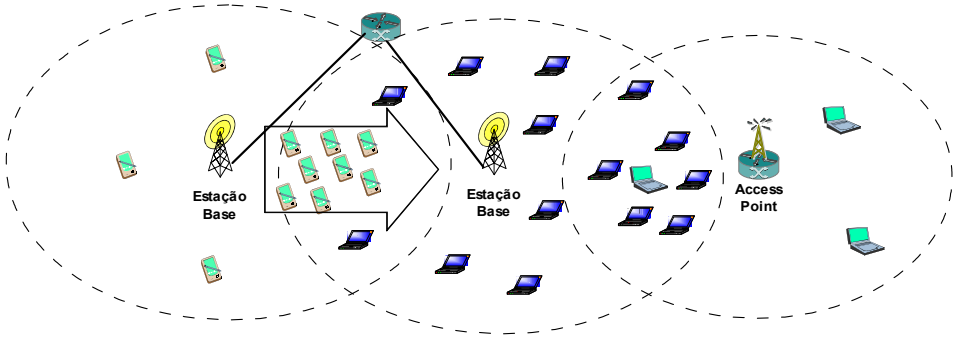


Figura 36: Cenário para gestão avançada de recursos pelo QoS *Broker*

O facto de se utilizarem QoS *Brokers* traz ainda outras vantagens:

- Uma mudança de política de gestão e controlo de recursos, implica mudanças em poucos elementos (os QoS *Brokers*) em vez de alterações em todos os *routers* de acesso.
- Do ponto de vista económico, é mais vantajoso ter concentrados em poucos sistemas as grandes necessidades de processamento e de manutenção de estado, querendo isto dizer, que é economicamente preferível para os operadores investirem em (poucos) sistemas potentes (QoS *Brokers*) do que ter essas necessidades espalhadas por todos os elementos da rede (os *routers*), que rapidamente poderiam ficar subdimensionados.

- A abordagem com QoS *Brokers* permite ao operador um distanciamento e independência dos fabricantes de equipamentos de acesso. Um QoS *Broker* poderá virtualmente controlar qualquer modelo de qualquer fabricante, através de interfaces normalizadas (por exemplo, COPS) sem necessitar de nenhuma alteração (bastando para tal desenvolver um módulo de *router* virtual para o modelo em causa).

5.5.6 Políticas de gestão de recursos nos QoS *Brokers*

Conforme foi dito, esta arquitectura baseada nos QoS *Brokers* é bastante flexível, permitindo implementar a política que o operador considerar melhor. Por exemplo, se o operador quiser optimizar os recursos, recorrendo a uma ligeira (ou não) degradação do serviço dos utilizadores, o QoS *Broker* é facilmente parametrizável para permitir determinados graus de sobre-reservas de recursos. Num acesso em que haja apenas 1Mbit/s de recursos disponíveis para serviços de alta prioridade, o operador poderá permitir um total de serviços aceites de, por exemplo 110%, "jogando" com multiplexagem estatística, e com alguma eventual degradação de serviço. Esta taxa de sobre-reservas depende obviamente da política do operador e poderá variar no tempo. Por exemplo, numa altura em que exista uma campanha de promoção com grandes reduções de tarifas, é natural que a degradação permitida seja maior, de modo a "rentabilizar" melhor a promoção. Os utilizadores, que estão a pagar bastante menos pelo serviço, não estão tão predispostos a reclamar por eventuais degradações. No entanto, isto não deve ser encarado como regra, e poucos operadores recorrerão a este tipo de estratégias.

A utilização dos QoS *Brokers* associados a um sistema de gestão e monitoria de rede em tempo real, permite que as decisões tomadas pelos QoS *Brokers* sejam não só baseadas no registo de utilizadores e serviços activos mas também, mais importante, no estado real da rede. Contudo, estas decisões não podem também ser tomadas inteiramente com base no estado da rede, pois desta forma poder-se-á facilmente cair numa situação de violação de SLA. Mais uma vez, a decisão dos QoS *Brokers* dependerá da política do operador e do risco de violação de SLA que este pretende correr. Como regra geral, poderemos afirmar que os recursos associados a serviços de tempo real (S1 na Tabela 7) nunca devem ser sobre-reservados. Os recursos associados a serviços do tipo prioritário (S2 e S3) podem ter alguma sobre alocação, enquanto que os recursos associados a serviços do tipo de melhor esforço (S4, S5 e S6), podem ter muita sobre alocação.

Contudo, se o QoS *Broker* recorrer a informação de estado real da rede, a quantidade de sobre alocação permitida poderá ser variável e não fixa, permitindo cumprir os SLAs dos utilizadores, minimizando o risco de os violar.

5.5.7 Controlo de recursos em meios partilhados

A arquitectura proposta garante os recursos associados a cada serviço apenas a partir do nível 3, do IP. Para cada tecnologia de nível 2, existe um mapeamento entre os recursos IP e os recursos físicos. No entanto, muitas das tecnologias (nível 2) disponíveis poderão não implementar ou não ter disponíveis mecanismos de qualidade de serviço. Nos casos em que a tecnologia não suporta ou não tem implementados mecanismos de controlo de QoS, todo o tráfego, quer seja pertencente a serviços mais ou menos prioritários, concorrerá em igualdade no acesso ao meio de transmissão. Para que se continue a garantir a QoS de todos os serviços, em especial dos de tempo real, ter-se-á que garantir que o total de recursos reservados não excede a capacidade do acesso. Assim, cabe também aos QoS *Brokers*, impedir que o agregado das reservas autorizadas para aquele acesso não exceda o total da largura de banda disponível. Na prática, irá dar-se tratamento preferencial a todo o tráfego, incluindo até o de melhor esforço. Ou seja, para não prejudicar o tráfego prioritário, beneficia-se todo o restante. Por outras palavras, será também necessário reservar recursos para as ligações de "melhor esforço", de modo a que estas não prejudiquem as mais prioritárias. O QoS *Broker*, com base nos registos de autorizações de serviço concedidas em determinada célula, e com base em informação de utilização recolhida, quer no *router* de acesso, quer por elementos específicos de medição, sabe quais os recursos ocupados e quais os ainda disponíveis na célula em questão, e assim pode tomar decisões sobre pedidos futuros de acesso à célula. No entanto, uma vez que não há controlo efectivo dos recursos de nível 2, cada utilizador poderá tentar consumir ou "ocupar" mais recursos do que aqueles que o seu serviço permite. De facto, não havendo mecanismos de controlo de QoS no nível 2, não se poderá impedir que isso aconteça. Contudo, uma vez que logo à entrada da rede, no *router* de acesso, são implementadas políticas de controlo rígido de utilização de recursos, se um utilizador tentar utilizar mais recursos do que o seu perfil permite, estes serão policiados. O utilizador será desta forma obrigado a reduzir o ritmo de transmissão para os valores contratados, de modo a garantir a qualidade do seu serviço. Portanto, garante-se desta forma a qualidade de todos os serviços de todos os utilizadores.

A introdução de mecanismos de controlo de prioridades, tais como o 802.11e (*Wireless* LAN) ou o 802.1p/q (*Ethernet*) mapeando-os em serviços de nível IP (à semelhança do realizado no caso do TD-CDMA) resolve este problema de uma forma mais eficaz e com optimização da utilização dos recursos.

CAPÍTULO 6

AVALIAÇÃO DE DESEMPENHO DA ARQUITECTURA PROPOSTA

6.1 Introdução

A arquitectura apresentada no capítulo anterior e todos os seus elementos constituintes foram desenvolvidos, implementados e sujeitos a vários testes de validação (incluindo simulação) e conformidade. Esta secção é dedicada à avaliação da arquitectura proposta, tendo esta avaliação sido realizada em três vertentes distintas: uma decomposição analítica dos diferentes termos relevantes para o desempenho da arquitectura, simulação de comportamento da sinalização durante o *handover*, e avaliação de uma implementação prática. São apresentados alguns limites teóricos para o desempenho da arquitectura, assim como os limites rígidos impostos por esta. A simulação incidiu sobre o aspecto mais importante e fulcral desta arquitectura e da sinalização: o *handover*.

A implementação real (muito semelhante ao apresentado na Figura 23) desta arquitectura foi realizada com base em implementações para *Linux*, tendo sido utilizado

hardware especificamente desenvolvido para o suporte de TD-CDMA. Há parâmetros de desempenho bastante importantes que determinam de forma explícita a percepção dos utilizadores: o atraso de registo, tempo de autorização de serviço e, o mais importante de todos, o atraso de *handover*. Este último parâmetro terá que ser bastante reduzido de modo a que a mobilidade seja transparente para o utilizador e aplicações mais sensíveis. Todos estes tempos estão intrinsecamente limitados pelo desempenho das entidades intervenientes nos diversos cenários e que executam operações mais ou menos complexas e demoradas, dependendo do âmbito. Desta forma, antes de entrar na análise dos cenários compostos, vamos apresentar valores do desempenho das diversas entidades na execução de funções distintas. Estes valores representam um limite físico do desempenho da arquitectura, passível de ser melhorado, de acordo com os componentes utilizados (*hardware* e *software*). Todos os valores de tempos medidos e apresentados neste capítulo foram obtidos utilizando instâncias do *Ethereal* [Ethe] a operar nos diversos equipamentos. A sincronização dos relógios dos equipamentos foi obtida utilizando NTP (*Network Time Protocol*). Devido ao processo utilizado não ser completamente isento de interferências, os valores apresentados podem ser vistos como margens superiores para os valores reais. Na prática, a própria utilização de instâncias de *software* (Ethereal) sobrecarrega de certa forma as máquinas utilizadas, e por ser um tratamento por *software*, introduz algum atraso adicional na própria medida.

De seguida é apresentada a avaliação do processo do *handover* recorrendo à sua simulação.

6.2 Avaliação por simulação do processo de *handover* com QoS

De modo a validar a arquitectura, e em especial a sinalização de *handover* proposta, foi efectuado um estudo de simulação, recorrendo ao simulador de redes de pacotes, o ns2 [NS-2].

A Figura 37 apresenta o cenário de simulação utilizado para avaliar o processo de *handover*. Este cenário consiste num terminal móvel que se desloca no espaço entre duas estações base (*routers* de acesso). Estas duas estações base estão ligadas, cada qual, a um *router* do núcleo da rede. Existe um QoS *Broker* com ligação às estações base através dos *routers* do núcleo. Fazem ainda parte deste cenário, um *home agent* e um nó

correspondente, fazendo o papel de um servidor ao qual o terminal acede durante o seu processo de movimento e de *handover*.

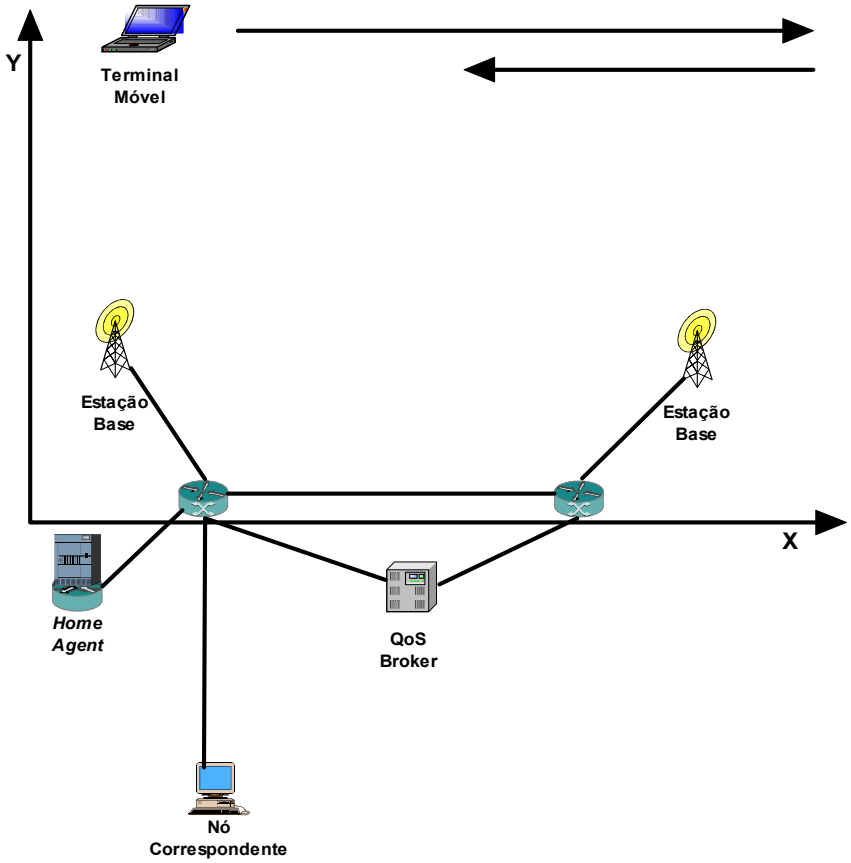


Figura 37: Cenário de simulação do processo de *handover*

Todas as ligações, na parte da rede fixa, foram definidas de modo a terem uma largura de banda de 5 Mbps e um atraso de 2 ms. No entanto, como veremos adiante, houve casos em que alguns destes parâmetros foram variados de modo a observar o comportamento da arquitectura em situações distintas. A parte rádio da arquitectura é baseada na implementação da norma 802.11, presente no simulador ns2.

O simulador de rede, ns2, foi devidamente adaptado de modo a incluir as entidades e funcionalidades requeridas, em termos de sinalização e controlo de *handover*, definidas nesta arquitectura.

Este estudo de simulação teve dois grandes objectivos:

- obter valores representativos dos tempos necessários à troca da sinalização e de execução do processo de *handover*;

- avaliar o desempenho do processo de *handover* enquanto o terminal recebe tráfego (de aplicações TCP e UDP).

6.2.1 Variantes do cenário de simulação

Foram realizadas várias simulações, tendo um cenário "base" de partida, e tendo sido feitas algumas variações de modo a melhor avaliar a influência de alguns parâmetros no processo de *handover*.

O cenário base de simulação (que foi sucintamente descrito atrás) tinha as seguintes características:

- A simulação tem uma duração de 50 segundos;
- O cenário é composto por uma área de 350x200 metros;
- A largura de banda das ligações fixas é de 5 Mbps;
- O atraso nas ligações fixas entre dois nós é de 2 ms;
- As estações base estão colocadas nos pontos [50;25] e [300;25];
- O terminal móvel está inicialmente localizado no ponto [10;185];
- Aos 5 segundos o terminal inicia uma deslocação em direcção ao ponto [340;185];
- Aos 35 segundos o terminal inicia uma deslocação em direcção ao ponto [10;185];
- O terminal móvel desloca-se a uma velocidade de 10 metros por segundo.
- Aos 3 segundos de simulação é iniciada uma aplicação (TCP ou UDP) que termina aos 47 segundos de simulação.
- O QoS *Broker* está programado para introduzir um tempo de processamento de *handover* de 17 ms.
- O fluxo TCP, pela própria natureza do TCP, ocupará o máximo da largura de banda disponível.
- O fluxo UDP foi dimensionado de modo a simular uma vídeo-chamada, gerando pacotes de 100 bytes (de dados) com um intervalo de 0.005 segundos, equivalendo a uma largura de banda média ocupada de cerca de 190 Kbps (dados e *overhead*).

As variações realizadas ao cenário base foram as seguintes:

- Cenário "afastamento eléctrico das estações base" (logo, do tempo de propagação) – As estações base que inicialmente se encontravam a 6ms, foram afastadas a 64ms;
- Cenário "nó correspondente a 50ms" – o nó correspondente, que inicialmente se encontrava ligado a um dos *routers* do núcleo a 2ms, foi afastado a 50ms;
- Cenário "terminal móvel a 80 metros por segundo" – a velocidade do terminal móvel foi incrementada para 80 m/s;
- Cenário "tempo de processamento do QoS *Broker* de 3ms" – o tempo de processamento da decisão de *handover* do QoS *Broker* foi reduzido para 3ms;
- Cenário "tempo de processamento do QoS *Broker* de 37ms" – o tempo de processamento da decisão de *handover* do QoS *Broker* foi incrementado para 37ms.

6.2.2 Resultados de simulação

Nesta secção são apresentados os resultados das várias variantes de simulação. Começa pela apresentação dos resultados do "cenário base". A seguir são apresentadas apenas as diferenças entre os cenários subsequentes e o "cenário base". São apresentados os resultados obtidos utilizando um fluxo TCP e os resultados obtidos por utilização de um fluxo UDP.

Na Tabela 10 são apresentados os instantes da simulação em que ocorre a troca das diversas mensagens de sinalização.

	Mensagem/acção	Tempo (s)	
		TCP	**UDP**
1	Anúncio de *Router*	15.3006	15.5022
2	Solicitação ao *Router para* actuação como *proxy*	15.303	15.5027
3	Início de processo de *handover*	15.3284	15.5037
4	Indicação de pedido de *handover*	15.3284	15.5037
5	Indicação de decisão de *handover*	15.35	15.5253
6	Resposta ao pedido de *handover*	15.3543	15.5297
7	Resposta de *handover* (actuação como *proxy*)	15.3606	15.536
8	Indicação de execução de *handover*	15.4315	15.6092
9	Início de replicação (*bicasting*) de pacotes e de temporizador	15.4569	15.6305
10	Confirmação de indicação de execução de *handover*	15.4569	15.6305
11	Estabelecimento de nova ligação (nível 2)	15.5477	15.6371
12	Anúncio de vizinho	15.5477	15.6371
13	Confirmação de actualização de endereço (e localização)	15.5546	15.6473

Tabela 10: Resultados de simulação das temporizações do processo de *handover*

Da análise aos resultados apresentados na Tabela 10 destacam-se as seguintes conclusões:

- A preparação do processo de *handover* (prévia à mudança de nível 2) leva menos de 150 ms (tempo decorrente entre a mensagem 2 e a mensagem 10);
- Todo o processo de *handover* (incluindo o *binding update* ao *home agent*) fica concluído em cerca de 250 ms;
- No caso da transmissão UDP que, ao contrário do TCP, não ocupa toda a largura de banda da ligação, estes tempos são ainda inferiores (130 ms e 230 ms respectivamente).

Há que realçar o facto de neste cenário de simulação, as mensagens de sinalização competirem nos diversos nós com a mesma prioridade do restante tráfego (TCP e UDP). Desta forma, todas as mensagens de sinalização são enviadas apenas quando o tráfego que já está nas filas de espera tiver sido enviado. Como consequência, a sinalização pode sofrer alguns atrasos.

Estes resultados permitem afirmar que, num terminal que se desloque a uma velocidade de 250 km/hora, o espaço de preparação do *handover* não excede os 11 metros, sendo todo o processo concluído em 17 metros. Ou seja, atendendo à velocidade considerada, a sobreposição de células rádio que é necessária é bastante reduzida. Daqui se

pode desde já concluir a viabilidade desta solução para uma arquitectura de mobilidade com suporte de QoS.

No entanto, para uma melhor análise do que se passa ao nível das aplicações, atentemos nas figuras seguintes, que mostram o comportamento dos fluxos de tráfego TCP e UDP em termos de largura de banda (instantânea e média), latência e *jitter*. No caso dos fluxos UDP é ainda possível fazer uma análise dos pacotes chegados fora de ordem.

A Figura 38 apresenta o gráfico da largura de banda instantânea do fluxo TCP entre o nó correspondente e o terminal móvel. Como foi apresentado atrás, nesta simulação ocorreu um *handover* aos 15,4 s (aproximadamente). Existe ainda um outro *handover* quando o terminal faz o caminho de regresso. Esse *handover* ocorreu aos 42,9 segundos (aproximadamente) de simulação. Podemos, embora não seja muito visível, constatar que existem dois picos na largura de banda instantânea logo após os *handovers*. Ora, isto deve-se ao facto de durante o *handover*, e devido ao próprio processo de *bicasting*, os pacotes sofrerem um pequeno atraso na chegada ao terminal, o que provoca uma diminuição na largura de banda que é compensada logo que a situação se regulariza. No entanto, verifica-se que o processo de *handover* é praticamente transparente para a aplicação, pois não se denota uma diminuição no débito, o que em caso de haver quebra momentânea da conectividade IP, seria evidente, com o TCP a diminuir drasticamente o débito. Este facto confirma a importância do *bicasting*, pois as conclusões deste processo sem *bicasting* são muito diferentes [Lina03].

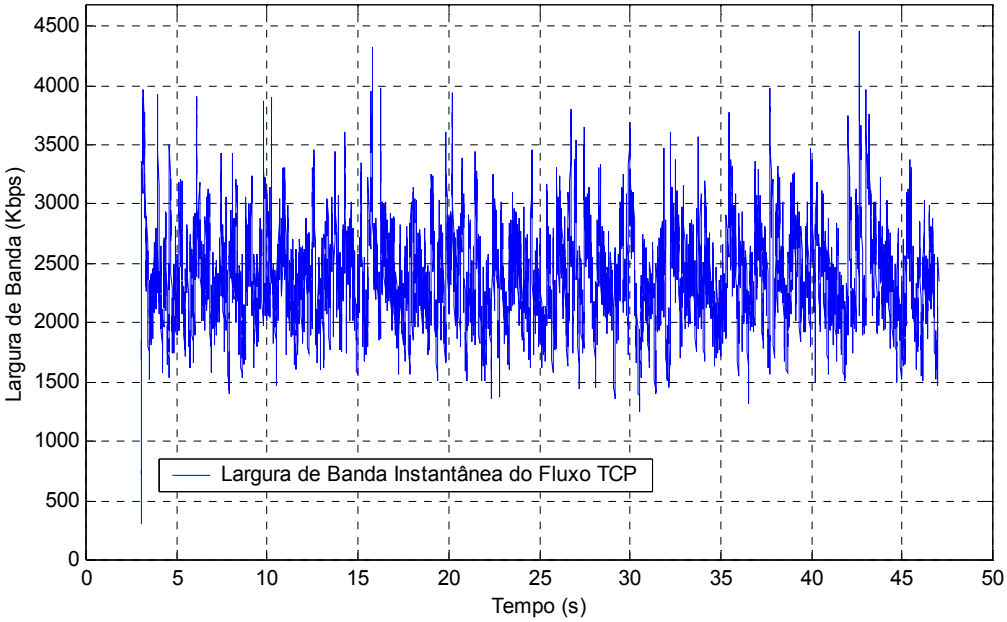


Figura 38: Largura de banda instantânea do fluxo TCP

A Figura 39 vem mostrar que, de facto a largura de banda média ocupada pelo fluxo TCP praticamente não varia, não se notando quaisquer efeitos particulares nos instantes em que decorre o processo de *handover*. Um facto curioso a registar, que nada tem a ver directamente com este processo, é o facto de a largura de banda média disponível na ligação sem fios (802.11) ser de cerca de 1,3Mbps.

Um último comentário a respeito da largura de banda média é o facto de se registar um pico no início da transmissão TCP. Este pico deve-se ao facto de o nó correspondente e o terminal estarem a uma distância eléctrica muito reduzida. Assim, com a chegada dos primeiros *acknowledges*, a janela de transmissão TCP aumenta muito rapidamente, fazendo com que a largura de banda suba de igual forma. O subsequente processo de adaptação à largura de banda efectivamente existente provoca o ondular seguinte, até que se dá a estabilização em torno da largura de banda efectivamente disponível.

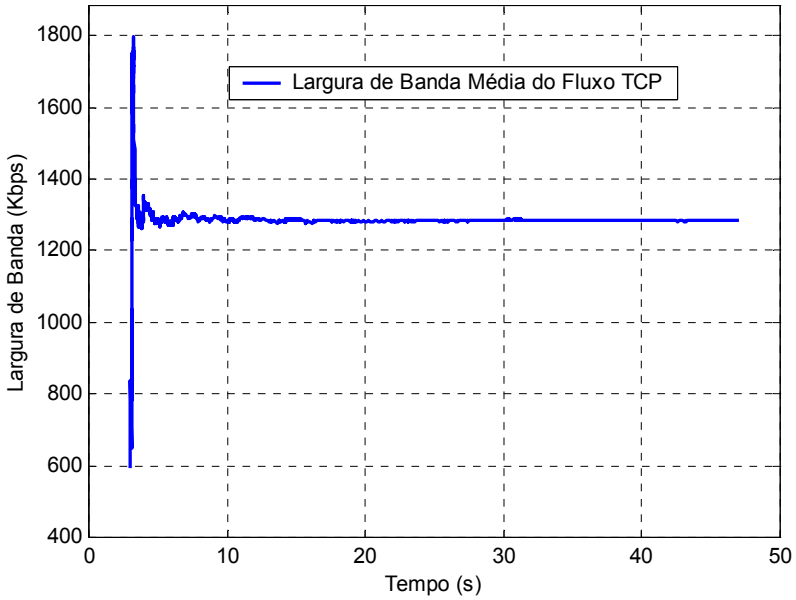


Figura 39: Largura de banda média do fluxo TCP

As Figura 40 e Figura 41 apresentam, respectivamente, a latência e o *jitter* sofridos pelos pacotes do fluxo TCP.

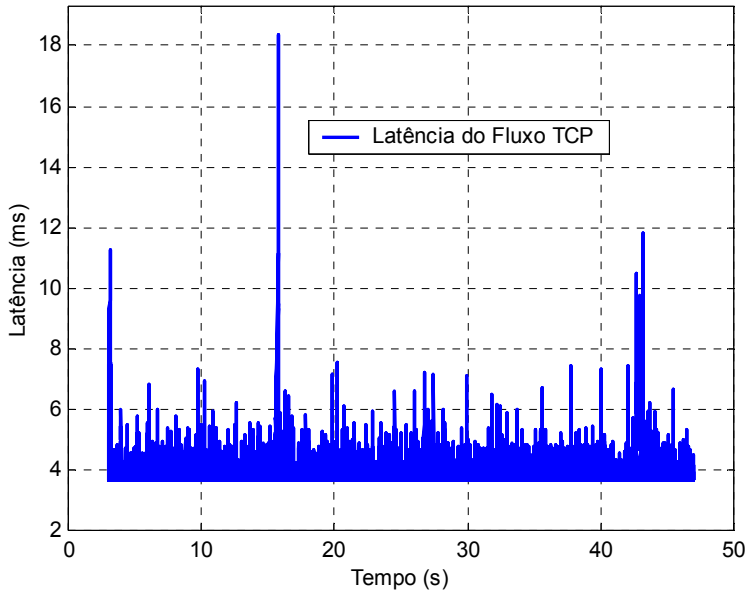


Figura 40: Latência do fluxo TCP

Podemos facilmente constatar a existência de 3 picos que se destacam. O primeiro, logo no início da transmissão do fluxo, deve-se ao facto de, no modo de funcionamento da mobilidade IPv6, o primeiro pacote que viaja do nó correspondente para o terminal móvel

ter de ser encaminhado através do *home agent*. O segundo (cerca dos 15,4 s) e terceiro picos (cerca dos 42,9 s), devem-se aos *handovers*, pois, como é natural, e devido ao próprio processo de *bicasting*, haverá pacotes que levaram mais tempo a chegar ao terminal, justificando um aumento da latência. Logo após a recepção dos pacotes que sofreram o *bicasting*, o terminal recebe pacotes que já não passaram por esse processo, e que portanto, estarão mais "próximos" dos anteriores. É desta forma que se justifica os picos de *jitter* negativos da Figura 41.

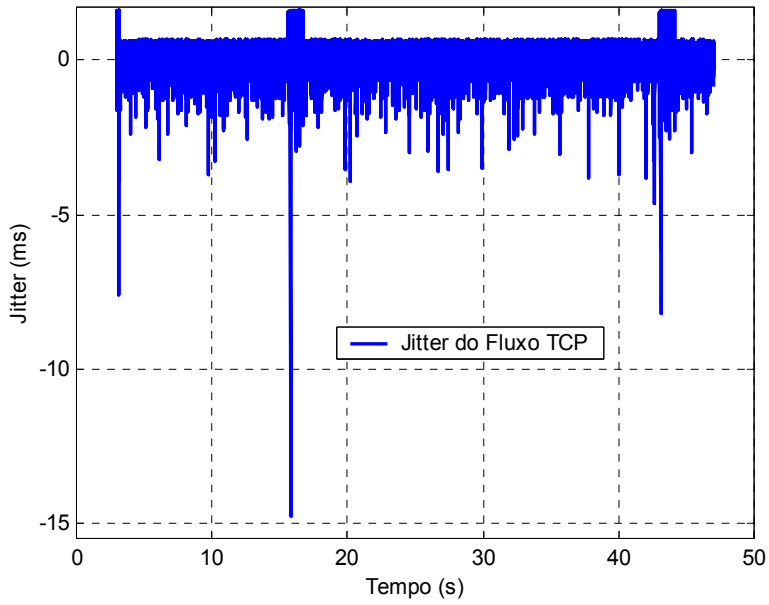


Figura 41: *Jitter* do fluxo TCP

As Figura 42 e Figura 43 apresentam os gráficos da largura de banda (instantânea e média) do fluxo UDP enviado do nó correspondente para o terminal móvel. A análise destas figuras revela um pico no instante inicial que é justificado pelo funcionamento da mobilidade IPv6. Os primeiros pacotes são encaminhados através do *home agent* enquanto que, após a recepção do primeiro pacote, o terminal móvel envia um *binding update* ao nó correspondente, que passa a enviar os pacotes directamente ao terminal. Durante uns escassos instantes, o terminal receberá pacotes que passaram pelo *home agent* em simultâneo com pacotes que já foram enviados directamente para o terminal. Assim, esta chegada simultânea de pacotes provoca um aumento da largura de banda ocupada nos instantes iniciais. Ainda na Figura 42 podemos observar mais dois picos na largura de banda instantânea. Estes dois picos devem-se aos *handovers* que se deram aos 15,5 e 42,6 segundos. Estes picos (antecedidos por uns ligeiros abaixamentos) devem-se ao processo

de *bicasting* que força a que existam alguns pacotes a chegarem com intervalos de tempo mais próximos do que aqueles com que foram gerados.

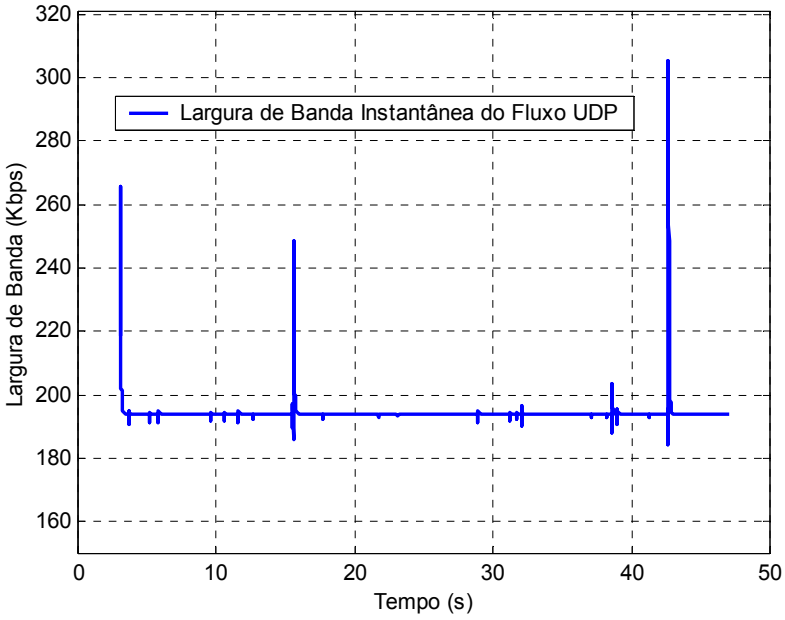


Figura 42: Largura de banda instantânea do fluxo UDP

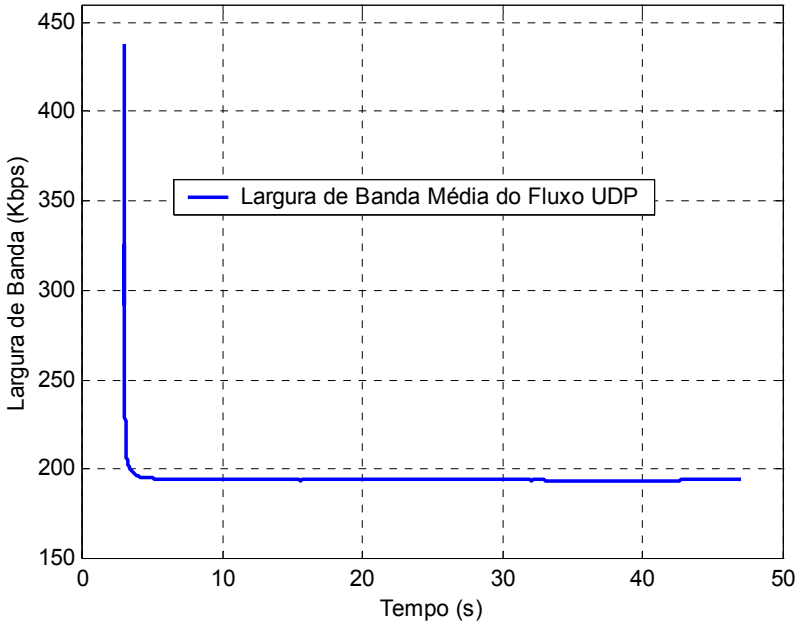


Figura 43: Largura de banda média do fluxo UDP

As Figura 44 e Figura 45 apresentam a latência e o *jitter* sofrido pelo fluxo UDP. Estas figuras vêm confirmar o que atrás foi dito em termos de largura de banda.

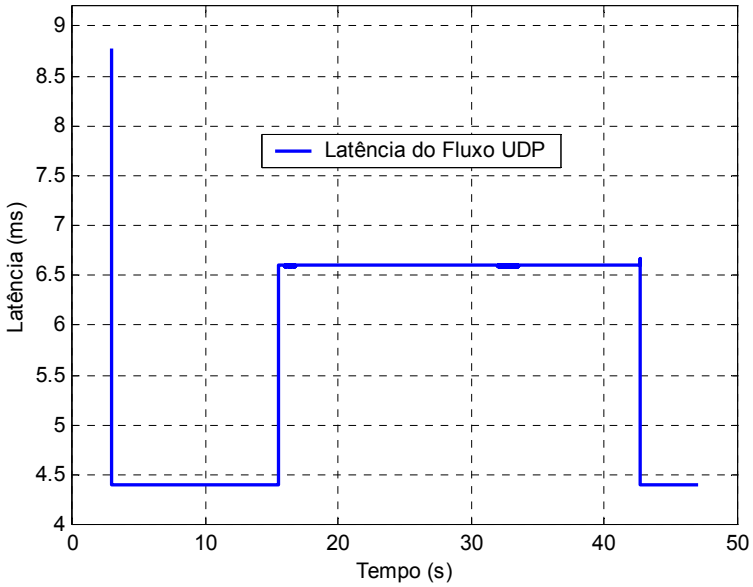


Figura 44: Latência do fluxo UDP

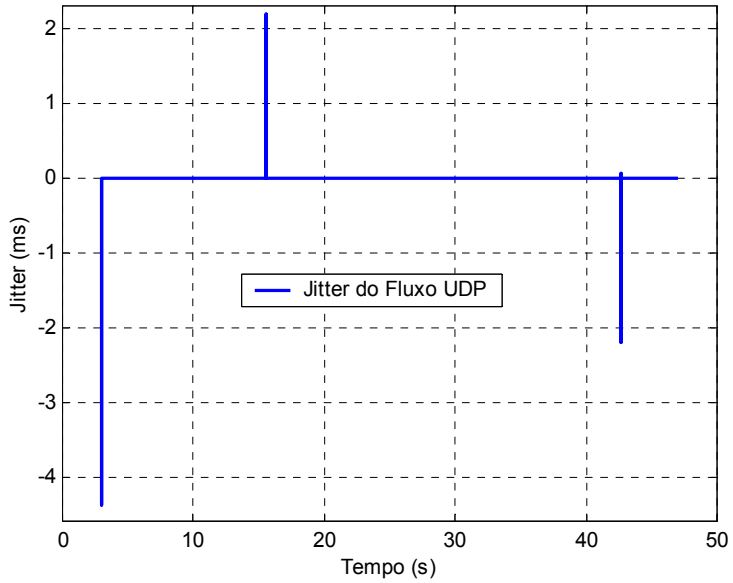


Figura 45: *Jitter* do fluxo UDP

Em termos de latência, existe um pico nos instantes iniciais, forçado pelo facto dos pacotes viajarem através do *home agent* antes de chegarem ao terminal móvel. Depois podemos observar que durante o tempo em que o terminal está ligado na segunda estação base, a latência cresce 2ms, relativos ao afastamento entre os *routers* de núcleo que os pacotes atravessam. A diferença entre a distância eléctrica entre os dois *routers* é a

responsável pelo *jitter* mostrado, sendo que, quando o terminal se move para a estação base mais distante do nó correspondente o *jitter* é positivo, e quando o terminal móvel regressa à estação base inicial, o *jitter* é negativo de valor simétrico (e igual ao afastamento eléctrico entre os *routers* de núcleo).

A Figura 46 apresenta os pacotes que chegaram fora de ordem durante o processo de *handover*. Os pacotes fora de ordem têm uma maior influência em aplicações UDP (normalmente aplicações de áudio e vídeo), especialmente se não houver armazenamento de informação. Ou seja, em caso de uma aplicação sem armazenamento, um pacote fora de ordem pode ser considerado um pacote perdido. A Figura 46 mostra que há pacotes fora de ordem apenas quando o terminal se desloca da estação base mais afastada para a estação base mais próxima do nó correspondente. Estes pacotes fora de ordem devem-se ao efeito do atraso sofrido pelos pacotes sujeitos a *bicasting*. Estes pacotes chegam ao terminal depois de alguns que foram entretanto enviados directamente do nó correspondente para o terminal móvel, após o *handover* e após o *binding update*.

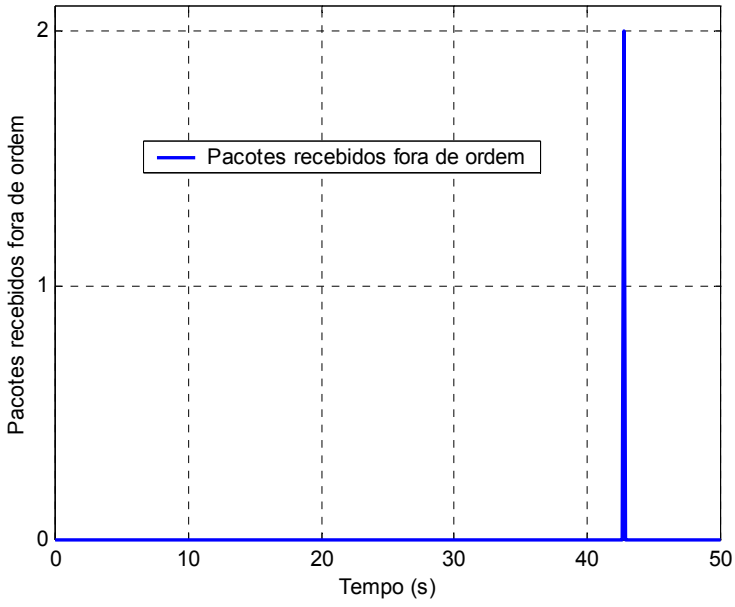


Figura 46: Pacotes do fluxo UDP chegados fora de ordem

De seguida é feita uma análise sucinta aos restantes cenários simulados.

Cenário "afastamento eléctrico das estações base"

As estações base que inicialmente se encontravam a 6ms, foram afastadas a 64ms.

Em termos da transmissão do fluxo TCP, os resultados mais significativos são os esperados à partida: uma redução da largura de banda média e um aumento da latência inerente ao afastamento quando o terminal muda para a estação base mais distante do nó correspondente. A redução da largura de banda ocupada pelo fluxo (observável a partir dos 16 segundos na Figura 47) é um reflexo natural do funcionamento do próprio TCP (altamente dependente das distâncias entre os nós envolvidos na comunicação).

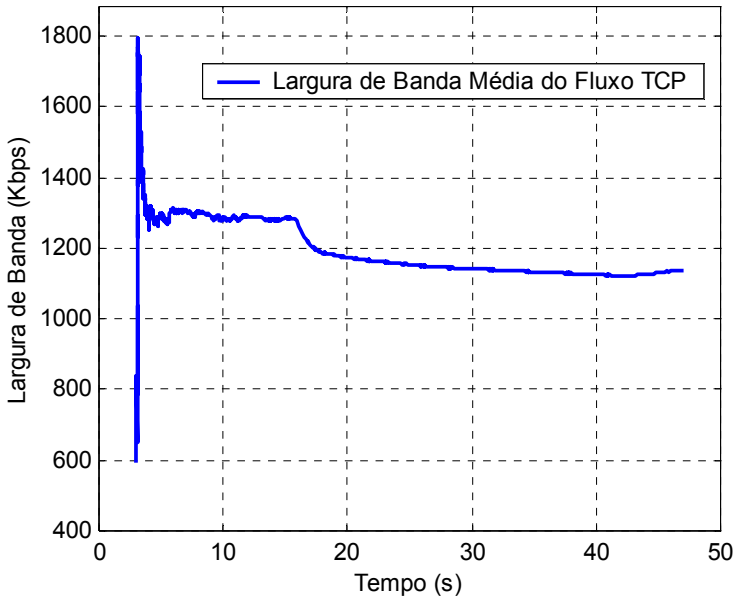


Figura 47: Largura de banda média do fluxo TCP com as estações base afastadas 64 ms

Em termos do fluxo UDP, o facto mais marcante deste afastamento é a quantidade de pacotes recebidos fora de ordem pelo terminal aquando do segundo *handover*. Neste caso, e devido ao afastamento e ao *bicasting*, há muitos pacotes a chegarem ao terminal após alguns pacotes que foram enviados já após o nó correspondente ter recebido o *binding update*. A Figura 48 ilustra os pacotes chegados fora de ordem.

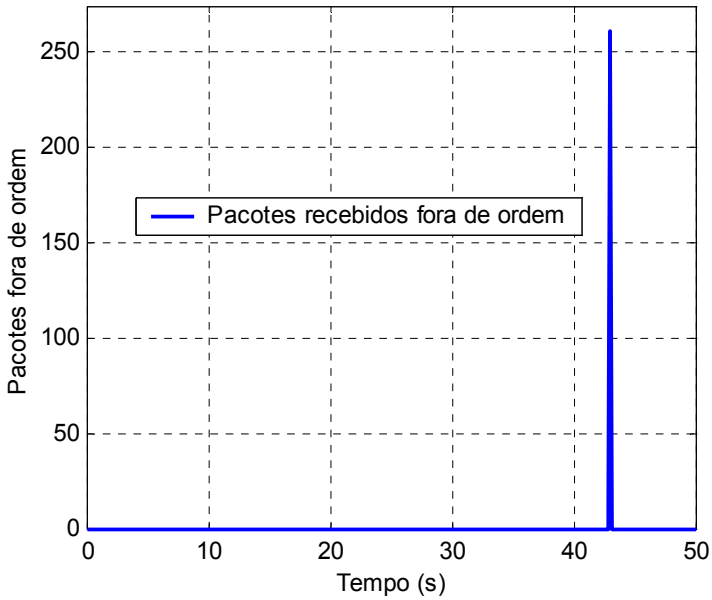


Figura 48: Pacotes do fluxo UDP chegados fora de ordem (estações base a 64 ms)

A Figura 49 e a Figura 50 mostram, respectivamente, os gráficos do *jitter* e latência sofridos pelos pacotes do fluxo UDP. As conclusões que se podem retirar destes gráficos são as mesmas que foram tiradas em relação aos gráficos apresentados nas Figura 44 e Figura 45, com a condicionante de o afastamento eléctrico entre as estações base ter aumentado 58ms.

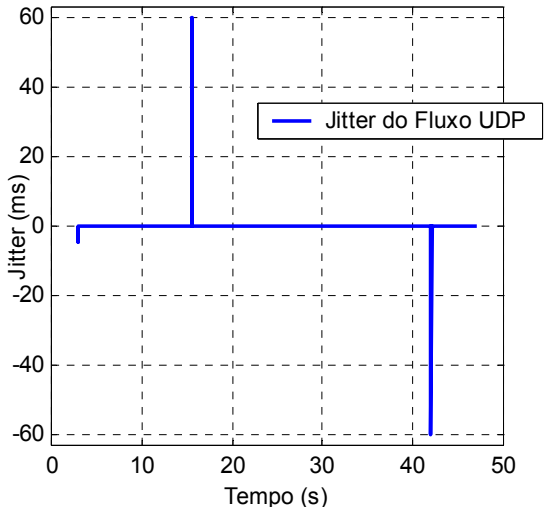


Figura 49: Jitter do fluxo UDP (estações base a 64 ms)

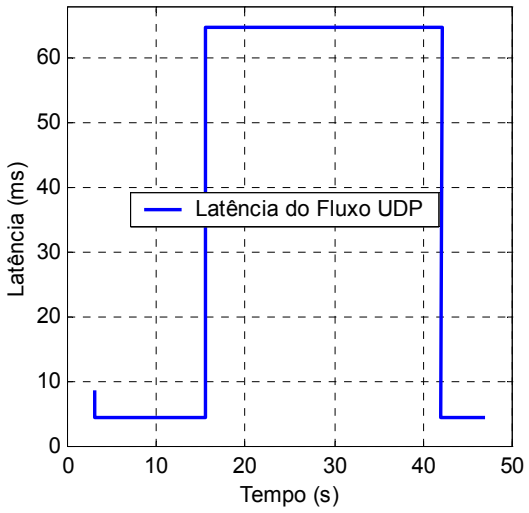


Figura 50: Latência do fluxo UDP (estações base a 64 ms)

Cenário "nó correspondente a 50ms"

O nó correspondente, que inicialmente se encontrava ligado a um dos *routers* do núcleo a 2ms, foi afastado para 50ms. Neste caso, como seria de esperar, os efeitos deste afastamento são o arranque mais lento da transferência do fluxo e a latência que é incrementada em 50 ms. A Figura 51 ilustra a largura de banda média ocupada pelo fluxo (de referir também a redução do débito máximo para 1,2Mbps) que, mais uma vez, é o reflexo do próprio funcionamento do TCP.

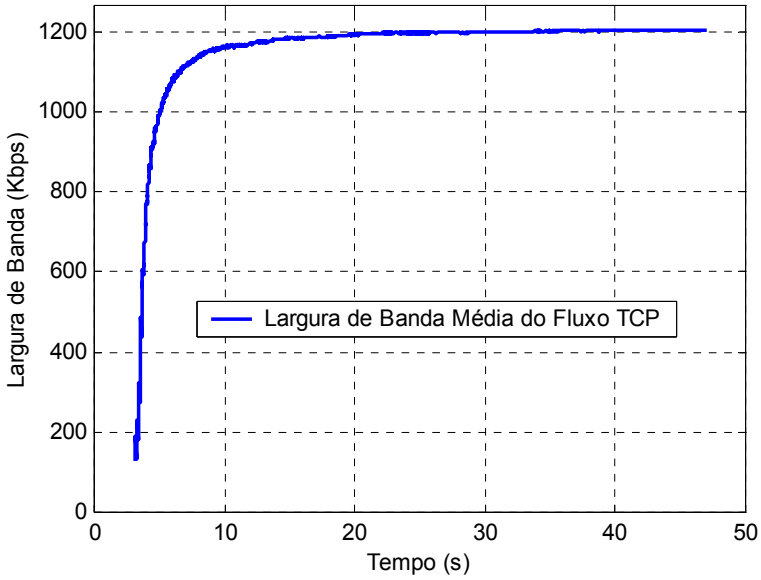


Figura 51: Largura de banda média do fluxo TCP com o nó correspondente a 50 ms

Cenário "terminal móvel a 80 metros por segundo"

A velocidade do terminal móvel foi incrementada para 80 m/s.

Neste cenário, apesar da velocidade do terminal ter incrementado oito vezes, para 288km/h, não houve qualquer distinção em relação ao cenário base, para além dos *handovers* se darem mais cedo no tempo. Estas conclusões são válidas quer para o caso do fluxo TCP quer para o fluxo UDP. A Figura 52 ilustra a latência sofrida pelo fluxo UDP neste caso, onde se pode ver claramente que os *handovers* se efectuam bastante mais cedo do que nos restantes casos.

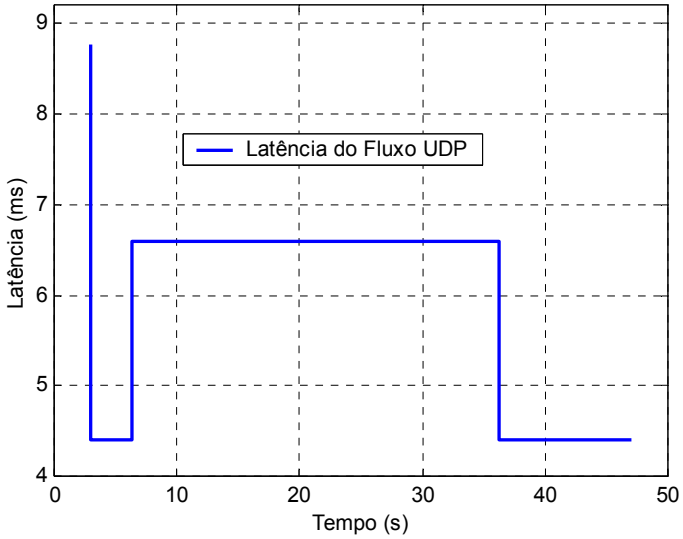


Figura 52: Latência do fluxo UDP quando o terminal se desloca a 80m/s

Cenário "tempo de processamento do QoS *Broker* de 3ms e 37ms"

Estes dois cenários não revelaram qualquer alteração em relação ao cenário base, para além da esperada redução (em 14 ms) da duração do processo de *handover* no primeiro caso e aumento (em 20 ms) no segundo caso. Nenhum destes casos prejudicou o desempenho da arquitectura, o que valida a opção de recorrer à figura do QoS *Broker* para o controlo do processo de *handover* com QoS. Assim, é de esperar que o comportamento da arquitectura não varie significativamente com a carga do QoS *Broker*.

Como comentário final aos resultados obtidos por simulação há um facto importante a reter que é o facto de não terem sido detectadas quaisquer perdas de pacotes, que representa o factor mais importante para se considerar um *handover* bem sucedido e transparente para o utilizador e aplicações. Registe-se que nem mesmo quando o terminal móvel se desloca a 80 metros por segundo (288 km/h) se verificam quaisquer perdas.

6.3 Avaliação da arquitectura em demonstrador real

A arquitectura aqui apresentada foi alvo de uma implementação real, baseada em PCs *desktop* e PCs portáteis, com o sistema operativo *Linux*. O cenário de demonstração utilizado encontra-se esquematizado na Figura 53. Este demonstrador foi utilizado para se

efectuarem medidas de desempenho a vários níveis. As mais relevantes em termos do trabalho realizado no âmbito desta tese são apresentadas nesta secção.

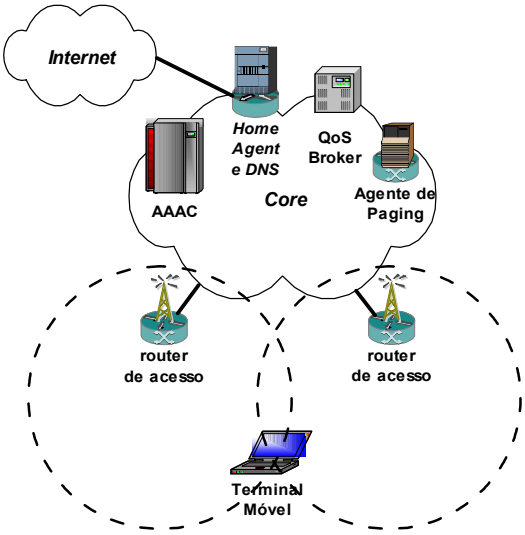


Figura 53: Demonstrador da arquitectura

6.3.1 Desempenho das principais entidades de rede

6.3.1.1 Sistema de AAAC

O sistema de AAAC, embora seja um dos elementos mais importantes numa arquitectura de operador, não tem, habitualmente, requisitos de tempos de resposta demasiado exigentes. Na arquitectura aqui proposta, a autorização de serviço e o *handover*, talvez os aspectos onde a rapidez da resposta é mais importante, estão delegados nos QoS *Brokers* e *routers* de acesso. Assim, o aspecto onde o desempenho do sistema de AAAC é mais importante em termos de qualidade de serviço observada pelo utilizador, é a autenticação. Desta forma, os parâmetros mais relevantes de desempenho do AAAC nos aspectos de prestação directa de serviço ao utilizador, é o tempo de resposta entre um pedido de autenticação e a resposta, e o tempo que passa desde o pedido de autenticação, até que a NVUP do utilizador seja passada ao QoS *Broker*. Esse valor, medido em condições reais no demonstrador da arquitectura proposta, é apresentado na Tabela 11.

Mensagem	Tempo de resposta em ms
Resposta a pedido de autenticação	56
Envio da NVUP ao QoS *Broker* após pedido de autenticação	40

Tabela 11: Parâmetros de avaliação de desempenho do sistema de AAAC

6.3.1.2 Router de Acesso

O *router* de acesso desempenha um papel fundamental na arquitectura proposta. Existem muitos parâmetros de avaliação de desempenho do *router* de acesso, sendo que uns influenciam de forma mais determinante o desempenho da arquitectura. Os valores considerados mais relevantes e limitativos do desempenho da arquitectura são:

- Tempo de reacção a pedido de autenticação – quando o cliente executa um pedido de autenticação, o atendedor de AAAC do *router* de acesso deve interceptá-lo e enviá-lo ao servidor de AAAC. Este tempo é o que decorre desde a chegada do pedido de autenticação, até ao envio desse pedido ao servidor de AAAC.
- Tempo de reacção a confirmação de autenticação – quando o servidor de AAAC responde ao pedido de autenticação essa resposta é enviada até ao terminal do utilizador, passando pelo módulo de AAAC do *router* de acesso. Este tempo é o que a mensagem demora a atravessar o *router* de acesso.
- Tempo de reacção a pedido de serviço – desde que o *router* de acesso recebe um pedido de novo serviço até que desencadeia o processo de consulta do PDP (QoS *Broker*).
- Tempo de reacção a instalação de política de serviço – tempo que decorre após uma decisão favorável do QoS *Broker* até que o *router* de acesso (PEP) instala a política do serviço pedido (este tempo é reduzido, pois corresponde à activação de uma linha nas IPTABLES do *Linux*). Este tempo é medido como o tempo que decorre entre a recepção do pacote COPS DEC (proveniente do QoS *Broker*) indicando a aceitação do serviço, e o envio do pacote COPS RPT (pelo *router* de acesso) a informar o QoS *Broker* que o serviço está instalado. Na realidade, a instalação do serviço ocorre antes do envio da resposta (só após a configuração realizada é que a resposta é enviada com o resultado da operação), pelo que o tempo efectivo será ainda menor do que o medido.
- Tempo de reacção a pedido de *handover* – este é o tempo que decorre desde que o terminal móvel indica ao seu actual *router* de acesso que pretende executar um *handover* até que o *router* de acesso envia essa indicação ao QoS *Broker* e àquele que é o candidato a *router* de acesso "novo".

- Tempo de reacção a indicação de resposta de *handover* – Existem duas situações em que o *router* de acesso deve responder a indicações de *handover*, uma na qualidade de *router* "antigo" e outra na qualidade de *router* "novo". O primeiro destes tempos (1), é o que decorre desde a recepção da resposta de *handover* do *router* "novo", até que essa resposta é enviada ao terminal móvel. O segundo (2), é o tempo que decorre desde a recepção da decisão do QoS *Broker* até que essa decisão é enviada ao *router* de acesso "antigo".
- Tempo de reacção a confirmação de execução de *handover* – este é o tempo que o *router* de acesso demora a confirmar a decisão do terminal móvel de execução *handover*.

Como é natural, todos estes tempos devem ser o mais baixos o possível. Convém no entanto frisar que todos estes factores são dependentes do desempenho do *hardware* (e qualidade do *software*) onde o *router* de acesso está implementado.

A Tabela 12 apresenta os valores de desempenho do *router* de acesso, medidos no demonstrador da arquitectura proposta.

Mensagem	Tempo de resposta em ms
Pedido de autenticação	170
Confirmação de autenticação	570
Pedido de novo serviço	750
Instalação de Serviço	0,4
Pedido de *handover*	7
Indicação de decisão de *handover* (1)	5
Indicação de decisão de *handover* (2)	7
Confirmação de execução de *handover*	12

Tabela 12: Parâmetros de avaliação de desempenho do *router* de acesso

6.3.1.3 QoS Broker

O QoS *Broker*, à semelhança do *router* de acesso, desempenha um papel determinante em termos de desempenho da arquitectura nomeadamente nas autorizações de serviço e no processo de *handover*.

Os parâmetros de resposta do QoS *Broker* mais determinantes em termos de influência no desempenho da arquitectura são:

- Tempo de recepção e tratamento da definição dos serviços – este é o tempo que o QoS *Broker* leva a processar a informação recebida do sistema de AAAC, contendo a definição dos serviços disponibilizados no seu domínio.
- Tempo de recepção e processamento de perfil de utilizador (NVUP) – este é o tempo que o QoS *Broker* leva a receber e processar os dados de um utilizador que se autenticou no domínio, até estar preparado para satisfazer pedidos de serviço para esse utilizador.
- Resposta a pedido de configuração – os *routers* de acesso (PEPs), quando são ligados, devem pedir ao QoS *Broker* (PDP) que os configure, tendo em conta a rede onde estão. Este tempo é aquele que o QoS *Broker* leva desde que recebe um pedido de configuração de um *router* de acesso até que responde a esse pedido. Para realizar esta operação, o QoS *Broker* necessita procurar na sua base de dados a informação relativa ao *router* de acesso em questão e após isso, procurar a configuração que lhe deve entregar. Uma vez que este processo só é realizado quando o PEP se liga pela primeira vez, não representa um valor crítico.
- Resposta negativa a pedido de serviço – este é o tempo que o QoS *Broker* leva até responder a um pedido de autorização de serviço no caso deste ser negado.
- Resposta a pedido de configuração de serviço proveniente do *core* – quando um serviço extremo-a-extremo é autorizado, na rede de destino, o QoS *Broker* local deverá configurar o *router* de acesso com os parâmetros relativos a esse serviço, de modo a que os pacotes cheguem ao terminal destino.
- Resposta positiva a pedido de serviço – este é o tempo que o QoS *Broker* leva desde que recebe um pedido de autorização de serviço, até que responde com a política que deve ser instalada no *router* de acesso para esse serviço.
- Resposta a pedido de revalidação de serviço – este é o tempo que o QoS *Broker* leva para renovar os serviços que estão autorizados. O *router* de acesso deve emitir pedidos de manutenção de serviço periodicamente, e o QoS *Broker* deverá renová-los. Em caso de serviços pré-pagos, poder-se-á, desta forma, cortar o serviço a utilizadores que tenham excedido o seu saldo.

- Resposta a pedido de *FastHandover* – este tempo é dividido em duas componentes. A primeira (1), no caso do QoS *Broker* desempenhar o papel de QoS *Broker* "antigo", é o tempo que decorre desde que o QoS *Broker* recebe indicação que o terminal pretende fazer um *handover*, até que comunica àquele que é o candidato a QoS *Broker* "novo" esse pedido. A segunda (2) componente, é o tempo que o QoS *Broker* "novo" leva até verificar se o *handover* é possível e o indica ao *router* de acesso "novo". Esta segunda componente inclui a detecção de duplicação de endereço, a verificação de disponibilidade de recursos no *router* "novo" e a construção da resposta que, em caso positivo, terá de levar a informação necessária à configuração do *router* de acesso "novo".

A Tabela 13 apresenta os valores de desempenho do QoS *Broker* medidos em campo de ensaio de validação da arquitectura.

Mensagem	Tempo de resposta em ms
Tempo de recepção de definição de serviços	385
Tempo de processamento de perfil (NVUP)	4
Pedido de configuração do *router* de acesso	54,3
Recusa de pedido acesso	0,733
Aceitação de pedido proveniente do *core*	1,5
Aceitação de pedido proveniente do acesso	3,8
Pedido de manutenção de serviço	0,1
FastHandover – Recepção de pedido e emissão de decisão	17,7

Tabela 13: Parâmetros de avaliação de desempenho do *QoS Broker*

6.3.1.4 Conclusões da avaliação real das diversas entidades

Os valores apresentados nas Tabela 11, Tabela 12 e Tabela 13 têm um carácter indicativo uma vez que as implementações utilizadas (baseadas puramente em *software*) e o suporte para as mesmas (PCs vulgares) não permitem um desempenho óptimo. Por outro lado, a carga a que o demonstrador foi sujeito também não se aproxima do que um cenário real de produção poderá atingir. No entanto, podemos verificar que os valores obtidos permitem concluir que o desempenho destas entidades se coaduna com o desempenho pretendido para um funcionamento correcto e eficaz do demonstrador e para permitir avaliar o desempenho da arquitectura apresentada.

6.3.2 Cenários chave

Após a validação das entidades mais relevantes do ponto de vista da arquitectura de QoS, vamos analisar o comportamento da arquitectura em cenários chave.

6.3.2.1 Tempo de registo

O atraso de registo é o tempo que decorre desde que o utilizador (após ligar o terminal) inicia o processo de autorização e o terminal fica pronto para o serviço. Neste contexto, existem atrasos de dois tipos: os atrasos das camadas inferiores (1 e 2), dependentes da tecnologia e do meio de transmissão e os atrasos das camadas superiores, dependentes do(s) protocolo(s) utilizado(s) (atraso provocado pelo envio e processamento das mensagens 1 a 4 apresentadas na Figura 32).

Assim, o tempo total de registo poderá ser indicado pela fórmula seguinte:

$$AtrasoRegisto = AtrasoNíveisInferiores + AtrasoNíveisSuperiores$$

O primeiro termo (*AtrasoNíveisInferiores*) é totalmente dependente da tecnologia e não depende do protocolo ou arquitectura utilizada, impondo assim um limite inferior inultrapassável. O segundo termo (*AtrasoNíveisSuperiores*) é dependente da velocidade das ligações utilizadas, do desempenho dos sistemas (*router* de acesso e sistema de AAAC), do protocolo e número de mensagens trocadas e, no caso de um utilizador em *roaming*, da distância eléctrica e velocidade de transmissão da ligação entre os sistemas de AAAC do seu domínio origem e do domínio onde se encontra:

$$AtrasoNíveisSuperiores = \sum AtrasoLigações + \sum AtrasoProcessamento + DistânciaAAACs$$

O primeiro termo (Σ*AtrasoLigações*), relativo ao atraso das ligações, entre o terminal móvel e o *router* de acesso, e entre o *router* de acesso e o sistema de AAAC pode ser normalmente desprezável, uma vez que a infra-estrutura de gestão é normalmente sobredimensionada em termos de recursos necessários e as velocidades de transmissão entre os elementos de controlo são normalmente elevadas (por exemplo, ligações *FastEthernet*). Desta forma, os atrasos de processamento (Σ*AtrasoProcessamento*) nos sistemas de AAAC (e atendedores de AAAC dos *routers* de acesso) podem ser os factores

dominantes, especialmente no caso de sobrecarga de processamento (muitos pedidos em simultâneo) ou subdimensionamento de capacidade de processamento (no caso de terem de gerir bases de dados muito grandes). No entanto, com o aumento de capacidade de processamento dos sistemas actuais, estes atrasos podem ser facilmente reduzidos. Assim, o factor mais restritivo de todos, poderá ser, em caso de utilizadores em *roaming*, a distância entre os sistemas de AAAC envolvidos (*DistânciaAAACs*). No entanto, este parâmetro não é dos mais importantes em termos de desempenho da arquitectura, desde que seja mantido em valores relativamente baixos em termos de percepção humana e semelhante ao que se experimenta actualmente em termos de redes celulares. Esta arquitectura é desta forma similar em desempenho aos sistemas actuais, do ponto de vista de tempo de registo.

6.3.2.1.1 Resultados obtidos

A Tabela 15 apresenta os valores dos tempos de autenticação medidos no cenário de demonstração (Figura 53). A numeração das mensagens respeita a numeração apresentada nas Figura 32 e Figura 33.

Nº	Δt (s)
1	0
2	0.169
3	0.209
4	0.224
5	0.800

Tabela 14: Desempenho da autenticação

Estes resultados confirmam a aptidão desta arquitectura e respectiva sinalização como base para uma futura arquitectura de próxima geração.

6.3.2.2 Tempo de autorização de serviço

O tempo de autorização de serviço é o tempo decorrido desde que o utilizador decide usufruir de determinado serviço até que os recursos estejam disponíveis para isso. Na Figura 32, este será o tempo decorrido entre as mensagens 5 e 11. Este tempo é composto essencialmente pelos atrasos nas ligações (i) entre os terminais (origem e destino) e os seus *routers* de acesso, (ii) entre os *routers* de acesso e os QoS *Brokers* e (iii) entre os dois *routers* de acesso (distância eléctrica entre ambos) e ainda pelos atrasos de processamento nos *routers* de acesso (nos módulos gestores de QoS) e QoS *Brokers*.

$$AtrasoInícioSessão = \sum AtrasoLigações + \sum AtrasoProcessamento + DistânciaARs$$

O primeiro destes termos (Σ*AtrasoLigações*) é normalmente muito reduzido (e é de difícil medição) uma vez que os terminais e os *routers* estão normalmente muito próximos, o mesmo se passando entre os *routers* e os QoS *Brokers* (e neste caso as ligações são normalmente sobredimensionadas). Quanto aos tempos de processamento (Σ*AtrasoProcessamento*)(que foram medidos nas secções anteriores), pode-se repetir a ideia apresentada na secção anterior relativamente à capacidade de processamento dos dispositivos de hoje em dia, ou seja, o aumento da capacidade de processamento deixa antever que não haverá qualquer limitação a este nível. Pode ainda acrescentar-se que num cenário de produção, os dispositivos utilizados serão ainda mais eficazes e poderosos, podendo eventualmente realizar diversas tarefas em *hardware* específico. Assim, uma vez mais, a distância eléctrica entre a origem e o destino da ligação (*DistânciaARs*) poderá ser o factor com maior preponderância no atraso total. No entanto, este é um limite físico da rede que não depende de todo da arquitectura proposta e não é passível de ser reduzido em termos arquitecturais.

6.3.2.2.1 Resultados obtidos

A Tabela 15 apresenta os valores de tempo de sinalização na autorização de serviço medidos no demonstrador (Figura 53). A numeração das mensagens está de acordo com a apresentado na Figura 32 e Figura 33.

Nº	Δt (s)
5	0
6	0.754
7	0.767
8	3.494

Tabela 15: Desempenho da autorização

Os resultados obtidos demonstram a validade da arquitectura e desempenhos que ao nível dos esperados em cenários de produção. A diferença de tempos entre a mensagem 7 e a mensagem 8 deve-se a duas componentes distintas: o tempo de configuração do *router* de acesso e os instantes em que a aplicação do cliente gera tráfego. Devido à utilização de sinalização implícita, a mensagem 8 é um pacote de uma aplicação do cliente que foi apenas gerado naquele instante.

6.3.2.3 Atraso de Handover

O atraso no *handover* é, sem dúvida, o mais crítico de todos os atrasos, em termos de sensibilidade e percepção do utilizador. O *handover* terá de ser efectuado muito rapidamente de modo a ser transparente para o utilizador e também para as aplicações, especialmente as aplicações multimédia. O atraso total do processo de *handover* é composto por diversos factores independentes, sendo que alguns apresentam limites rígidos inalteráveis, enquanto que outros podem ser passíveis de serem reduzidos ou desprezados: atrasos de transmissão, atrasos de computação ou processamento, e atrasos de *handover* de nível 2.

$$AtrasoHandover = \sum AtrasoTransmissão + \sum AtrasoComputação + \sum HandoverNível2$$

Os atrasos de transmissão ($\Sigma AtrasoTransmissão$) são compostos pelos atrasos das ligações entre os diversos elementos: terminal móvel e *routers* de acesso, *routers* de acesso e QoS *Brokers*, entre os QoS *Brokers* envolvidos e entre os *routers* de acesso "novo" e "antigo".

O tempo total de *handover* é o tempo passado desde a mensagem 3 (Figura 34) até à mensagem 12. No entanto, o tempo que o terminal móvel está sem conectividade e sem recursos atribuídos é apenas o tempo despendido no *handover* de nível 2 ($\Sigma HandoverNível2$), ou seja, o tempo que o terminal demora a mudar de uma célula para outra (tempo passado entre as mensagens 10a e 12). Este é um comportamento fundamentalmente associado ao desempenho da camada física dos dispositivos.

6.3.2.3.1 Resultados obtidos

A Figura 53 ilustra o cenário utilizado para fazer as medições do desempenho da arquitectura no que diz respeito ao *handover*. A Tabela 16 apresenta os resultados obtidos, mais relevantes do ponto de vista do *handover*. No demonstrador foi utilizado apenas um QoS *Broker*. Desta forma, a mensagem 5 definida nas Figura 34 e Figura 35, não existe.

Nº	Δ (s)		Nº	Δ (s)
2	0		7	1,087
3	1,006		8	1,114
4a	1,029		9	1,126
4b	1,065		10	1,144
6	1,086		12	1,257

Tabela 16: Tempos da sinalização de *handover*

6.3.2.3.2 Latência Geral Durante o processo de *Handover*

Os resultados medidos mostram que o tempo de preparação do *handover* se situa na ordem dos 138ms (diferença de tempos entre as mensagens 3 e 10). Se considerarmos que o terminal se está a deslocar a 250km/hora, toda a negociação prévia do *handover* se realiza em apenas 9,6m. Atendendo às dimensões das células rádio, e às suas sobreposições (várias dezenas de metros), podemos concluir que o desempenho da arquitectura proposta é adequado a ser utilizado em cenários reais.

Dos resultados observados, o tempo de todo o processo de *handover* (mensagens 3 a 12), relevante do ponto de vista do terminal que se está a mover, é cerca de 250ms, que é um valor dentro do parâmetros aceitáveis.

6.3.2.3.3 Pacotes perdidos durante o *Fast Handover*

Os testes e medições realizados permitiram avaliar que, utilizando aplicações correntes baseadas em UDP (para voz e vídeo), graças à preparação do *handover* antes da sua realização (*make before break*) e graças ao mecanismo de *bicasting* implementado no *router* antigo, o número de pacotes perdidos não ultrapassa um a dois por aplicação, havendo mesmo casos em que não se perde qualquer pacote. A perda de pacotes ocorre apenas aquando do *handover* para a rede *Ethernet*. Verificou-se que a perda de pacotes ocorre devido ao tempo que a própria interface física *Ethernet* do PC *Linux* leva a ficar disponível, partindo de uma situação sem ligação.

6.3.2.3.4 Atraso na preparação do *handover* introduzido pela componente de QoS

Foi feito um teste em que toda a componente de QoS (envolvendo os QoS *Brokers*) foi omitida do processo de *Fast Handover*. Esse teste permitiu avaliar o tempo de processamento extra que é introduzido pela componente de QoS na preparação do *handover*. Os resultados obtidos mostram que esse atraso extra se situa na ordem dos 8ms. Atendendo a que mesmo durante este tempo o terminal continua ligado na célula actual (e

que portanto não implica uma perda de serviço) e que com este processo se garantem os recursos necessários na célula para onde o terminal se deseja mover, pode considerar-se que este tempo é bastante satisfatório e que não compromete a solução proposta.

6.3.3 Testes baseados na percepção humana

De forma a avaliar a viabilidade desta arquitectura do ponto de vista da utilização real por humanos, foram utilizadas diversas aplicações representativas de vários tipos de serviços e foi recolhida a opinião dos utilizadores nas diversas fases.

As aplicações utilizadas foram telefonia IP, vídeo telefonia, *streaming* de áudio e vídeo, transferência de dados (mgen, ftp e http) e jogos multi-utilizador (quake 2) [IDSO].

Apesar de algumas destas aplicações serem de difícil teste enquanto o utilizador se move (por exemplo jogar quake), a percepção geral dos utilizadores foi a de que a arquitectura e o demonstrador conseguem suportar a transparência desejada em termos de mobilidade e QoS, uma vez que os utilizadores não conseguiram detectar qualquer problema ou interferência aquando dos instantes de *handover*.

Para a realização destes testes foram escolhidos estudantes com hábitos de utilização da *Internet*, contudo sem qualquer conhecimento da rede e da arquitectura que estavam a utilizar. Estes utilizadores foram sujeitos a situações de *handover* transparente, de *paging* e foram cobrados (virtualmente) de acordo com o perfil de QoS dos serviços utilizados e da própria utilização da rede. A rede foi sujeita a condições de sobrecarga forçada para se poder aferir a eficácia da diferenciação de serviço e de perfis. Conforme o esperado, os utilizadores com perfis de QoS de menor prioridade foram os que registaram queixas de degradação de serviço.

Os parágrafos seguintes descrevem testes formais realizados ao sistemas, e que foram seleccionados de forma a cobrirem situações de pior-caso para o comportamento da arquitectura, dadas as limitações logísticas existentes no demonstrador.

Teste 1 - Dois utilizadores, um com um perfil de serviço "*gold*" e outro com um perfil de serviço "*bronze*" acederam ao mesmo conteúdo (um fluxo áudio mp3) em condições de carga na rede. A aplicação que reproduziu o fluxo áudio não tinha qualquer armazenamento (*buffering*) temporário. O utilizador com perfil "*gold*" realizou vários *handovers* com QoS e AAAC enquanto a música tocava, sem que ele conseguisse detectar os instantes de *handover*, pois não houve qualquer perda de pacotes (mesmo com a sobrecarga efectuada na rede). Todos os *bytes* enviados e recebidos foram tarifados de

acordo com o DSCP, representativo do serviço utilizado. O utilizador com perfil "*bronze*" realizou o mesmo teste. Devido ao seu perfil ser de prioridade inferior e a rede estar sobrecarregada (com tráfego artificial), este utilizador detectou diversos cortes (mesmo sem efectuar *handovers*). No entanto, também nesta situação, o utilizador foi incapaz de detectar os instantes dos *handovers*, ou seja, ou níveis de QoS recebidos foram uniformes. Devido ao facto de este utilizador ter utilizado um perfil de serviço de menor prioridade, e de ter recebido menos *bytes* (alguns perderam-se antes de serem entregues) a sua tarifação foi inferior à do utilizador com o perfil "*gold*". De notar, no entanto, que os factores de custo de utilização podem ser dependentes de outros aspectos que não apenas os níveis de QoS. No limite, um utilizador "preferencial" poderá ter um serviço melhor a um custo mais reduzido que um outro utilizador "menos preferencial" (podemos ter um "grande" cliente empresarial como exemplo de um utilizador preferencial. Este é um aspecto associado aos problemas discutidos na secção 2.3.6 de problemas comerciais).

Teste 2 - Para aferir a real distinção e diferenciação de serviço e de QoS, foi realizado um teste em que dois utilizadores com perfis distintos ("*bronze*" e "gold") visualizaram um fluxo de vídeo simultaneamente, enquanto alguns *handovers* foram realizados. O cenário utilizado era composto por duas redes de acesso *wireless* (IEEE 802.11b e uma rede *Ethernet*. Inicialmente, os dois utilizadores, cada qual com o seu terminal, estavam ligados em cada uma das redes *wireless*. O fluxo de vídeo utilizado tinha características de largura de banda muito exigentes (cerca de 4 Mbps). Atendendo a que a largura de banda efectivamente disponível numa ligação 802.11b é apenas de cerca de 5,5 Mbps é facilmente dedutível que não há suficiente largura de banda disponível para acomodar os dois fluxos (*unicast*) na mesma célula *wireless*. Assim, enquanto cada utilizador estava numa rede de acesso *wireless* distinta não se observou qualquer anomalia na visualização do fluxo de vídeo em qualquer dos terminais. O passo seguinte consistiu em mover o utilizador com o perfil "*gold*" para a proximidade da rede de acesso *wireless* ocupada pelo utilizador com perfil "*bronze*". Ao perder sinal relativamente à célula que ocupava e detectando uma outra célula com melhor condições de sinal, o terminal do utilizador com o perfil "*gold*" realizou um *handover* rápido com QoS e AAAC para a célula ocupada pelo outro utilizador. Uma vez que a largura de banda disponível não era suficiente para ambos os fluxos, detectou-se que o utilizador com o serviço de perfil "*bronze*" começou a não receber todos os pacotes, acabando por forçar a aplicação de

visualização do vídeo a não exibir o filme. Em contrapartida, o utilizador com o serviço de perfil "*gold*" realizou o *handover* sem notar qualquer interferência, e continuou a visualizar o vídeo em condições perfeitas. Após isto, este utilizador, moveu-se para a rede de acesso Ethernet. Mais uma vez o *handover* foi realizado sem qualquer interferência ao nível da visualização do vídeo, e uma vez que os recursos na rede *wireless* onde o utilizador com o perfil serviço de "*bronze*" aumentaram, este viu a sua aplicação de vídeo a mostrar de novo o filme. O último passo deste teste consistiu em forçar um *handover* do utilizador com perfil "*bronze*" para a rede Ethernet. O *handover* foi realizado sem que o utilizador pudesse registar qualquer interferência na visualização, e uma vez que os recursos disponíveis na rede Ethernet eram suficientes para acomodar os dois utilizadores com os dois fluxos de vídeo, ambos puderam continuar a observar os vídeos sem qualquer problemas.

Teste 3 - Foram realizados testes com perfis de serviço "*gold*" mas com limitação de largura de banda. Nestes casos, verificou-se que sempre que os utilizadores tentaram utilizar aplicações que exigiam mais largura de banda do que aquela que o seu perfil permitia, mesmo tendo alta prioridade, havia perdas de pacotes correspondente aos pacotes descartados pelos *routers* de acesso. Ou seja, o serviço tinha de facto prioridade elevada e não era afectado pela introdução de tráfego concorrente menos prioritário, mas também não permitia a utilização de mais recursos do que os especificados no perfil.

É de salientar que nestes testes, foram envolvidas todas as entidades e módulos definidos na arquitectura apresentada, à excepção dos agentes e módulos de *paging*. Também se deve notar que estes testes foram realizados em ambiente semi-controlado, e que em alguns casos, problemas associados com as placas de rede (nomeadamente passarem de modo *stand-by* para modo activo) eram aparentes. Obviamente que estas situações, tendo sido devidamente identificadas, e sendo reprodutíveis, não são de relatar neste ponto.

6.4 Conclusões

Os resultados obtidos, quer por simulação quer em ambiente de demonstração real, permitem concluir que a arquitectura apresentada é transparente para o utilizador e permite o suporte de um conjunto vasto de serviços de rede de uma forma flexível e realista em termos de cenário reais em ambientes de produção. É ainda de salientar a boa aproximação

entre os valores dos tempos do processo de *handover* obtidos no demonstrador real e por simulação. A diferença entre os valores de tempo de *handover* obtido por simulação (cerca de 150ms para tráfego TCP e 130ms para tráfego UDP) e o valor em cenário real (cerca de 140ms) pode ser explicada devido a diversos factores: (i) o cenário de simulação não é exactamente igual ao cenário real, nomeadamente devido à existência no cenário de simulação de dois *routers* adicionais entre as estações base; (ii) no cenário de simulação, o tráfego TCP e UDP tem a mesma prioridade que o tráfego de sinalização – no demonstrador real, a sinalização tem uma classe de serviço independente; (iii) as entidades presentes no simulador, são modulações de entidades reais, mas que podem apresentar algumas diferenças. Em relação ao tempo total do *handover*, os resultados de simulação e os obtidos no demonstrador são ainda mais próximos.

A análise feita nesta secção recaiu sobre os aspectos mais críticos da arquitectura que podem influenciar o desempenho das entidades ao nível da QoS. Não foram focados os aspectos inerentes ao fornecimento de serviços com qualidade diferenciada (baseados na arquitectura DiffServ), mas estes foram implementados com sucesso no demonstrador. O demonstrador permitiu verificar o fornecimento simultâneo de serviços com características distintas. Foram realizados testes com introdução de tráfego concorrente, de modo a verificar que os serviços mais prioritários não são degradados. Por fim, foram também realizados testes em que se verificou que a implementação não permite que as aplicações utilizem recursos que os serviços contratados não contemplem.

CAPÍTULO 7

CONCLUSÕES

Neste capítulo são apresentadas as principais conclusões do trabalho efectuado e descrito nesta Tese. É também feita uma análise sumária a propostas de arquitecturas alternativas para redes de próxima geração. Por fim é apresentada uma breve descrição de como parte deste trabalho está a ser utilizada para o desenvolvimento de produtos reais, assim como os tópicos em aberto, e que se consideram importantes como propostas de trabalho futuro.

7.1 Principais Conclusões

A provisão de serviços multimédia em redes heterogéneas de comutação de pacotes é um desafio premente. Esta problemática reveste-se de diversas envolventes, cuja relevância depende fortemente do tipo de cenários que se pretendem atacar.

Por exemplo, um ambiente de operador de fornecimento de serviço é certamente mais complexo que um ambiente de *campus* ou de *intranet*. Os avanços em termos de capacidade de processamento e de capacidade de comunicação, acompanharam os desenvolvimentos nas aplicações, também mais ricas e mais exigentes em termos de

recursos necessários. Nesta tese foram abordadas as questões, relacionadas com aspectos de QoS, consideradas pertinentes num ambiente de fornecimento de serviço, em redes de próxima geração. Foi discutida a problemática do fornecimento de serviço, em várias vertentes, e foram apresentadas tecnologias e soluções que podem contribuir para suportar mecanismos de diferenciação de QoS em redes de próxima geração. Foi dado especial ênfase aos aspectos críticos de suporte de QoS em ambientes com mobilidade.

Neste sentido, foi proposta uma arquitectura de rede de próxima geração, dotada de funcionalidades que permitem o fornecimento de serviços diversificados, em ambientes com infra-estruturas de acesso heterogéneas. Esta arquitectura pode já ser considerada como uma arquitectura preliminar de quarta geração, com suporte de banda larga e mobilidade heterogénea. O protocolo IP, na sua versão 6, é utilizado como elemento fundamental na agregação das diferentes tecnologias de níveis inferiores. Toda a sinalização, controlo, e acesso aos recursos são efectuados recorrendo apenas a protocolos suportados pela camada IP. Este foco no IP cria um nível de abstracção relativamente às camadas inferiores, pelo que deverão existir funções em diversos elementos que mapeiem a informação de controlo presente no nível IP em diferentes aspectos de níveis inferiores.

O grande objectivo desta arquitectura é dar suporte de QoS em simultâneo com o suporte a mobilidade entre diversas tecnologias, de uma forma totalmente transparente para aplicações e utilizadores, e sob o controlo do operador. Para que isto aconteça foi necessário integrar três campos habitualmente separados como são a mobilidade, a Qualidade de Serviço e AAAC, de uma forma eficiente. A simplicidade e potencialidade desta arquitectura resultam da utilização quase exclusiva de normas e protocolos definidos no âmbito do IETF, em detrimento de protocolos dependentes de tecnologias específicas. Foram contudo realizadas algumas modificações a vários protocolos de modo a os tornar úteis nos cenários pretendidos. Esta arquitectura recorre a IPv6 Móvel (MIPv6) com optimizações de *Fast Handover* a correr sobre uma infra-estrutura baseada em *DiffServ* e controlada por QoS *Brokers*, e com um sistema de AAAC capaz de lidar com cenários complexos de próxima geração. Todos estes protocolos foram modificados de forma a interoperarem entre si.

Recorrendo à utilização do simulador de redes, ns2, o aspecto à partida eventualmente mais limitativo desta arquitectura, o *handover*, foi simulado. Os resultados

obtidos permitem validar as escolhas tomadas, para múltiplos cenários. Estes resultados estão também perfeitamente alinhados com os resultados obtidos num demonstrador real.

A arquitectura apresentada foi implementada e demonstrada utilizando três tecnologias de acesso distintas: LAN, *Wireless* LAN e TD-CDMA. No entanto, a arquitectura é suficientemente genérica para suportar qualquer outra tecnologia de acesso. Os testes realizados sobre uma implementação protótipo desta arquitectura vieram provar que será possível utilizar em cenários futuros as ideias e conceitos aqui desenvolvidos. Os resultados obtidos mostraram bons resultados de *handover*, utilizando aplicações multimédia, mesmo tendo PCs vulgares para suportar a maioria dos elementos da rede. Os tempos de resposta são suficientemente reduzidos, evitando que os utilizadores se apercebam ou notem qualquer interferência ou degradação aquando do processo de *handover*. Apesar dos resultados apresentados terem sido obtidos em cenários de carga baixa, é de notar que a rede de testes não era de todo uma referência em termos de desempenho, que poderá facilmente ser melhorado com hardware adequado.

Esta arquitectura, baseada na noção de serviço prestado ao utilizador, tem a vantagem de utilizar os mesmos elementos independentemente da tecnologia de acesso, mas permitindo optimizações posteriores nos níveis físicos. Conceptualmente, trata-se de uma arquitectura aberta e flexível, oferecendo uma separação clara entre a tecnologia e os domínios administrativos, mantendo a capacidade de fornecer serviços específicos a utilizadores específicos. Esta abordagem facilita a introdução de diferentes modelos de fornecimento de serviço, uma vez que desagrega a noção de serviço (associada a um contrato com um utilizador), das tarefas de gestão de rede.

7.2 Análise de propostas de arquitecturas alternativas

Este trabalho surge num período particularmente activo nesta área. Os trabalhos de normalização nas comunicações celulares, o UMTS [UMTSF], têm sido muito activos, com poucos anos a separar as primeiras versões estáveis (UMTS *release* 99), muito focadas na transição dos sistemas GPRS, das versões criando IP *Multimedia Subsystem* (IMS), entidades optimizadas para tratar do controlo de aplicações multimédia baseadas em IP (UMTS *release* 4), e das versões que assumem todo o tráfego como sendo tráfego IP (*releases* 5 e 6). Tal como neste trabalho, as classes de tráfego são suportadas usando uma abordagem de serviços diferenciados [Carl98].

As propostas apresentadas neste âmbito levam a sistemas complexos, fundamentalmente por causa da necessidade de manutenção de uma evolução suave a partir dos sistemas existentes. Outras propostas para redes de 4G têm aparecido na literatura, por exemplo [Joac04][KIM03][Wise02]. Estas propostas têm geralmente em comum o facto de assumirem a existência de redes baseadas em sistemas IP, mas são muito diversas, com diferentes níveis de detalhe, desde (as mais frequentes) propostas que enunciam os princípios subjacentes a estas futuras redes, a propostas que focam numa maior exploração das potencialidades do protocolo IP (mas mantendo a interoperabilidade com as redes actuais), a propostas de mapeamento de QoS, até a redes que expandem os conceitos de rede local para a rede de acesso pública. De notar no entanto que nenhuma destas propostas juntava a noção de serviço contratado, própria das redes actuais, com a flexibilidade de uma rede completamente baseada no protocolo IP, e deste modo permitindo manter a diversidade actual de aplicações dentro de um enquadramento realista de oferta de serviços de comunicação por parte de um operador.

7.3 Exploração dos resultados

O enquadramento do trabalho realizado nesta tese teve, sempre que possível, uma orientação para as necessidades previsíveis e reais dos operadores. O facto de parte significativa do trabalho ter sido realizado num ambiente industrial e de consultoria dentro do Grupo Portugal Telecom, permitiu avaliar algumas das necessidades e também o modo de funcionamento e operação real dos operadores. Este dados foram cruciais para moldar a arquitectura proposta em função de modelos e necessidades reais. É assim de um modo natural e com alguma satisfação que se pode observar a aplicação de alguns dos conceitos desenvolvidos nesta arquitectura a novos produtos.

Alguns dos produtos onde estes conceitos e módulos desta arquitectura estão ou vão ser aplicados são (de uma forma necessariamente breve, dados os interesses comerciais existentes):

- Controlador de acessos e portal de autenticação – este é um sistema que visa a aplicação específica de *hotspots*, mas que pode facilmente ser expandido a outros cenários mais complexos, incorporando a junção de QoS com AAAC.

- Gestor de Qualidade de Serviço – o QoS *Broker* foi definido de modo a poder facilmente ser utilizado em cenários reais no presente. Existem à data da escrita desta tese, planos para a implementação de um QoS *Broker* comercial.

A par do direccionamento deste trabalho para o mercado actual das telecomunicações, há a realçar também a sua continuidade no âmbito de um contínuo aperfeiçoamento e enriquecimento no âmbito de projectos de I&D a nível internacional, em particular no projecto Daidalos - *Designing Advanced network Interfaces for the Delivery and Administration of Location independent, Optimised personal Services* [Daid] (projecto integrado IST-2002-506997 do 6º programa quadro).

7.4 Sugestões para trabalho futuro

Esta arquitectura foi uma primeira abordagem a um problema complexo. A arquitectura, tal como foi definida, não aparenta problemas de viabilidade de implementação ou de escalabilidade, mas podem ser facilmente identificadas lacunas que deveriam ser estudadas e ultrapassadas, antes de se poder atingir uma implementação em larga escala e, de alguma forma, "universal". Alguns desses pontos são identificados nos parágrafos seguintes.

Se do ponto de vista conceptual se pode considerar que a arquitectura aqui apresentada suporta a existência de qualquer tipo de tecnologia de acesso, há no entanto que fazer essa integração. Ou seja, por cada tecnologia de nível 2 que se pretenda colocar ao serviço desta arquitectura, ter-se-á de fazer o mapeamento dos parâmetros IP, em recursos dessa tecnologia. A questão do transporte da sinalização também é fortemente dependente da tecnologia de acesso. Uma tecnologia que seja orientada ao "circuito" ou "canal", deverá possuir mecanismos que permitam que a sinalização IP se faça de um modo transparente (foi o caso do TD-CDMA na arquitectura aqui apresentada, que teve de ser readaptada em termos de sinalização de controlo de QoS).

Ao nível dos elementos de rede, a arquitectura apresentada tem algumas lacunas importantes, nomeadamente a falta de detalhe do NMS (*Network Management System*), que apesar de estar previsto, não tem as suas interacções com os restantes elementos da arquitectura suficientemente detalhadas. Este sistema, para além da gestão global da rede e recursos, deverá ser o agregador de dados colectados por diversas entidades de auditoria. Desta forma, será possível detectar anomalias e desencadear acções correctivas. Ao nível

dos SLSs e SLAs dos clientes, podem ser gerados relatórios e disponibilizadas interfaces para que os clientes possam verificar o cumprimento do seu contrato, e em caso de falha, podem também ser detectadas violações dos SLSs e SLAs dos clientes, ou gerar alarmes em caso de violação. O NMS deverá não só operar ao nível da rede, mas também ao nível dos serviços. Por exemplo, se um serviço de "vídeo a pedido" tem uma falha poucos segundos em uma hora, poderá ser o suficiente para que se possa considerar que o serviço não foi prestado de acordo com o SLA. Se a análise for feita apenas ao nível da rede, é provável que em termos de "bits" transportados, a falha não seja suficiente para considerar que houve uma violação do contrato. Torna-se então necessário que o NMS possua mecanismos fortes de detecção de violação de serviço, não apenas dependentes do nível da rede.

De modo a que a gestão dos recursos se possa fazer de um modo mais optimizado, o QoS *Broker* apresentado nesta tese prevê interfaces quer para o NMS, quer para sistemas de monitorização de rede. De modo a ter controlo, por exemplo, de atrasos e variações de atraso dentro de um domínio, ou mesmo extremo-a-extremo, pode (e deve) enveredar-se por um esquema de monitoria da rede em tempo real, com colectores activos [Além01][Sarg02] e passivos, distribuídos por pontos estratégicos da rede [VMar03c]. Um sistema de monitoria activa enviará pacotes nas diversas classes de serviço, espaçados e com dimensões segundo distribuições estatística que os aproxime dos serviços reais. Estes pacotes são enviados entre diversas "pontas de prova", localizadas com predominância nas fronteiras dos domínios (*routers* de acesso e *routers* de fronteira com outros domínios), e serão registados os tempos de partida e chegada de cada um deles. Desta forma é possível detectar perdas, atrasos e variações de atraso, e com estes dados alimentar o sistema de gestão (NMS) e os QoS *Brokers*. Um sistema de monitoria passiva é também importante numa arquitectura deste género. Embora a monitoria activa possa servir para determinar o estado da rede, a monitoria passiva, "observando" o tráfego real, pode detectar falhas reais e também, pelo lado positivo, verificar que a rede pode ainda suportar mais clientes e serviços activos, permitindo um determinado grau de multiplexagem estatística.

O QoS *Broker* apresentado nesta tese não está ainda dotado de mecanismos que lhe permitam adaptar-se a alterações nas tabelas de encaminhamento, especialmente dos *routers* de acesso. Neste momento, o QoS *Broker* prevê que cada *router* de acesso tenha apenas uma interface de ligação ao core e todos os seus algoritmos de controlo e gestão de

recursos se baseiam nesse facto. Desta forma, será importante no futuro que o QoS *Broker* possa participar e "entender" vários protocolos de *routing*, para que seja possível fazer, por exemplo, optimização dos percursos, balanceamento de carga, detecção e reacção a falha de ligações, integração com protocolos como MPLS, etc.

Em termos de mobilidade, a arquitectura aqui apresentada resolveu problemas de mobilidade de utilizadores e portabilidade de terminais. Isto é, um utilizador poderá usufruir dos seus serviços em terminais distintos, não havendo nenhum laço entre um terminal específico e o utilizador. Um ponto de investigação futura, será permitir a mobilidade e portabilidade do serviço, isto é, permitir que o utilizador possa transferir um serviço em tempo real de um terminal para outro. Poder-se-á ainda permitir que um utilizador possa estar a usufruir de serviços distintos, com qualidade de serviço distinta, em terminais distintos, e que cada serviço possa estar a recolher dados de origens distintas. Modelos como o DMIF poderão vir a desempenhar um papel fundamental neste tipo de arquitecturas, uma vez que prevêem a possibilidade de uma sessão ou serviço ser composta por conteúdos localizados em diversos pontos da rede distintos, e serem acessíveis através de várias tecnologias de acesso em simultâneo. A sincronização de conteúdos que modelos como o MPEG-4, MPEG-7 e MPEG-21 fornecem, são essenciais nestes cenários.

No que respeita à mobilidade, a sinalização apresentada focou-se no caso de mobilidade transparente dentro de um mesmo domínio administrativo. Existem diversas alternativas para a sua extensão de modo a permitir a mobilidade entre domínios administrativos distintos. Importa fazer um estudo idêntico ao apresentado nesta tese, dos pontos de vista analítico, de simulação e de experimentação. Neste caso, o principal constrangimento poderá ser a distância eléctrica entre as entidades dos domínios envolvidos. A solução poderá passar por acordos especiais entre os diversos operadores envolvidos no processo.

LISTA DE ABREVIATURAS E ACRÓNIMOS

1G	Primeira Geração
2G	Segunda Geração
2,5G	Entre a Segunda e a Terceira Gerações
3G	Terceira Geração
4G	Quarta Geração
AAA	*Authentication, Authorization and Accounting*
AAAC	*Authentication, Authorization, Accounting and Charging*
AAL	*ATM Adaptation Layer*
ABR	*Available Bit Rate*
ACELP	*Algebraic Code-Exited Linear Prediction*
ADSL	*Asymmetrical Digital Subscriber Line*
AF	*Assured Forwarding*
ALG	*Application Layer Gateway*
API	*Application Program(ming) Interface*
ARP	*Address Resolution Protocol*
AS	*Autonomous System ou Sistem Autónomo*
ASM	*Application Specific Module*
ATM	*Asynchronous Transfer Mode*
AVP	*Attribute Value Pair*

BA	*Behavior Aggregate*
BE	*Best Effort*
BL	Banda Larga
BB	*Bandwidth Brokers*
BGP	*Border Gateway Protocol*
BR	*Border Router*
CAC	*Call Admission Control* – Controlo de Admissão de Chamadas
CAS	*Channel Associated Signalling*
CATV	*Community Antenna TeleVision*
CBR	*Constant Bit Rate*
CCITT	*Comité Consultatif International Téléphonique et Télégraphique*
CCS	*Common Channel Signalling*
CDMA	*Code Division Multiple Access*
CDV	*Cell Delay Variation*
CELP	*Code Excited Linear Prediction*
CER	*Cell Error Ratio*
CHAP	PPP *Challenge Handshake Authentication Protocol*
CIDR	*Classless Inter-Domain Routing*
CIR	*Committed Information Rate*
CLI	*Command Line Interface*
CLNP	*ConnectionLess Network Protocol*
CMR	*Cell Misinsertion Rate*
CN	*Correspondent Node*
CoA	*Care-of Address*
COPS	*Common Open Policy Service*
CoS	*Classes of Service* – Classes de Serviço
CS-ACELP	*Conjugate Structure Algebraic-Code-Excited Linear-Prediction*
CTD	*Cell Transfer Delay*
DAI	*DMIF Application Interface*
DARPA	*Defense Advanced Research Projects Agency*
DHCP	*Dynamic Host Configuration Protocol*
DiffServ	*Differentiated Services* – Serviços Diferenciados

DMIF	*Delivery Multimedia Integrated Framework*
DNI	*DMIF Network Interface*
DNS	*Domain Name System ou Domain Name Service*
DS	*Differentiated Services field*
DSCP	*Differentiated Services Codepoint* ou *DiffServ CodePoint*
DTR	*Delay*, *Throughput* e *Reliability*
E-LSP	*Exp-inferred-LSP*
EAP	PPP *Extensible Authentication Protocol*
EF	*Expedited Forwarding*
ESI	*End Station Identifier*
ETSI	*European Telecommunications Standards Institute*
FA	*Foreign Agent*
FCAPS	*Fault, Configuration, Auditing, Performance and Security*
FDDI	*Fiber Distributed Data Interface*
FHm	*Fast Handover module*
FH	*Fast Handover*
FHO	*Fast HandOver*
FN	*Foreign Network*
FTP	*File Transfer Protocol*
GPRS	*Generic Packet Radio Service*
GSM	*Global System for Mobile communications*
Ha	*Home address*
HA	*Home Agent*
HDTV	*High Definition TeleVision*
HN	*Home Network*
HTTP	*Hyper Text Transfer Protocol*
ICMP	*Internet Control Message Protocol*
IETF	*Internet Engineering Task Force*
IMS	*IP Multimedia Subsystem*
IntServ	*Integrated Services* – Serviços Integrados
IP	*Internet Protocol*
IPSec	*IP SECurity*

IPv4	*Internet Protocol version 4*
IPv6	*Internet Protocol version 6*
ISP	*Internet Service Provider*
IST	*Information Society Technology*
ITU	*International Telecommunication Union*
ITU-T	ITU – *Telecommunication Standardization Sector*
L-LSP	*Label-only-inferred-LSP*
LAN	*Local Area Network*
LANE	*LAN Emulation*
LDAP	*Lightweight Directory Access Protocol*
LDP	*Label Distribution Protocol*
LSP	*Label Switching Path*
LSR	*Label Switching Router*
MAC	*Medium Access Control*
MGEN	*Multi-Generator Toolset*
MIB	*Management Information Base*
MIP	*Mobile IP*
MIPv4	*Mobile IPv4*
MIPv6	*Mobile IPv6*
MN	*Mobile Node*
MP-MLQ	*MultiPulse Maximum Likelihood Quantization*
MPEG	*Moving Picture Experts Group*
MPLS	*Multi-Protocol Label Switching*
MT	*Mobile Terminal*
MTNM	*Mobile Terminal Network Manager*
MTU	*Maximum Transmission Unit*
NAS	*Network Access Server*
NASREQ	*Network Access Server Requirements*
NAT	*Network Address Translation*
NCP	*Networking Control Panel*
NMS	*Network Management System*
NNTP	*Network News Transfer Protocol*

NSIS	*Next Steps in Signalling*
NTP	*Network Time Protocol*
NTSC	*National Television System Committee*
nrt-VBR	*non real time VBR*
ns	*network simulator*
NVUP	*Network View of the User Profile*
OSI	*Open Systems Interconnection*
OSPF	*Open Shortest Path First*
PAP	*PPP Password Authentication Protocol*
PAL	*Phase Alternating Line*
PBM	*Policy Based Management*
PCM	*Pulse Code Modulation*
PDA	*Portable Digital Assistant*
PDCP	*Packet Data Convergence Protocol*
PDH	*Plesiochronous Digital Hierarchy*
PDP	*Policy Decision Point*
PDU	*Packet Data Unit*
PEP	*Policy Enforcement Point*
PHB	*Per Hop Behavior*
PHY	*Physical Layer*
PIN	*Personal Identifier Number*
PPP	*Point-to-Point Protocol*
PS	*Policy Server*
PSTN	*Public System Telephone Network*
PT	Portugal Telecom
QoS	*Quality of Service* – Qualidade de Serviço
RA	*Router* de Acesso
RADIUS	*Remote Access Dial In User Service*
RAR	*Resource Allocation Request*
RCF	*Radio Convergence Function*
RDIS	Rede Digital com Integração de Serviços
RDIS-BL	Rede Digital com Integração de Serviços de Banda Larga

RED	*Random Early Detection*
RIO	*RED – with In and Out*
RLC	*Radio Link Control*
ROAMOPS	*Roaming Operations*
ROCS	*Regional Bell Operating Companies*
RRC	*Radio Resource Control*
RRM	*Radio Resource Management*
RSVP	*resource ReSerVation Protocol*
RTCP	*Real-Time Control Protocol*
RTFM	*Real-Time Flow Measurement*
RTP	*Real-time Transport Protocol*
rt-VBR	*real time VBR*
SAP	*Service Access Point*
SCTP	*Stream Control Transmission Protocol*
SECBR	*Severely Errored Cell Block Ratio*
SDH	*Synchronous Digital Hierarchy*
SIP	*Session Initiation Protocol*
SLA	*Service Level Agreement*
SLIP	*Serial Line Internet Protocol*
SLS	*Service Level Specification*
SMTP	*Simple Mail Transfer Protocol*
SNMP	*Simple Network Management Protocol*
SP	*Service Provider*
SONET	*Synchronous Optical NETwork*
SS7	*Signalling System number 7*
STP	*Shielded Twisted Pair*
SVC	*Switched Virtual Connection*
TC	*Traffic Control*
TCA	*Traffic Conditioning Agreement*
TCP	*Transmission Control Protocol*
TCS	*Traffic Conditioning Specification*
TD-CDMA	*Time Division CDMA*

TM	Terminal Móvel
ToS	*Type of Service*
TTI	*Transmission Time Interval*
UAC	*User Agent Client*
UAS	*User Agent Server*
UBR	*Unspecified Bit Rate*
UDP	*User Datagram Protocol*
UMTS	*Universal Mobile Telecommunications System*
UPC	*Usage Parameter Control*
URL	*Uniform Resource Locator*
URP	*User Registration Protocol*
UTP	*Unshielded Twisted Pair*
VBR	*Variable Bit Rate*
VC	*Virtual Channel*
VCC	*Virtual Channel Connection*
VoIP	*Voice over IP*
VP	*Virtual Path*
VPC	*Virtual Path Connection*
VPN	*Virtual Private Network*
WAN	*Wide Area Network*
WAP	*Wireless Access Protocol*
WDM	*Wavelenght Division Multiplex*
WEB	*World Wide Web*
WFQ	*Weighted Fair Queueing*
Wi-Fi	*Wireless Fidelity*
WLAN	*Wireless LAN*
WWW	*World Wide Web*
xDSL	*any Digital Subscriber Line technology*

REFERÊNCIAS

[3GPP] Endereço *Internet*: http://www.3gpp.org/

[ATMF] Endereço *Internet*: http://www.atmforum.com/

[Além01] R. Além e M. Carmo, "Mecanismos de *Probing* para Controlo de Admissão de Chamadas em Redes IP com QoS", Relatório de projecto em Eng. Electrónica e de Telecomunicações sob a orientação do Prof. Associado Rui Valadas com a colaboração da Eng. Susana Sargento e Eng. Victor Marques, Universidade de Aveiro, Setembro 2001.

[Almq92] P. Almquist "*Type of Service in the Internet Protocol Suite*", IETF RFC 1349, Julho 1992

[Arms] "*Armstrong IPv6 deployement*", Endereço *Internet*: http://www.eurescom.de/~public-webspace/P1000-series/P1009/index.html, acedido em 31 de Dezembro de 2003

[Beau03] Christophe Beaujean, Nesrine Chaher, Victor Marques, Rui L. Aguiar, Carlos García, José Ignacio Moreno, Michelle Wetterwald, Thomas Ziegler, "*Implementation and Evaluation of an End-to-End IP QoS Architecture for Networks Beyond 3rd Generation*", IST Mobile Summit 2003, Aveiro, Portugal, Junho 2003

[Bill] Endereço *Internet*: http://www.cclab.com/billhist.htm, acedido em 31 de Dezembro de 2003

[Borr02] Filipa Borrego, "Organização e estrutura das redes de comunicação: uma panorâmica actual", Dissertação de mestrado, Universidade de Aveiro, 2002

[Boyl00] J. Boyle et al., "*The COPS (Common Open Policy Service) Protocol*", IETF RFC 2748, 2000.

[BR5U04] 3G Americas, "*The Evolution of UMTS - 3GPP Release 5 and Beyond*", Junho de 2004

[Brad94] R. Braden, D. Clark e S. Shenker, "*Integrated Services in the Internet Architecture: An Overview*", RFC 1633, *IETF*, Junho 1994.

[Brad97] R. Braden, L. Zhang, S. Berson, S. Herzog e S. Jamin, "*Resource Reservation Protocol (RSVP) – Version 1 Functional Specification*", RFC 2205, *IETF*, Setembro 1997.

[Calh01] P. Calhoun et al., "*DIAMETER Base Protocol*". IETF RFC 3588, Setembro 2003.

[Calh02] Calhoun, P. et al, "*Diameter Framework Document*", draft-ietf-aaa-diameter-framework-01.txt, IETF Draft, Março 2001.

[Carl98] M. Carlson, W. Weiss, S. Blake, Z. Wang, D. Black e E. Davies, "*An Architecture for Differentiated Services*", RFC 2475, *IETF*, Dezembro 1998.

[CERN] Endereço *Internet*: http://www.cern.ch, acedido em 31 de Dezembro de 2003

[Cuev03] Antonio Cuevas, José Ignacio Moreno, Rui Aguiar, Victor Marques, Carlos García, Ignacio Soto, "*Mechanisms for AAA and QoS Interaction*", ASWN 2003, Zurique, Suiça, Maio 2003

[Daid] "*Designing Advanced network Interfaces for the Delivery and Administration of Location independent, Optimised personal Services* ", IST-2002-506997, Endereço *Internet*: http://www.ist-daidalos.org

[Deer98] S. Deering, R. Hinden, "*Internet Protocol, Version 6 (IPv6) Specification*", RFC 2460, IETF, Dezembro 1998

[Drom97] R. Droms, "*Dynamic Host Configuration Protocol*", RFC 2131, IETF, Março 1997

[Ethe] Endereço *Internet*: http://www.ethereal.com/

[ETSI] Endereço *Internet*: http://www.etsi.org/

[Fram98] "*Service Level Definitions Implementation Agreement FRF.13*", Frame Relay Forum Technical Committee, Agosto 1998

[Full93] V. Fuller, T. Li, J. Yu, K. Varadhan, "*Classless Inter-Domain Routing (CIDR): an Address Assignment and Aggregation Strategy*", RFC 1519, IETF, Setembro 1993

[GPRS00] *"3GPP TS 23.060 v3.3.1 General Packet Radio Service (GPRS); Service description; Stage 2" (Release 1999),* Maio 2000.

[Gozd03] Janusz Gozdecki, Piotr Pacyna, Victor Marques, Rui L. Aguiar, Carlos Garcia, Jose Ignacio Moreno, Christophe Beaujean, Eric Melin, Marco Liebsch, "*An IP QoS architecture for 4G networks*", "Art-QoS 2003", Warsaw, Poland, Março de 2003

[Hart02] H. Hartenstein, M. Liebsch, X. P. Costa, R. Schmitz, "*A MIPv6, FMIPv6 and HMIPv6 handover latency study: analytical approach*", IST Mobile & Wireless Telecommunications Summit 2002, Thessaloniki, Grécia, Junho 2002, pp. 100-105.

[Hein99] J. Heinanen, F. Baker, W. Weiss e J. Wroclawski, "*Assured Forwarding PHB Group*", RFC 2597, *IETF*, Junho 1999.

[Hist] Endereço *Internet*: http://www.historyalive.com/essays/ha20cah/topic14.asp, acedido em 31 de Dezembro de 2003

[IDSO] Endereço *Internet*: http://www.idsoftware.com/

[IETF] Endereço *Internet* do organismo IETF, http://www.ietf.org

[INTQ] Página do *Internet2 QoS Working Group*, http://www.*internet*2.edu/QoS/wg, acedido em 31 de Dezembro de 2003

[IP81] "*Internet Protocol, Darpa Internet Program, Protocol Specification*", IETF RFC 791, Setembro 1981.

[IPSE] Endereço *Internet*: http://www.ietf.org/html.charters/ipsec-charter.html, acedido em 31 de Dezembro de 2003

[IPv6] Endereço *Internet*: http://www.ipv6.org, acedido em 31 de Dezembro de 2003

[IPv6a] Endereço *Internet*: http://www.ietf.org/html.charters/ipv6-charter.html, acedido em 31 de Dezembro de 2003

[ITUT] Endereço *Internet*: http://www.itu.org

[Jaco99] V. Jacobson, K. Nichols, K. Poduri, "*An Expedited Forwarding PHB*", RFC 2598, *IETF*, Junho 1999.

[Joac04] Joachim Hillebrand, et al., "*Quality-of-Service Signaling for Next-Generation IP-Based Mobile Networks*", IEEE Communications Magazine, Junho 2004, páginas 72 a 79.

[Jonh03] D. Jonhson, Perkins C., J. Arko, "*Mobility Support in IPv6*", IETF *Internet* Draft: draft-ietf-mobileip-ipv6-24,30, Junho 2003

[Jupi] "*Jupiter-II - Joint Usability, Performability and Interoperability Trials in Europe*", Endereço *Internet*: http://www.eurescom.de/~public-webspace/P800-series/P807/main/main.htm, acedido em 31 de Dezembro de 2003

[Kaar01] H. Kaaranen et al.; "*UMTS Networks: Architecture, Mobility, and Services*". John Wiley & Sons, 2001

[KIM03] G. Kim et al, "*Beyond 3G: vision, requirements, and enabling technologies*", IEEE Communications Magazine, Volume 41, Março 2003, páginas 120 a 124.

[Kood02] Rajeev Koodli, "*Fast Handovers in Mobile IPv6*", IETF *Internet* Draft, <draft-ietf-mobileip-fast-mipv6-05.txt>, Setembro 2002

[Kram01] M. Krampell et al., "*Interaction of transition mechanisms*", IETF Draft, draft-krampell-v6transition-interaction-00.txt, Fevereiro 2001

[Laat00] C. de Laat, G. Gross, L. Gommans, J. Vollbrecht, D. Spence: "*Generic AAA Architecture*"; IETF Experimental RFC 2903, Agosto 2000

[Laub94] M. Laubach, "*Classical IP and ARP over ATM*", RFC 1577, IETF, Janeiro 1994

[Lina03] Lina Brito, "Qualidade de Serviço para Comunicações IP em Redes Móveis", Dissertação de mestrado, Universidade de Aveiro, 2003

[*Linux*] Endereço *Internet*: http://www.linux.org

[M.3400] Recomendação ITU-T, Série M, "*M.3400 - TMN Management Functions*", Fevereiro de 2000

[Marc02] Marco Liebsch et al, "*Solutions for IPv6-based mobility in the EU project Moby Dick*", World Teletraffic Congress, Paris, França, Setembro, 2002

[Marc04] Marco Liebsch, Bernd Lamparter, "*A generic IP Paging Architecture and Protocol*", European Wireless Conference 2004, Barcelona, Fevereiro, 2004

[MGEN] Endereço *Internet*: http://manimac.itd.nrl.navy.mil/MGEN/, acedido em 31 de Dezembro de 2003

[Micr] Endereço *Internet*: http://www.microsoft.com

[MPEG4a] Endereço *Internet*: http://www.chiariglione.org/mpeg/standards/mpeg-4/mpeg-4.htm, acedido em 31 de Dezembro de 2003

[MPEG4b] ISO/IEC JTC1/SC29/WG11 N4668, "*Overview of the MPEG-4 Standard*", Março 2002

[Moby] "*Mobility and Differentiated Services in a Future IP Network*", IST-2000-25394, Endereço *Internet*: http://www.ist-mobydick.org

[Neil99] R. Neilson, J. Wheeler, F. Reichmeyer e S. Hares, "*A Discussion of Bandwidth Broker Requirements for Internet2 Qbone Deployment*", *Internet* Qbone BB Advisory Council, Version 0.7, Agosto 1999.

[Nich98] K. Nichols, S. Blake, F. Baker, D. Black, "*Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers*", RFC 2474, *IETF*, Dezembro 1998.

[Nich99] K. Nichols, V. Jacobson e L. Zhang, "*A Two-Bit Differentiated Services Architecture for the Internet*", IETF RFC 2638, Julho 1999.

[NS-2] Endereço *Internet*: http://www.isi.edu/nsnam/ns/index.html, acedido em 31 de Dezembro de 2003

[NSIS] Endereço *Internet*: http://www.ietf.org/html.charters/nsis-charter.html, acedido em 31 de Dezembro de 2003

[Perk02] C. Perkins, Ed., "*IP Mobility Support for IPv4*", IETF RFC 3344, Agosto 2002

[QoSF] Endereço *Internet*, http://www.qosforum.com/

[QoSF98] QoSForum, "Introduction to QoS Policies", *White Paper*, 1998.

[Regi02] Regis J. "Bud" Bates, "*GPRS General Packet Radio Service*", McGraw-Hill, 2002.

[Rign00] C. Rigney et al., "*Remote Authentication Dial In User Service (RADIUS)*", IETF RFC 2865, 2000.

[Rose01] E. Rosen, A. Viswanathan, R. Callon, "*Multiprotocol Label Switching Architecture*", RFC 3031, Janeiro 2001

[Rosb02] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, E. Schooler "*SIP: Session Initiation Protocol*", IETF RFC 3261, Junho 2002

[Sarg02] Susana Sargento, Roger Salgado, Miguel Carmo, Victor Marques, Rui Valadas, Edward Knightly, "*An Experimental Study of Probing-Based Admission Control for DiffServ Architectures*"; Apresentação oral na Networking2002, Pisa, Itália, 19 a 24 de Maio de 2002

[Sarg03] Susana Sargento, "Gestão de Recursos em Redes com Suporte de Qualidade de Serviço", tese de doutoramento, Universidade de Aveiro, Fevereiro de 2003

[Schu34] J. Schumpeter, "*The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle*" (*translated by Redvers Opie, with a special preface by the author*), Cambridge Mass: Harvard University Press, 1934. Second Printing 1936; third printing 1949.

[Shen97] S. Shenker, C. Partridge e R. Guerin, Specification of Guaranteed Quality of Service", RFC 2212, *IETF*, Setembro 1997.

[Sris01] P. Srisuresh e K. Egevang, "*Traditional IP Network Address Translator (Traditional NAT)*", IETF RFC 3022, 2001.

[TCAPI] TC API Project (Fevereiro 2003), Endereço *Internet*: http://oss.software.ibm.com/developerworks/projects/tcapi, acedido em 31 de Dezembro de 2003

[Tiph00] ETSI TR 101 329-1 V3.1.1 - "*General Aspects of Quality of Service*", Tiphon (*Telecommunications and Internet Protocol Harmonization over Networks*), ETSI, Julho, 2000

[UMTS00] "*UMTS, 3GPP TS 23.002 v5.0.0: Network Architecture (Release 5)*", Outubro 2000.

[UMTSF] Endereço *Internet*: http://www.umts-forum.org

[UQoS00] "*UMTS, 3GPP TS 23.107 v3.0.0: UMTS QoS Concept and Architecture*."

[Vieira03] Jacinto Vieira, Victor Marques, Carlos Parada, Carlos Rodrigues, Francisco Fontes, "*A Glance of the Current IPv6 Status*", ConfTele 2003, Aveiro, Portugal, Junho 2003

[VMar01] Victor Marques, Ricardo Cadime, Amaro Sousa, A. Oliveira Duarte, "*DMIF based QoS Management for MPEG-4 Multimedia Streaming: ATM and RSVP/IP Case Studies*"; Apresentação oral na ConfTele 2001, Figueira da Foz, Portugal, Abril 2001

[VMar01a] Victor Marques, Rui Aguiar, Francisco Fontes, Jürgen Jähnert, Hans Einsiedler ; "*Enabling IP QoS in Mobile Environments*"; Apresentação oral na "Mobile Summit 2001", Barcelona, Espanha, Setembro 2001

[VMar01b] Victor Marques, Rui Aguiar, Jürgen Jähnert, Karl Jonas, Marco Liebsch, Hans Einsiedler, Francisco Fontes; "*An Heterogeneous Mobile IP QoS-aware Network*"; "CRC 2001", Novembro 2001

[VMar02] Victor Marques, Rui L. Aguiar, Piotr Pacyna, Janusz Gozdecki, Christophe Beaujean, Nesrine Chaher, Carlos García, José Ignacio Moreno, Hans Einsiedler, "*An Architecture Supporting End-to-End QoS with User Mobility for Systems Beyond 3rd Generation*"; "IST Mobile Summit 2002", Tessalónica, Grécia, Junho de 2002

[VMar03a] Victor Marques, Rui L. Aguiar, Antonio Cuevas Casado, Jose Ignacio Moreno, Nesrine Chaher, "*A simple QoS service provision framework for beyond 3rd generation scenarios*", "10th International Conference on Telecommunications - ICT'2003", Tahiti, Papeete, French Polynesia, Março 2003.

[VMar03b] Victor Marques et al., "*An IP-based QoS architecture for 4G operator scenarios*", IEEE Wireless Communications Magazine, Junho 2003.

[VMar03c] Victor Marques, Carlos Parada, Pedro Gonçalves, Rui L. Aguiar, Francisco Fontes, "*Next Generation Network Provider Architecture Demonstrator*", ConfTele 2003, Aveiro, Portugal, Junho 2003

[VoIP04] Endereço *Internet*: http://www.iec.org/online/tutorials/vfoip/topic01.html, acedido em 31 de Dezembro de 2003

[Wise02] D. Wisely, E. Mitjana, "*Paving the Road to Systems Beyond 3G - The IST MIND Project*", Journal of Communication and Networks, Dezembro 2002.

[Wroc97] J. Wroclawski, "*Specification of the Controlled-Load Network Element Service*", RFC 2211, *IETF*, Setembro 1997.

[WWW] Endereço *Internet*, http://www.w3.org/WWW, acedido em 31 de Dezembro de 2003

[Yava00] R. Yavatkar, "*A Framework for Policy-Admission Control*", IETF RFC 2753, 2000.

[Zhan93] L. Zhang, S. Deering, D. Estrin, S. Shenker e D. Zappala, "*RSVP: A New Resource ReSerVation Protocol*", *IEEE Network*, pp. 8-18, Setembro 1993