

LETICIA AZEVEDO GENOVEZ DE MESQUITA BRAGA

**ESTUDO DE CONVERGÊNCIA DE
REDES,
*NEXT GENERATION NETWORK E
IP MULTIMEDIA SUBSYSTEM***

Trabalho de Conclusão de Curso
apresentado à Escola de Engenharia de São
Carlos, da Universidade de São Paulo

Curso de Engenharia Elétrica com ênfase
em Eletrônica

ORIENTADORA: Mônica Lacerda Rocha

São Carlos
2011

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica preparada pela Seção de Tratamento
da Informação do Serviço de Biblioteca – EESC/USP

B813e Braga, Leticia Azevedo Genovez de Mesquita
Estudo de convergência de redes, *Next Generation Networks e IP Multimedia Subsystem* ; orientador Mônica de Lacerda Rocha -- São Carlos, 2011.

Monografia (Graduação em Engenharia Elétrica com ênfase em Eletrônica) -- Escola de Engenharia de São Carlos da Universidade de São Paulo, 2011.

1. Redes convergentes. 2. *Next Generation Networks*.
3. *IP Multimedia Subsystem*. 4. Redes de telecomunicação.
I. Título.

Agradecimentos

Agradeço primeiramente a Deus pela vida que me deu, que sem ela nenhuma das realizações feitas até hoje e no futuro a seriam possíveis. Agradeço também a minha família: meus pais Carlos e Nilde, meus irmãos Cecília, Marília, Livia, João Paulo, Álvaro José e José Augusto, avós, tios e primos por fazerem parte da minha vida e por sempre me apoiarem em tudo. À minha orientadora, Profª Mônica, por aceitar desenvolver esse trabalho comigo e me incentivar sempre superar os meus limites. A Trópico pela oportunidade de trabalhar e conhecer melhor o ambiente das telecomunicações não menos importante a todos os meus amigos que me ajudaram e deram forças nessa graduação, em especial a Maria Luiza, a Natália Passone, Larissa, Rafael, Ciro, Nishizawa e Guilherme Mazanti, entre muitos outros que podem ser citados aqui.

Obrigada a todos vocês e a muitos outros que tornaram esse trabalho e essa graduação possíveis.

Resumo

A convergência das redes de telecomunicações já é uma realidade no cenário mundial. As operadoras, em busca da conquista de mercado, já adotaram o modelo NGN (*Next Generation Networks*), que se baseia na rede IP em operações. A grande vantagem dessa solução é a facilidade de prover serviços diferenciados aos usuários.

A fim de melhorar a forma como os serviços são oferecidos, órgãos de padronização internacionais se juntam constantemente para decidir novos conceitos e tecnologias que tragam facilidades tanto para o usuário, quanto para as operadoras e fornecedores de equipamento. Um dos modelos mais promissores é o IMS (*IP Multimedia Subsystem*), que entre suas principais premissas estão tarifação diferenciada, integração de serviços e provisão de qualidade de serviço.

Neste trabalho é apresentado um estudo sobre um ambiente de convergência das redes de telecomunicações, motivações e razões que apontam para o atual modelo adotado pelas operadoras. Também apresenta o conceito de IMS, uma arquitetura promissora que vem sendo apontada por especialistas como uma melhoria significativa para serviços oferecidos pela rede IP com suas principais características e funcionamento.

Palavras-chave: Redes Convergentes, NGN, Redes de Telecomunicação, IMS.

Abstract

The convergence of telecommunications networks is a reality in the world stage. The service providers, in search of market conquest, already adopted the model NGN (Next Generation Network), which is based on the IP network operations. The great advantage of this solution is the ease of providing differentiated services to users.

In order to improve the way services are offered, international standardization groups are added constantly to define new concepts and technologies that bring same facilities to the user, to the service providers and equipment suppliers.

One of the most promising is the IMS (IP Multimedia Subsystem), which among its main premises are differentiated charging, integrated services and provision of Quality of Service.

This research presents a study on the environment of convergence of telecommunication networks, motivations and reasons that point to the current model adopted by the service providers. It also introduces the concept of IMS, a promising architecture that has been identified by experts as a significant improvement to services offered by the IP network with its main characteristics and performance.

Key-words: Converged Networks, NGN, Telecommunication networks, IMS.

Sumário

Agradecimentos	iii
Resumo	i
Abstract	iii
Sumário	v
Lista de Figuras	viii
Lista de Tabelas	x
Lista de Siglas	xi
1. Introdução.....	1
1.1. Objetivo do Trabalho	1
1.2. Organização do Texto.....	1
2. Evolução das Telecomunicações	3
2.1. O Homem e a Comunicação	3
2.2. Evolução da Comunicação de Voz através da Rede de Telefonia.....	4
2.2.1. Comutação das Chamadas	7
2.2.2. Protocolos de Sinalização entre Centrais Telefônicas	8
2.3. Evolução na Comunicação de Dados.....	9
2.3.1. Modelo OSI	11
2.3.2. Modelo TCP/IP	13
2.4. Evolução na Comunicação Móvel	14
2.4.1. Primeira Geração	15
2.4.2. Segunda Geração.....	16
2.4.3. A Transmissão de dados via Sistemas Celulares	16
2.4.4. A Tecnologia de 3ª Geração de Rede Celular	17
3. Convergência de redes	18
3.1. Ambiente Pré-Convergência	18
3.1.1. Limitações de Infraestrutura	19
3.1.2. Outras Dificuldades.....	19
3.2. Justificativas para a Convergência de Redes e Serviços	20
3.3. Premissas para a Convergência de Redes.....	20
3.4. Início da convergência	21
3.4.1. Voz sobre Frame Relay	22
3.4.2. ATM	24
3.4.3. VoIP - Voz sobre IP.....	26
3.5. Necessidades de uma Rede Convergente	28
3.5.1. Codificação do Sinal de Voz- Processo de Digitalização da Voz.....	29
3.5.2. Largura de Banda e Compressão.....	30
3.5.3. Atrasos, Congestionamento e Perda de Pacotes	30
3.5.4. Qualidade de serviço	31
3.6. QoS na rede IP	32
3.6.1. IntServ - Arquitetura de Serviços Integrado	32
3.6.2. DiffServ - Arquitetura de Serviços Diferenciados.....	33
3.6.3. Protocolo MPLS (MultiProtocol Label Switching)	34
3.6.4. Alternativas para QoS na rede IP	35
3.7. Evolução das redes convergentes.....	35
4. NGN - Next Generation Networks.....	37
4.1. Características.....	37
4.2. Padronização	39
4.3. Arquitetura.....	39

4.3.1.	Camada de transporte	41
4.3.2.	Camada de Controle	41
4.3.3.	Camada de Aplicação e Serviços	41
4.4.	Equipamentos da rede NGN	42
4.4.1.	<i>Media Gateways</i>	42
4.4.2.	<i>Softswitch</i>	43
4.5.	Protocolos de controle da rede	44
4.5.1.	MGCP - <i>Media Gateway Control Protocol</i>	44
4.5.2.	Protocolo Megaco/ H.248 [6]	45
4.5.3.	H.323	45
4.5.4.	Protocolo SIP	47
4.5.4.1.	Comparações entre o SIP e HTTP	48
4.5.4.2.	Principais características do protocolo SIP	49
4.5.4.2.1.	SIP Server (Servidores SIP)	49
4.5.4.2.2.	Identificação SIP	49
4.5.4.2.3.	Relação entre SIP e outros protocolos	50
4.5.4.2.4.	Mensagens SIP	51
4.5.4.2.4.1.	Mensagens SIP <i>Request</i>	52
4.5.4.2.4.2.	Mensagens SIP <i>Response</i>	53
5.	IP MULTIMEDIA SUBSYSTEM – IMS	55
5.1.	Justificativas para o IMS	55
5.2.	Premissas IMS	57
5.3.	Exemplos de Aplicação	60
5.4.	Padronização IMS	62
5.4.1.	Órgãos responsáveis	62
5.4.2.	Evolução da Padronização do IMS	63
5.4.3.	Visão geral da padronização do IMS	64
5.5.	Estrutura / Arquitetura IMS	66
5.5.1.	Estrutura	67
5.5.2.	Interfaces	67
5.5.3.	O núcleo IMS	69
5.5.3.1.	Bases de Dados: HSS e SLF	69
5.5.3.2.	Controle das sessões – <i>Call/Session Control Function</i> – CSCF	70
5.5.3.3.	<i>Proxy-CSCF</i> (P-CSCF)	71
5.5.3.4.	<i>Interrogating</i> – CSCF (I-CSCF)	73
5.5.3.5.	<i>Serving</i> - CSCF (S-CSCF)	75
5.5.3.6.	Plano de Aplicação - <i>Application Servers</i> – AS	77
5.5.3.7.	<i>Media Resource Function</i> (MRF)	79
5.5.3.8.	Conexão do IMS com as redes comutadas PSTN	81
5.5.3.8.1.	<i>Breakout Gateway Control Function</i> (BGCF)	81
5.5.4.	Versões diferentes de IP	83
5.6.	Identidade do usuário no IMS	84
5.6.1.	A identidade pública	84
5.6.2.	Identidade Privada	84
5.7.	Cobranças no IMS (<i>Charging</i>)	85
5.7.1.	Cobrança Online	86
5.7.2.	Cobrança Offline	86
5.8.	Caso de uma sessão IMS	86
6.	Situação Atual e Futuro da NGN e do IMS	88
6.1.	Redes convergentes e NGN	88
6.2.	IMS	89
7.	Conclusão	92
	Referências Bibliográficas	94

Lista de Figuras

Figura 1 - Topologia " <i>Fully Meshed</i> " [4]	4
Figura - 2 – Mesa de Operação e Telefonista [3]	5
Figura 3 - Central telefônica com diversos operadores [5]	5
Figura 4 - Fios de telefonia "poluindo" a cidade [2]	6
Figura 5 - Hierarquia de centrais [4]	6
Figura 6- Evolução da Transmissão da Telefonia [6].....	8
Figura 7 - Interação de sinalizações de uma chamada telefônica em uma rede TDM	9
Figura 8 - Computador Harvard Mark 1 [7]	10
Figura 9 - Modelo OSI com os principais componentes de cada camada.	12
Figura 10 - Comparação Modelo OSI e Modelo TCP/IP	14
Figura 11 - Martin Cooper e o primeiro modelo de celular comercial [12]	15
Figura 12 - Redes própria para telefonia fixa, móvel e dados [17]	22
Figura 13 - Rede com Tecnologia Frame Relay [16]	24
Figura 14 - Célula ATM	25
Figura 15 - Convergência de voz em dados via IP	26
Figura 16 - Endereços IP [16]	27
Figura 17 - Processo PCM [6]	29
Figura 18 - Rede com MPLS [16]	34
Figura 19 - Camadas de uma rede NGN.....	40
Figura 21 - Topologia de uma rede NGN.....	40
Figura 21 - Representação de um bloco Media Gateway [6]	43
Figura 22 - Representação de um Softswitch [6]	44
Figura 23 - Componentes do Protocolo H.323	47
Figura 24 - Protocolos utilizados em conjunto com o SIP [22]	51
Figura 25 - Estrutura de mensagens SIP [22]	52
Figura 26 - Mensagem SIP Request [22].....	52
Figura 27 - Mensagem SIP Resposta [22]	53
Figura 28 - Redes conectadas com IMS	59
Figura 29 - Tecnologias convergentes no IMS.....	59
Figura 30 - Sistema NGN e IMS [27]	60
Figura 31 - Lista de contatos de um usuário com IMS [27]	61
Figura 32 - Exemplo Aplicação IMS [27]	62
Figura 33 - Release apresentados ao longo dos anos para o IMS.....	64
Figura 34 - Comparação - Modelo OSI e Protocolos IMS [27]	65
Figura 35 - Separação em camadas dos componentes em uma rede IMS [24]	66
Figura 36 - Modelo em Camadas do IMS	67
Figura 37 - Localização das Interfaces na rede IMS [30].....	68
Figura 38 - Pilha HSS [27]	70
Figura 39 - Acesso ao Servido P-CSCF	72
Figura 40 - Resposta ao Pedido de Acesso.....	73
Figura 41 - Informações trocadas pelo I-CSCF.....	74
Figura 42 - Controle feito pelos CSCFs	76
Figura 43 - Sessão Completa de Controle IMS [27]	76
Figura 44 - Localização do AS na rede IMS	78
Figura 45 - Pedido de serviço ao AS pelo S-CSCF [27]	79
Figura 46 - Componentes do MRF [27]	80

Figura 47 - - Localização dos MRF na rede IMS	81
Figura 48 - Ligação mundo PSTN e IMS	82
Figura 49 - Conversão do IMS com diferentes versões do IP	83
Figura 50 - Identidades Públicas e Privadas do Usuário [27].....	85
Figura 51 - Evolução dos SIM [27]	85
Figura 52 - Caso de uma sessão IMS [32].....	87
Figura 53 - Dificuldades do IMS [36]	91

Lista de Tabelas

Tabela 1 - Interfaces do IMS [27]	68
Tabela 2 - Assinantes 3G por total de celulares	90

Lista de Siglas

3G	<i>Third Generation</i>
3GPP	<i>Third Generation Partnership Project</i>
3GPP2	<i>Third Generation Partnership Project 2</i>
AAL	<i>ATM Adaptation Layer</i>
ADSL	<i>Asymmetric Digital Subscriber Line</i>
ANATEL	<i>Agência Nacional de Telecomunicações</i>
API	<i>Application Programming Interfaces</i>
AS	<i>Aplication Server</i>
ASR	<i>Automated speech recognition</i>
ATIS IIF	<i>Automatic Terminal Information Service – IPTV Interoperability Forum</i>
ATM	<i>Assynchronous Transfer Mode</i>
B2BUA	<i>Back-to-back User Agent</i>
BGCF	<i>Breakout Gateway Control Function</i>
BICC	<i>Bearer Independent Call Control</i>
CAMEL	<i>Customized Application for Mobile Network Enhanced Logic</i>
CBR	<i>Constant Bit Rate</i>
CDMA	<i>Code Division Multiple Access</i>
CDR	<i>Clock and Data Recovery</i>
CIR	<i>Committed Information Rate</i>
COPS	<i>Common Open Policy Service</i>
CRM	<i>Customer Relationship Management</i>
CSCF	<i>Call/Session Control Function</i>
DHCP	<i>Dynamic Host Configuration Protocol</i>
DLCI	<i>Data Link Connection Identifier</i>
DNS	<i>Domain Name System</i>
DSCP	<i>Differentiated Service Code Point</i>
DSL	<i>Digital Subscriber Line</i>
DVB	<i>Digital Video Broadcasting</i>
ETSI	<i>European Telecommunications Standards Institute</i>
FCM	<i>Fixed-Mobile Convergence</i>
FCMA	<i>Fixed-Mobile Convergence Alliance</i>
FGNGN	<i>Focus Group on NGN</i>

FQDN	<i>Fully Qualified Domain Name</i>
FRAD	<i>Frame Relay Access Device</i>
FTP	<i>File Transfer Protocol</i>
GGSN	<i>Gateway GPRS Support Node</i>
GPRS	<i>General Packet Radio System</i>
GSM	<i>Global System for Mobile Communications</i>
HDTV	<i>High Definition Television</i>
HGI	<i>Home Gateway Initiative</i>
HLR	<i>Home Location Register</i>
HSS	<i>Home Subscriber Server</i>
HTTP	<i>Hypertext Transfer Protocol</i>
I-CSCF	<i>Interrogating - Call/Session Control Function</i>
IEEE	<i>Institute of Electrical and Eletrcronics Engineering</i>
IETF	<i>Internet Enginnering Task Force</i>
IMS	<i>IP Multimidia Subsystem</i>
IM-SSF	<i>IMS Service Switching Function</i>
INAP	<i>Intelligent Network Application Part</i>
IP	<i>Internet Protocol</i>
IPTV	<i>IP Television</i>
IPX	<i>Internet Packet Exchange</i>
ISDN	<i>Integrated Services Digital Network</i>
ISO	<i>International Organization for Standardization</i>
ISUP	<i>ISDN User Part</i>
ITU-I	<i>International Telecommunication Union Telecommunication Standardization</i>
LAN	<i>Local-area Network</i>
LIA	<i>Location-Information-Answer</i>
LIR	<i>Location-Information-Request</i>
LTE	<i>Long Term Evolution</i>
MAN	<i>Metropolitan Area Network</i>
MCU	<i>Multipoint Control Unit</i>
MEGACO	<i>Media Gateway Control Protocol</i>
MG	<i>Media Gateway</i>
MGC	<i>Media Gateway Controler</i>
MGCF	<i>Media Gateway Control Function</i>
MGCP	<i>Media Gateway Control Protocol</i>
MMUSIC	<i>Multiparty Multimedia Session Control</i>

MPLS	<i>Multiprotocol Label Switching</i>
MRF	<i>Media Resource Function</i>
MRFC	<i>Media Resource Function Controller</i>
MRFP	<i>Media Resource Function Processor/Plataform</i>
NGN	<i>Next Generation Networks</i>
NGN-FG	<i>NGN Focus Group</i>
NGN-GSI	<i>NGN-Global Sdandards Initiative</i>
OSA	<i>Open Service Access</i>
OSA-SCS	<i>Open Service Access-Service Capability Server</i>
OSI	<i>Open System Interconnection</i>
PCM	<i>Modulação por Codificação de Pulsos</i>
P-CSCF	<i>Proxy-Call/Session Control Function</i>
PoC	<i>Push to Talk over Cellular</i>
PSTN	<i>Public Switch Telephony Network</i>
PVC	<i>Permanent Virtual Connection</i>
PVC	<i>Permanent Virtual Circuit</i>
QoS	<i>Quality of Service</i>
RI	<i>Rede Inteligente</i>
RIP	<i>Routing Information Protocol</i>
RSVP	<i>Resource Reservation Protocol</i>
RTCP	<i>Real Time Control Protocol</i>
RTP	<i>Real Time Protocol</i>
SAE	<i>System Architecture Evolution</i>
SCP	<i>Service Control Point</i>
S-CSCF	<i>Serving- Call/Session Control Function</i>
SDP	<i>Session Description Protocol)</i>
SGW	<i>Signaling Gateway</i>
SIP	<i>Session Initiation Protocol</i>
SLA	<i>Service Level Agremment</i>
SLF	<i>Subscription Locator Function</i>
SMTP	<i>Simple Mail Transfer Protocol</i>
SRVCC	<i>Single Radio Voice Call Continuity</i>
SS7	<i>Sistema de Sinalização 7</i>
SSF	<i>Service Switching Function</i>
SVC	<i>Switched Virtual Circuit</i>
TCP	<i>Transmission Control Protocol</i>

TDM	<i>Time Division Multiplexing</i>
TDM	<i>Time Division Multiplexing</i>
TDMA	<i>Time Division Multiple Access</i>
THIG	<i>Thopology Hiding Inter-Network Gateway</i>
TISPAN	<i>Telecoms & Internet Converged Services & Protocols for Advanced Networks</i>
TU	<i>Terminal do Usuário</i>
UA	<i>User Agent</i>
UDP	<i>User Datagram Protocol</i>
UHF	<i>Ultra High Frequency</i>
URI	<i>Uniform Resource Identifier</i>
VC	<i>Virtual Circuits</i>
VHF	<i>Very High Frequency</i>
VoATM	<i>Voz sobre ATM</i>
VoFR	<i>Voz sobre Frame Relay</i>
VoIP	<i>Voz sobre IP</i>
VPC	<i>Virtual Path Connection</i>
VPN	<i>Virtual Private Network</i>
WAN	<i>Wide-area Network</i>
WAP	<i>Wireless Application Protocol</i>
WiMAX	<i>Worldwide Interoperability for Microwave Access</i>

1. Introdução

As redes de telecomunicações estão sendo aperfeiçoadas para suportar a transmissão de informações com a introdução de novas tecnologias, tanto do lado dos equipamentos da rede (elementos de rede), quanto dos meios de transmissão (redes de transporte) e dos sistemas de operação para gerenciamento. Desta forma, é possível perceber que as diversas redes existentes, como as de voz, vídeo e dados, estão evoluindo para apenas uma rede comum, garantindo maior quantidade de serviços, oferecidos com qualidade sempre superior.

Essas novas tecnologias visam aperfeiçoar e aumentar o oferecimento de serviços e aplicações nessa nova rede comum, mais conhecida como rede convergente. A esse termo de convergência de redes se dá o nome de NGN (*Next Generation Networks* ou Redes de Próxima Geração). Com elas é possível a transmissão de informações e serviços (voz, dados e mídias como o vídeo) encapsulando-os em pacotes, semelhante ao tráfego de dados da Internet. Normalmente utiliza-se como base de transporte o protocolo IP.

No contexto de evolução das redes, as operadoras de telecomunicações buscam alternativas para garantir, com qualidade, o atendimento à demanda crescente de serviços, o que incentiva a pesquisa de novas tecnologias.

Entre muitas novas tendências, pode-se destacar o IMS (*IP Multiservice Subsystem*) como uma das mais promissoras. O IMS é definido como uma arquitetura de referência que possibilita a entrega de serviços de comunicação de voz, dados, e vídeo, com mobilidade, sobre uma rede IP. Sua utilização apresenta grandes melhorias em relação a outros tipos de rede, principalmente em relação às redes de telefonia.

1.1. Objetivo do Trabalho

O objetivo dessa monografia é apresentar primeiramente, o cenário das redes de telecomunicações a partir de uma breve introdução ao histórico da evolução das redes de telefonia fixa, telefonia móvel e de dados. Em seguida, discutir conceitos da evolução das redes, motivações e razões que levaram às novas tendências dessas redes, focando especialmente o NGN e o IMS.

1.2. Organização do Texto

O texto está organizado em capítulos que pretendem mostrar os uma breve introdução aos meios de comunicações mais utilizados até os dias atuais, com o conceito de NGN e IMS.

No Capítulo 2 é apresentado um breve histórico das telecomunicações, dando uma maior ênfase a evolução da telefonia fixa e suas características e padrões mais importantes; as redes de dados até os modelos de arquiteturas utilizados; e redes móveis, que foram a principal motivação para o IMS. O Capítulo 3 revela o ambiente de convergência de redes, revelando motivos, razões e premissas que levaram as operadoras a adotarem esse sistema. Também apresenta de uma forma geral as algumas das principais tecnologias e alternativas para correção de questões dessa convergência. O Capítulo 4 apresenta o principal modelo adotado pelas operadoras, conhecido como NGN, revelando suas principais características, seus componentes de rede e protocolos mais usados. Já o capítulo 5 aborda o objeto principal de estudo, a arquitetura de referência IMS, descrevendo suas principais características e o processo envolvido para a sua atuação. Por fim, o Capítulo 6 apresenta uma breve descrição da situação da NGN e do IMS, assim como qual deve ser o futuro para esses dois padrões.

2. Evolução das Telecomunicações

Neste capítulo busca-se comentar os alguns dos principais eventos históricos relacionados à evolução das telecomunicações, evidenciando os aspectos que colaboraram para a situação atual e o futuro das redes de comunicação, conhecidas também por NGN e IMS. O objetivo principal desse capítulo é apresentar uma breve introdução das invenções e tecnologias mais importantes que contribuíram para o momento presente das redes de telecomunicações.

2.1. O Homem e a Comunicação [1] [2] [3]

O ser humano como ser social necessita se comunicar com seus iguais. Isso está registrado desde as primeiras pinturas rupestres pelos primeiros hominídeos, que estimam ser de mais de 40.000 A.C. Essas pinturas, por si só, já podem ser entendidas como uma forma de comunicação, em que esses homens procuravam retratar para outras conquistas, caçadas e até mesmo histórias. Essa é considerada uma das primeiras formas de comunicação conhecidas. Desde então, assim como a evolução do ser humano, os meios de comunicação evoluíram. Assim, pode-se dividir a história das comunicações em dois períodos distintos: o das comunicações pré-eletricidade e os da comunicação pós-eletricidade.

As comunicações do período pré-eletricidade são as que foram desenvolvidas com instrumentos, meios ou artefatos que não faziam uso da eletricidade para seu funcionamento. Para essa categoria, o homem fez uso de recursos como sons, através de tambores e sinos, por exemplo; sinais de fumaça; luzes de fogueiras; sinalizações visuais através de bandeiras e até mesmo através de redes de elaborados semáforos mecânicos

Já as comunicações do período pós-eletricidade têm seu marco inicial com a invenção do telégrafo elétrico por Samuel Morse em 1837. Nesse processo, foi criado um código que permitia o envio de mensagens. A primeira linha comercial de telégrafo foi implantada em 1843 nos EUA.

Em 1876 ocorre a invenção do telefone por Alexander Graham Bell, dando início ao período de comunicação de voz à distância. Basicamente os meios de comunicação até então faziam uso de códigos que precisavam ser convertidos para serem entendidos pela linguagem humana, que era o caso do telégrafo. Isso tornava necessário um operador do sistema que conhecesse os códigos utilizados para compreensão da mensagem, algo que restringia sua utilização a um grupo restrito de especialistas, que deveriam dar suporte as necessidades de

comunicações das demais pessoas. Com o telefone utilizando diretamente a voz humana, abriu-se a oportunidade de pessoas comuns poderem utilizar o serviço de maneira mais natural.

2.2. Evolução da Comunicação de Voz através da Rede de Telefonia

O telefone se popularizou rapidamente. Logo isso também demonstrou a inviabilidade da estrutura da interligação que ele apresentava, em que todos os telefones se interligavam, em uma topologia conhecida como “*fully meshed*” (Figura 1). Nesse modelo, para haver uma conversação, era preciso uma ligação direta de fios entre o originador e o receptor.

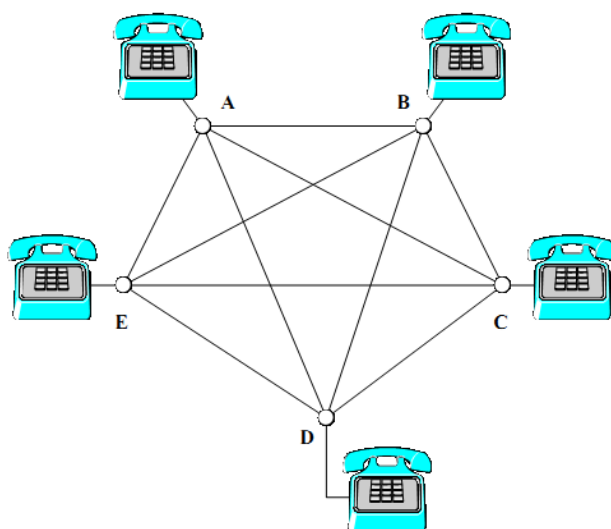


Figura 1 - Topologia "Fully Meshed" [4]

A solução encontrada para esta questão foi a criação de uma estrutura intermediária, uma central telefônica, instalada em um prédio no qual todos os telefones seriam ligados. Para fazer uma ligação, o usuário, ao retirar o fone do gancho, fazia com que o seu aparelho telefônico gerasse uma corrente elétrica que sinalizava em uma mesa de operação, em que trabalhava um/uma telefonista. (Figura 2).

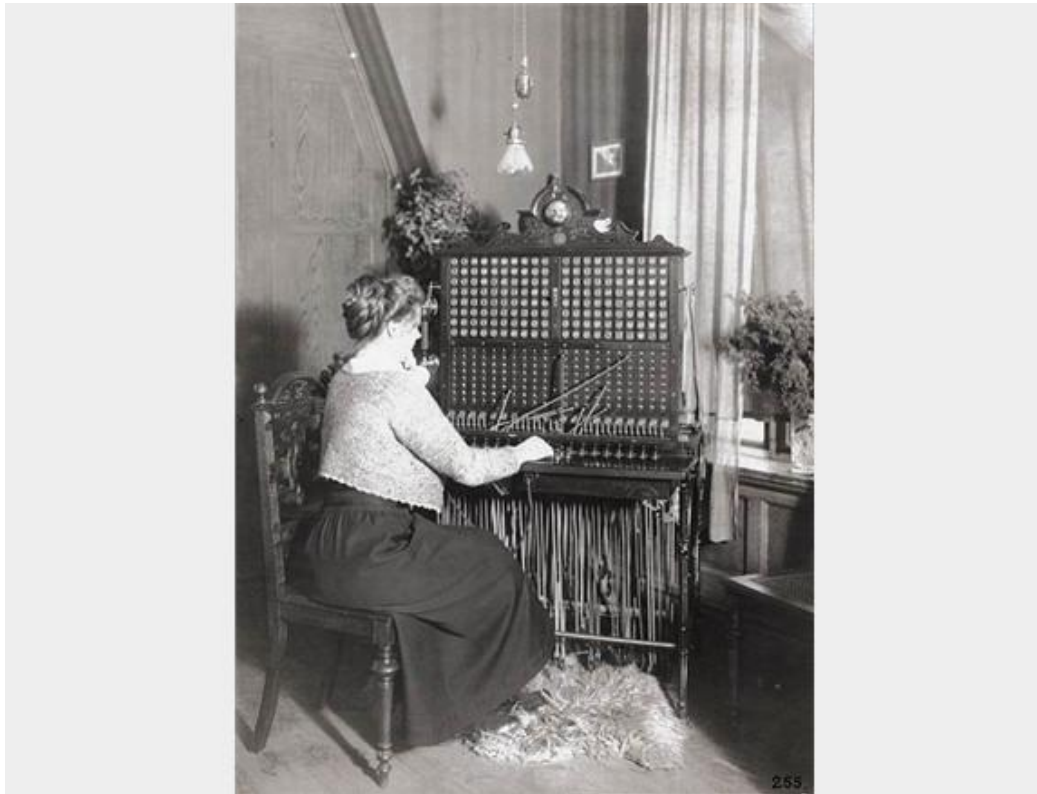


Figura - 2 – Mesa de Operação e Telefonista [3]

Essa pessoa atendia a ligação então e fazia uma conexão física do fio do telefone desejado através de fio “*jumper*”, disponível na mesa de operações. A primeira central desse tipo foi implantada em 1878 em New Haven, Connecticut, EUA (Figura 3).



Figura 3 - Central telefônica com diversos operadores [5]

Com o crescimento do serviço de telefonia, houve a necessidade de se instalar mais centrais telefônicas em áreas geográficas distintas. Isso foi necessário devido às limitações de distância para se ligar todos os telefones, principalmente sob dois aspectos: comprimento dos fios e a quantidade deles chegando a um mesmo prédio (Figura 4).



Figura 4 - Fios de telefonia "poluindo" a cidade [2]

A próxima necessidade decorrente do rápido crescimento da telefonia foi a interligação as diversas centrais telefônicas entre si para permitir a comunicação entre diversas localidades. Como ocorreu com a ligação dos primeiros aparelhos telefônicos, logo se verificou que seria inviável interligar todas as centrais diretamente umas com as outras, assim como a topologia "*fully meshed*". Para resolver esse problema, criou-se hierarquias de centrais telefônicas (Figura 5) onde as centrais nos níveis hierárquicos inferiores faziam as ligações com os telefones dos clientes e as centrais nos níveis superiores faziam as ligações das chamadas entre as centrais dos níveis inferiores [2][4].

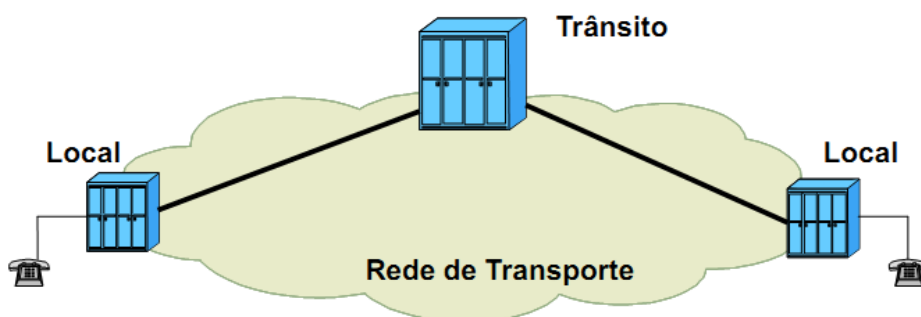


Figura 5 - Hierarquia de centrais [4]

Essa estrutura hierárquica das redes de telefonia permaneceu praticamente inalterada por mais de cem anos e ainda é a base da rede de telefonia fixa comutada, também conhecida como PSTN (*Public Switched Telephone Network*). No entanto isso não significa que não houve evolução tecnológica nesse período. Essa ocorreu de forma sistemática e contínua, gerando novas padronizações, tecnologias e produtos que permitiram a efetiva interligação do mundo através de uma rede que presta um serviço de voz robusto e confiável. Além disso, muitas dessas tecnologias geradas pelas necessidades da rede de telefonia são hoje a base para as redes de nova geração que buscam a integração dos diversos serviços, como é o caso das fibras ópticas e das redes de comunicação móvel. As evoluções tecnológicas mais significativas nessa rede podem ser encontradas nas seguintes formas:

- Conexão de ligações telefônicas internamente às centrais, a chamada “comutação” das chamadas,
- Troca de informações entre centrais para estabelecimento de chamadas (os chamados protocolos de sinalização) e
- Transmissão das informações de voz entre as centrais (ou meios de transmissão).

A seguir são relacionados os principais avanços ocorridos em cada um desses grupos.

2.2.1. Comutação das Chamadas

A primeira evolução na forma de conectar, ou comutar, chamadas telefônicas ocorreu com a invenção da primeira central automática de comutação, que não dependia dos operadores para estabelecer as chamadas, desenvolvido por Almon B. Strowger em 1891 [2]. Era um equipamento eletromecânico que possuía engrenagens e relês chaveavam as ligações a partir de pulsos elétricos recebidos dos aparelhos telefônicos que originavam as chamadas. Essa invenção permitiu que as chamadas fossem realizadas mais facilmente, com maior velocidade e com um grau de privacidade que não era possível com as centrais operadas manualmente.

A invenção do transistor em 1947 pela *Bell Labs* permitiu o desenvolvimento de circuitos eletrônicos menores que vieram substituir a comutação eletromecânica pela totalmente elétrica. Com isso reduziu-se o espaço físico ocupado pelos equipamentos, aumentou-se a confiabilidade e reduziu-se o consumo de energia necessário para operar as centrais telefônicas.

A partir da invenção e rápida evolução do computador em 1945 e do circuito integrado em 1958. Isso possibilitou que ao longo dos anos houvesse o desenvolvimento das primeiras centrais telefônicas cujo controle de sua operação era realizado por computador, as denominadas CPA – Central de Programa Armazenado.

O desenvolvimento da eletrônica digital e a invenção do microprocessador em 1971 permitiram a próxima evolução nas centrais telefônicas. Isso possibilitou o desenvolvimento da digitalização completa das informações tratadas internamente nas centrais. Dessa forma, a informação analógica, que antes era transmitida do aparelho telefônico do assinante, é convertida para informação digital logo na interface de entrada da central telefônica, e todo o tratamento posterior dentro do sistema, incluindo a comutação, é feita digitalmente. Isso permitiu um incremento na capacidade de processamento de chamadas das centrais, redução de tamanho físico juntamente com redução do consumo de energia.

Surgiu então a tecnologia TDM (*Time Division Multiplexing*). Esse sistema, baseado no conceito de modulação PCM (*Pulse Code Modulation*) que converte o sinal analógico em um código binário para ser transmitido digitalmente, trouxe algumas melhorias para a conversação. Entre elas, a permissão de várias chamadas simultâneas entre os meios de comunicação; a regeneração de sinal não piora a relação sinal/ruído mesmo a longas distâncias; e a criação de níveis hierárquicos que conseguem transportar grande volume de canais de conversões.

A partir desse ponto, a evolução das redes de telefonia segue em direção ao sistema NGN, tema que será abordado no Capítulo 4 desse estudo.

A figura 6 apresenta a evolução da transmissão da telefonia, a partir do início feito apenas com sinais analógicos até o sistema totalmente digital.

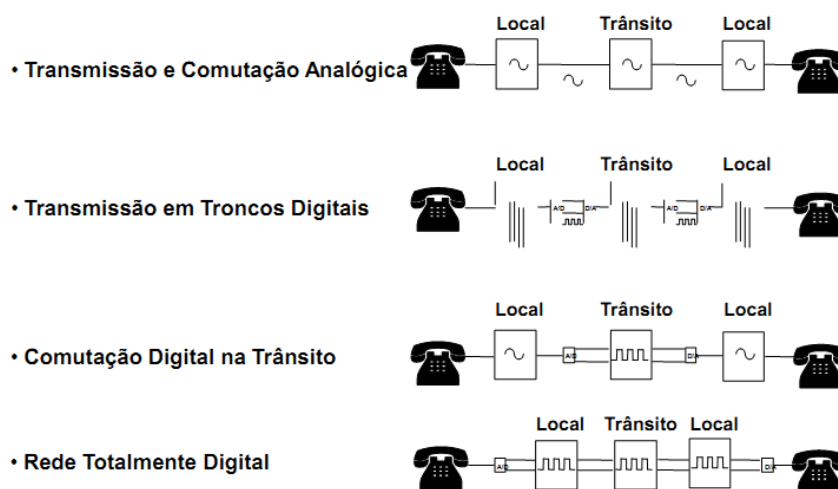


Figura 6- Evolução da Transmissão da Telefonia [6]

2.2.2. Protocolos de Sinalização entre Centrais Telefônicas

Os protocolos de sinalização (ou simplesmente: sinalização) definem de que forma os sistemas telefônicos interagem com os assinantes e entre si, visando estabelecer e completar uma ligação da origem ao destino.

A sinalização do assinante é a que permite a um usuário acessar o sistema através de seu telefone, ao tirar o telefone do gancho, selecionar o número do telefone desejado e identificar se a ligação está em andamento ou eventualmente se não será completada através de tons audíveis e padronizados (como o “tom de ocupado”). É ela que envolve as informações trocadas entre o aparelho telefônico do assinante e a central a qual ele está ligado. Essas informações utilizam como recurso o chaveamento de corrente, quando o telefone é “tirado do gancho” ou colocado de volta; e tons informativos, como o envio de dígitos do telefone desejado; e tons de controle gerados pela central telefônica, que são os tons de controle de chamada, como o de ocupado.

A sinalização entre as centrais telefônicas permite que a chamada para um número de telefone que não é atendido pela central de origem seja encaminhado para a central que atende o assinante chamado. Sua evolução e padronização permitiram que a rede telefônica se tornasse um serviço de alcance global.

A figura 7 abaixo exemplifica uma interação entre as sinalizações de assinante e entre centrais, durante uma chamada telefônica.

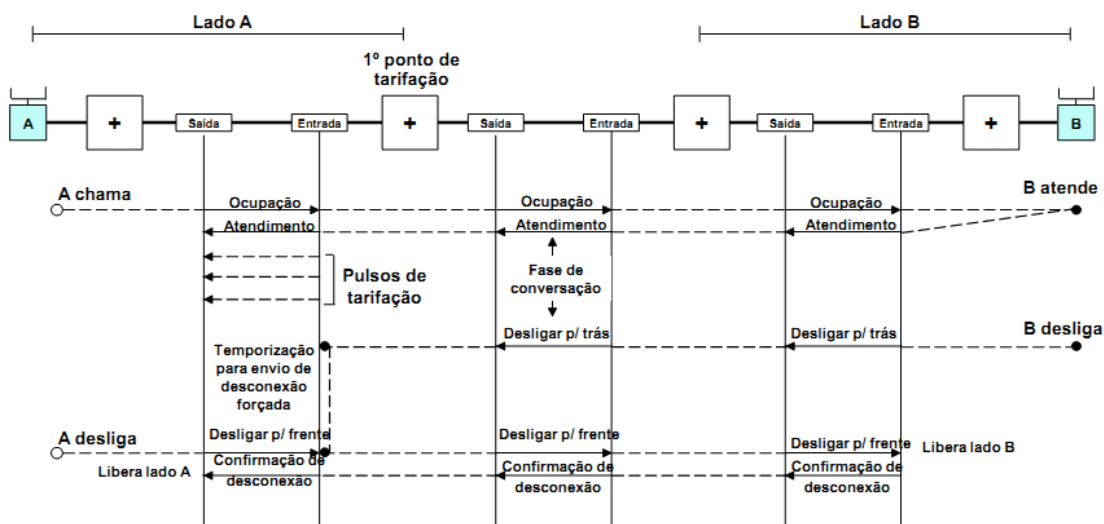


Figura 7 - Interação de sinalizações de uma chamada telefônica em uma rede TDM [6]

2.3. Evolução na Comunicação de Dados

A comunicação de dados, entendida como a troca de informações entre sistemas informatizados, teve seu início a partir da invenção do computador. Um dos principais modelos

iniciais foi o Harvard Mark I, que foi criado em 1944, em uma parceria entre a IBM e a Universidade de Harvard para fins militares (Figura 8) [7].

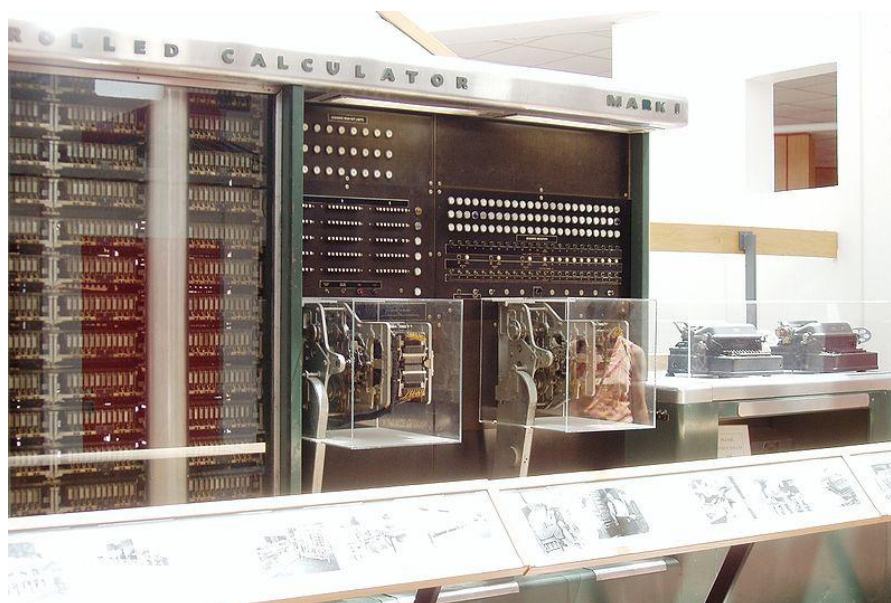


Figura 8 - Computador Harvard Mark 1 [7]

Somente na década de 1950 os computadores foram introduzidos comercialmente. Ao longo das próximas décadas, essa invenção se popularizou motivada pela redução de preço proporcionada pela evolução e miniaturização dos circuitos eletrônicos principalmente após a invenção do transistor.

Inicialmente os dados e informações eram introduzidos diretamente nos computadores através de chaves e botões integrados nos próprios equipamentos. Posteriormente foram desenvolvidos terminais remotos que permitiam enviar dados para o computador e receber relatórios do mesmo. Isso possibilitou que o terminal remoto ficasse em um local distinto de onde estava instalado o computador, limitado ao comprimento dos fios necessários para fazer a ligação entre eles.

Na década de 1960, desenvolveu-se o primeiro MODEM (Modulador/Demodulador), pela AT&T. A partir dessa inovação, era possível agora a transformação dos sinais elétricos das interfaces que ligavam os computadores em sinais acústicos de voz, possibilitando sua transmissão através da rede de telefonia. Desta forma, a comunicação de dados nasceu da necessidade de ligar os computadores aos terminais remotos que ficavam em locais geograficamente distantes para entrada de dados e recebimento dos relatórios processados. Essas foram consideradas as primeiras redes de comunicação de dados, que utilizavam a rede de telefonia pública e modems para transmissão de dados. Com a evolução dos sistemas computacionais, tanto em hardware como em software, os computadores passaram a trocar informações entre si para realizar as requisições que lhes eram feitas. Essa necessidade de

comunicação de dados entre os computadores permitiu o desenvolvimento do conceito de rede de computadores.

Assim, em 1969, surgiu a INTERNET, quando a rede ARPANET, patrocinada pela Agência de Defesa Americana do EUA, começou a operar ligando quatro computadores das principais universidades americanas que participavam do projeto [6].

Com a ARPANET foi introduzido o conceito de transmissão de pacotes, onde as informações em bits originais a serem enviadas como um relatório, por exemplo, eram separados em pequenos grupos, denominados pacotes, que eram então transmitidos através da rede de dados.

O conceito de transmissão de pacotes permitiu uma utilização mais eficiente dos meios de transmissão, sendo que enquanto esse meio não estava sendo utilizado para transmitir dados de um determinado cliente, teria a disponibilidade de transmitir os dados de um outro. No caso de comutação de circuitos, o meio fica reservado para a comunicação em curso mesmo que não haja nada sendo transmitido. É o caso da pausa entre uma conversa telefônica: apesar de nenhum dos envolvidos utilizarem o canal, este fica alocado do mesmo modo para essa conversação. Com isso, foram desenvolvidas as primeiras redes de comutação de pacotes, que embora não utilizassem as centrais de comutação de voz, faziam uso dos cabos e sistemas de transmissão de longa distância já implantados para telefonia.

A partir da década de 1970, foram desenvolvidos diversos protocolos para comunicação de dados, tendo sido o primeiro o X.25, seguindo-o o *Frame Relay* e posteriormente ATM (*Asynchronous Transfer Mode*). Embora esses tenham sido os principais, outros protocolos os substituíram e, com o tempo, prevaleceram [6].

A próxima evolução veio com a invenção do microcomputador em 1976, também conhecido como computador pessoal. Pouco tempo depois surgiram as LANs (*Local Area Network*) ou redes locais de computadores. Finalmente, com o crescimento e expansão global da Internet, juntamente com a adoção do conjunto de protocolos associados a ela, surgiu o conjunto de protocolos conhecido como TCP/IP, o que permitiu a rápida expansão das redes de comunicação de dados e iniciou a mudança dos meios de transmissão à longa distância das redes TDM para comutação de pacotes IP.

Uma importante inovação para as redes de comunicação de dados foi a criação de modelos de referência para a transmissão na rede. Surgiu então o conceito do Modelo OSI (*Open Systems Interconnection*) e o TCP/IP. Esses definiram a forma comum de conectar os computadores [9].

2.3.1. Modelo OSI [8] [9]

A arquitetura OSI apresenta a rede de computadores dividida em setes camadas (Figura 9), facilitando assim a implementação de protocolos que executam diferentes funcionalidades. Essa divisão se apresenta de forma hierarquizada, ou seja, cada camada usa as funções dela própria ou da camada anterior para evitar complexidade e revelar as operações aos usuários, podendo ele ser outra camada ou um programa [9].



Figura 9 - Modelo OSI com os principais componentes de cada camada.

A primeira camada, chamada de Camada Física, define as características técnicas das entidades elétricas e ópticas (físicos) do sistema. Ela contém os equipamentos de cabeamento ou outros canais de comunicação que se comunicam diretamente com o controlador da interface de rede. Com isso, ela permite uma comunicação simples e confiável, com controle de erros mais básicos.

A segunda camada é a de enlace ou ligação de dados. É ela quem faz a transmissão e recepção de pacotes e pelo controle de fluxos. Além disso, essa camada estabelece o protocolo de comunicação entre os sistemas conectados e pode opcionalmente corrigir erros que ocorreram na camada física.

A camada de Rede é responsável pelo endereçamento dos pacotes de rede, também conhecidos por datagrama, associando endereços lógicos em endereços físicos, fazendo com que os pacotes cheguem corretamente ao destino. Essa camada também determina a rota que os pacotes irão seguir para atingir o objetivo, utilizando fatores como condições de tráfego da rede e prioridades. Ela é utilizada quando já mais de um caminho para o pacote percorrer.

Já a camada de transporte divide os dados enviados pela camada de Sessão em pacotes que serão transmitidos para a camada de Rede na transmissão. Na recepção, ela é responsável por captar os pacotes recebidos da camada de Rede, remontar o dado original e enviá-lo à

camada de Sessão. Além disso, inclui ainda tarefas como controle de fluxo, ordenação de pacotes e correção de erros.

A camada de transporte ainda separa as camadas de aplicação (5 a 7) das camadas de nível físico (1 a 3). É o elo entre esses dois grupos e determina a classe de serviço necessária.

A camada de Sessão permite que duas aplicações em computadores diferentes estabeleçam uma sessão de comunicação. Nesta sessão, essas aplicações definem o modo como que será feita a transmissão de dados, além de colocar marcações nos dados que estão a ser transmitidos. Caso haja falha da rede, os computadores reiniciam a transmissão dos dados a partir da última marcação recebida.

A camada de Apresentação é responsável pela conversão do formato do dado recebido pela camada de Aplicação para outro que possa ser usado na transmissão que seja compreendido pelo protocolo usado.

Por fim, a camada de Aplicação responde por prover serviços para aplicações de modo a abstrair a existência de comunicação em rede entre processos de diferentes computadores. É nessa camada que ocorre a interação micro-usuário.

No entanto, a implementação do modelo OSI não é seguido integralmente. Por isso determinou-se como o modelo implantado o conjunto de protocolos TCP/IP, que contem apenas cinco camadas.

2.3.2. Modelo TCP/IP [8] [10]

O TCP/IP é um conjunto de protocolos que, baseando-se ao modelo OSI, é usado para comunicação entre computadores em rede. A grande importância dessa arquitetura é que a Internet atual baseia-se nesse conceito [4].

Seu nome é composto de dois protocolos principais o TCP (*Transmission Control Protocol*) e o IP (*Internet Protocol*). Pode-se fazer uma analogia e comparar os modelos TCP/IP e OSI. De modo geral, as camadas apresentadas na figura 10 são correspondentes.

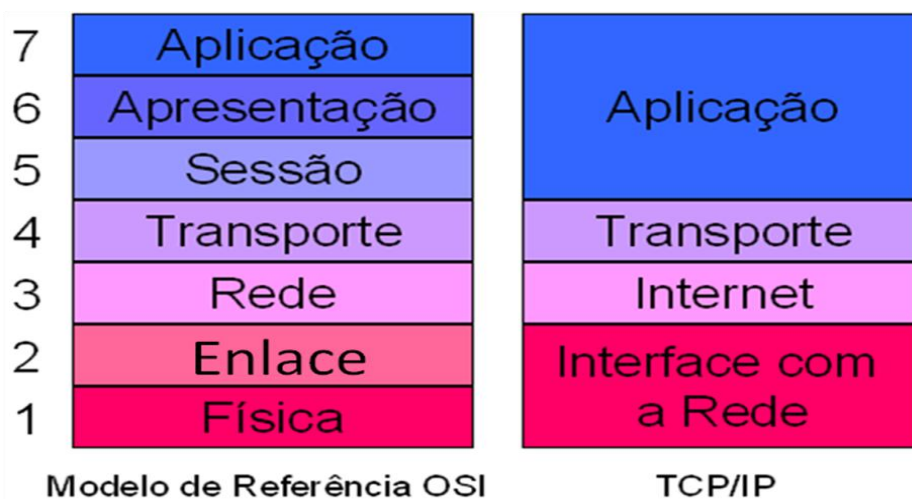


Figura 10 - Comparação Modelo OSI e Modelo TCP/IP

No TCP/IP, a camada de rede utiliza do protocolo IP, enquanto a camada de transporte pode se utilizar de protocolos como o TCP, que é um protocolo orientado a conexões confiável que permite a entrega sem erros de um fluxo de bytes; ou o UDP (*User Datagram Protocol*), que fornece integridade de dados, mas não fornece entrega garantida.

Atualmente, a maioria dos sistemas operacionais utiliza o modelo TCP/IP como padrão.

2.4. Evolução na Comunicação Móvel [11] [13] [14]

O conceito de transmissão de códigos de comunicação pelo ar surgiu no final do século XIX, através do desenvolvimento do conceito de ondas eletromagnéticas desenvolvido por Maxwell.. Graças a essa idéia inicial foi possível o desenvolvimento da comunicação móvel. O celular nada mais é que um aparelho de comunicação por ondas eletromagnéticas que permite a transmissão bidirecional de voz e posteriormente de dados. Sua invenção ocorreu em 1947, novamente nos laboratórios Bell.

Mas foi somente em 1956 que a Ericsson apresentou o primeiro protótipo comercial, conhecido por Ericsson MTA (*Mobilie Telephony A*). Esse primeiro modelo, pesando cerca de 40 quilos, foi desenvolvido para ser colocado em porta malas de carros. Somente no início da década de 70, a Motorola apresentou seu celular: *Motorola Dynatac 8000X*, responsável por fazer a primeira ligação de um aparelho celular (Figura 11). Esse protótipo se mostrou mais portátil, uma vez que pesava apenas um quilo e tinha cerca de 25 cm de comprimento.

No entanto, a implantação da primeira rede de telefonia celular ocorreu no Japão e na Suécia, em 1979. A rede norte americana só começou a operar em 1983.



Figura 11 - Martin Cooper e o primeiro modelo de celular comercial [12]

2.4.1. Primeira Geração

O primeiro padrão ou primeira geração dos sistemas celulares para a comunicação das redes móveis foi o AMPS (*Advanced Mobile Phone System*). Esse modelo, formado apenas por sistemas analógicos, utilizava o múltiplo acesso por divisão de frequência (FDMA). Foi implantado nos EUA em 1983 e utilizava a frequência de 800 MHz.

No AMPS, a banda era dividida em canais de RF (Radiofrequência), em que cada canal consiste de um par de frequências com 30 kHz de banda cada, em que uma faz a transmissão e a outra a recepção. Outra característica é que esse padrão utilizava modulação FM e era composto por 416 canais, sendo 21 para controle e o restante para voz. Esse canal de voz, ao ser alocado, permanecia dedicado somente à chamada durante toda a duração.

Já na Europa foram implantados outros sistemas para a 1ª geração. Baseando-se no mesmo princípio do FDMA, adotaram-se os seguintes sistemas:

- *Total Access Communications System* (TACS) no Reino Unido, Itália, Espanha, Áustria e Irlanda;
- *Nordic Mobile Telephone* (NMT) em vários países nórdicos;
- C-450 na Alemanha e em Portugal;
- Radiocom 2000 na França; e
- *Radio Telephone Mobile Systems* (RTMS) na Itália.

No entanto, por conta da diversidade de sistemas analógicos operando, surgiu a dificuldade da incompatibilidade entre eles. Como cada um tinha um padrão próprio, não era permitido que um telefone de um sistema operasse em outro. Também com a popularização do serviço, logo as bandas não comportavam mais os números de chamadas. Surgiu então a necessidade de aumento de capacidade [11] [13].

2.4.2. Segunda Geração

A segunda geração dos sistemas celulares surgiu com a digitalização nos meios de acesso na interface aérea, substituindo as transmissões analógicas por digital. Isso foi motivado pela tentativa de aumento da capacidade, uma vez que o sistema da primeira geração já não comportava o serviço.

Assim, surgiram os sistemas TDMA (*Time Division Multiplexing Access*) e CDMA (*Code Division Multiplexing Access*) que formaram a base das principais tecnologias adotadas. São elas [13]:

- TDMA IS-54: evolução do AMPS, padrão americano, permitiu que em cada canal analógico fosse utilizado para trafegar 3 canais digitais TDMA, ampliando assim a capacidade do sistema. Acabou sendo implantado em todo o continente americano;
- CDMA IS-95: utilizando técnica de espalhamento espectral, utilizava a mesma faixa de frequência do sistema analógico AMPS. Não era compatível com o TDMA IS-54. Foi adotado na América do Norte e também em alguns países da América Latina e da Ásia;
- GSM (*Global System for Mobile Communications*): foi o padrão criado na Europa. O GSM, em especial, visou ser adotado em todos os países daquele continente para substituir os inúmeros sistemas analógicos que não eram compatíveis entre si. Foi concebido adotando acesso TDMA com até 8 canais de voz para cada canal na interface aérea de largura de banda de 200 kHz definidos pelo GSM. Ele operava inicialmente na faixa de 1.800 MHz, mas em pouco tempo também começou a operar também nas faixas de 900MHz (Europa). Posteriormente foi utilizado nos EUA nas faixas de 850 MHz e 1.900 MHz. Desta forma se tornou um padrão mundial, sendo adotado em toda a Europa, diversos países da Ásia e também nas Américas, incluindo EUA e Canadá. Nos dias de hoje continua em operação, mesmo com a grande popularização da tecnologia 3G.

2.4.3. A Transmissão de dados via Sistemas Celulares

A partir das tecnologias de segunda geração foi introduzida a transmissão de dados via pacotes através das redes celulares. A partir dessa tecnologia, o conceito de convergência de rede já esteve presente na solução. Assim, as tecnologias disponibilizadas para transmissão convencionou-se chamar de geração 2,5. Entre as principais estão:

- GPRS (*General Packet Radio Service*): O GPRS foi desenvolvido para utilizar a infraestrutura de rede GSM para transmissão de dados via pacotes, permitindo taxas de transmissão de dados de 56 a 114 kbps a possibilidade de acesso à Internet em modo

contínuo (“*always on*”). A rede GPRS não utiliza as centrais de comutação de circuitos, utilizando sistemas de comutação de pacotes para sua implementação.

- EDGE (*Enhanced Data rates for GSM Evolution*): O EDGE é a evolução do GPRS que através de melhorias na modulação para transmissão de dados, permitiu um aumento para taxas para até 230 kbps.

Essas tecnologias continuam em uso, apesar das taxas de transmissão de dados consideradas baixas para os dias de hoje. Elas ainda atendem bem para aplicações como telemetria e terminais de ponto de venda, como cartões de crédito.

2.4.4.A Tecnologia de 3ª Geração de Rede Celular

Visando principalmente ampliar a capacidade de transmissão de dados em alta velocidade através da rede celular, foi desenvolvida a tecnologia de 3ª Geração de Redes Celulares. Entre as propostas estabelecidas a que foi amplamente adotada em nível global foi a UMTS (*Universal Mobile Telecommunications System*). Essa tecnologia, apesar de adotar conceitos definidos no GSM, utiliza novas estações radio base e novas alocações de frequência (2.100 MHz na Europa) e mais comumente a técnica de espalhamento espectral denominada W-CDMA. As taxas de transferências de dados previstas nas versões iniciais do padrão UMTS alcançavam 384 kbits/s. Como evolução do UMTS inicial foi desenvolvido o HSPA (*High Speed Packet Access*), que é conhecida como geração 3,5 e que permite atingir maiores taxas teóricas de *downlink*.

É nessa geração que desenvolveu-se o conceito do IMS, objeto principal desse estudo, visando melhorar a qualidade dos serviços oferecidos. Posteriormente o IMS será estudado no capítulo 5.

Como próxima etapa de evolução para as redes celulares, a 4ª Geração, está prevista taxas de até 100 Mbps. Essa velocidade poderá ser alcançada utilizando nova forma de codificação na interface aérea e cujos principais padrões concorrentes para adoção mundial a candidatos a se tornar padrão global são o LTE (*Long Term Evolution*) e WiMAX (*Worldwide Interoperability for Microwave Access*), padrão desenvolvido por empresas norte americanas. Nessa nova “guerra” de padrões, o LTE aparenta surgir com mais força para ser adotado como padrão em função da escolha das grandes operadoras mundiais, já que se aproveita de uma parte da infra-estrutura já existente nas redes 3G [14].

3. Convergência de redes

Pelo contexto da evolução dos meios de telecomunicações, houve a necessidade das prestadoras de serviço oferecerem serviços inovadores e diferenciados para poderem competir no mercado. Assim, elas não podiam ficar restritas oferecendo apenas um único serviço. Era necessário encontrar uma solução em que elas poderiam diversificar os serviços e conquistar seus consumidores.

Esse capítulo tem como objetivo descrever essa situação e mostrar quais foram as razões que motivaram as operadoras a buscar a convergência de suas redes. Também visa apontar quais foram as necessidades e premissas utilizadas, além de apresentar as principais alternativas e idéias para garantir a entrega dos serviços de maneira eficaz. Tudo isso para demonstrar como se chegou ao modelo atual de convergência de redes, também conhecida como NGN.

3.1. Ambiente Pré-Convergência [15]

Com a evolução das telecomunicações, surgiram diversos serviços que podiam ser oferecidos pelas operadoras aos usuários. Telefonia fixa, móvel, televisão a cabo, internet, entre outros, passaram a fazer parte do cotidiano das pessoas. No entanto, no início, para cada novo serviço criado, houve a necessidade de uma rede própria para ele ser oferecido.

Com a grande quantidade de novos projetos aparecendo, podem ser destacados três tipos de serviços que se generalizam:

- Voz
- Vídeo
- Dados

No caso da voz, destaca-se a telefonia. Essa pode ser separada em dois grupos distintos:

(i) Telefonia fixa: com a tradicional rede TDM que permite a transmissão simultânea de vários sinais, sendo que cada sinal tem um tempo definido para uso da banda;

(ii) Telefonia móvel: conhecida também como rede de telefonia celular, onde se encontram diversas tecnologias, como a TDMA, CDMA e GSM.

Para o caso de serviços de vídeo, outros conceitos foram desenvolvidos e empregados. Entre as mais importantes, pode-se citar:

(i) TV a cabo: sistema de distribuição de conteúdos audiovisuais de televisão, através de cabos coaxiais fixos. Utilizado principalmente como televisão por assinatura.

(ii) TV via satélite: utilização de comunicação via satélite para distribuição de canais de televisão.

E por fim, para oferecimento de serviços de dados, houve a criação de redes de computadores, sendo o conjunto de protocolos mais usado o TCP/IP.

Em um cenário em que cada serviço necessita de uma rede própria para ser oferecido, configura-se um ambiente com muitas limitações tanto por parte dos serviços oferecidos aos clientes quanto para as operadoras e controladores da rede. A seguir serão discutidos alguns pontos que tornam essas redes únicas tão limitadas.

3.1.1. Limitações de Infraestrutura [15]

Primeiramente, para a maioria das redes especializadas era necessário um meio próprio para a sua transmissão. Isso quer dizer que para cada serviço seria preciso uma rede específica para o transporte, podendo ela ser constituída de cabos coaxiais, pares trançados, fibras ópticas e o ar. Assim, se uma operadora quisesse ofertar mais de um serviço precisaria adotar outra infraestrutura diferente daquela que ela já possui. Por isso, para cada serviço a ser ofertado no mercado, sempre havia uma operadora específica para ofertá-lo.

O segundo ponto é o alto custo de operação e manutenção de uma rede específica. Era necessária a contratação de peritos especializados na infraestrutura, uma vez que esta, por ser muito especializada e fragmentada, não estava apta a oferecer um único padrão.

Ainda como restrição de infraestrutura própria, específica, observou-se uma grande dificuldade de oferecer novos serviços. Eles acabaram se tornando muito limitados e houve uma grande dificuldade de oferecimento de serviços com valor agregado. Também nas dificuldades dessas restrições de redes, havia a necessidade de um terminal específico para cada serviço. Por exemplo, para assistir um vídeo, havia a necessidade de uma televisão, ou para realizar uma ligação, a necessidade de um telefone.

Desse modo, para a obtenção de diversos serviços, o usuário precisaria contratar várias operadoras distintas, cada uma com a sua rede própria e cada uma delas com um perfil diferente. Além disso, outra dificuldade para usuário era o pagamento para cada serviço. Não era possível o estabelecimento de um padrão, uma vez que a cobrança dependia da região, da operadora e dos meios de transmissão, como em casos como a cobrança de *roaming* e mudanças de operadoras.

3.1.2. Outras Dificuldades [15]

Entre diversas dificuldades encontradas, um ponto importante a se destacar é que mesmo com várias operadoras ofertando o mesmo serviço, suas redes de transporte se tornaram muito semelhantes, ou seja, sem mudanças significativas entre a rede de transporte de uma ou de outra. Isso causou certo nivelamento entre as empresas que ofereciam os serviços, não tendo assim grandes diferenciais entre elas para disputarem os clientes.

E também, diante da exclusividade de algumas tecnologias desses serviços, a interconexão adotada posteriormente com outras redes e mesmo outras tecnologias tornou-se muito difícil. Portanto, mesmo com as interconexões houve uma enorme dificuldade na oferta de novos serviços, uma vez que a tecnologia e rede é muito específica.

Por isso, apesar dessas interconexões apresentarem uma saída para esse problema, a manutenção de redes muito específicas acabou se tornando inviável, principalmente pelas empresas que procuravam oferecer maior quantidade de serviços e aplicações. Assim, começou-se a desenvolver o conceito de redes que poderiam oferecer mais de um serviço, primeiramente voz e dados. Começou-se então o incentivo para o desenvolvimento desse projeto. Começou-se a desenvolver o conceito de rede convergente.

3.2. Justificativas para a Convergência de Redes e Serviços [15]

Surgiu inicialmente uma idéia de convergir para as redes de voz e dados, uma vez que a demanda de ambas era alta. No entanto, esse projeto também deveria ser aberto, ou seja, fosse uma solução viável para oferecimento também de novos serviços que pudessem surgir, como foi visto depois com a incorporação de serviços de vídeo. Além disso, devia apresentar um bom custo benefício para a operadora. Também com uma rede única, o controle de tráfego seria facilitado. Haveria uma melhor gerência, uma vez que a matriz é comum a todos os serviços, facilitando o suporte e controle de qualidade dos serviços ofertados.

Com uma nova rede, haveria a possibilidade de criação dos novos serviços de valor agregado que poderiam ser desenvolvidos para essa nova estrutura, aumentando o número de opções de fornecimento de produtos pela operadora, gerando assim mais receita.

Pelos motivos citados, verificou-se que a convergência de redes era a melhor opção para fornecimento de novos serviços e redução de custos no mercado de telecomunicações. Era uma alternativa viável, mas algumas observações importantes tiveram que ser observadas para garantir o bom funcionamento dessa rede.

3.3. Premissas para a Convergência de Redes [15]

Primeiramente, essa rede deveria oferecer QoS (*Quality of Service*) e transparência fim a fim, uma vez que muitos dos serviços como voz necessitam de um alto padrão de segurança e qualidade de entrega.

Outro fator importante é a interoperabilidade dessas novas redes com as exclusivas já existentes, via interfaces abertas, uma vez que é inviável substituir todas as redes de uma vez. Assim, uma interconexão entre serviços novos e os já existentes seria criada, permitindo à operadora a utilização de ambas as redes para oferecimento de seus produtos.

Além disso, a mobilidade na rede de usuários, terminais e serviços era desejável para que cada ponto da rede não ficasse especializado. Em outras palavras, era necessário que qualquer ponto da rede tivesse capacidade própria de oferecer o serviço.

Outra característica fundamental para que as redes pudessem ofertar diversos serviços era a necessidade de suporte para tantos terminais já existentes de outros tipos de rede quanto dispositivos para os novos terminais convergentes. Além disso, a rede deveria ter uma grande flexibilidade de configuração, uma vez que deveria suportar as diferentes tecnologias de acesso utilizadas. Por fim, por ser uma rede aberta, ela também deveria fornecer mecanismos que garantam a segurança a fim de proteger a troca de informações sobre essa nova infraestrutura.

Era necessário o suporte para uma grande variedade de serviços e aplicações, uma vez que a ideia era a criação de novas opções aos usuários, baseadas na construção modular de serviço, que incluem tráfego em tempo real, não real, *streaming* e multimídia. Também era preciso garantir a convergência de serviços fixos e móveis e integração de redes sem fio com outros tipos de rede, possibilitando assim uma maior abrangência para os usuários.

Levando em conta todas essas premissas, desenvolveu-se um conceito de convergência viável às operadoras. A seguir o início desse processo é descrito.

3.4. Início da convergência [4] [16] [17]

A grande motivação das empresas prestadoras de serviços para buscar o processo da convergência era a redução de custos de operação e manutenção do sistema, além de poder oferecer aos clientes maior quantidade de opções para oferecimento de serviços.

O início da convergência das redes se deu com os serviços de voz e dados. Isso devido à popularização da Internet e à necessidade da telefonia. Somente posteriormente outros tipos de mídia como o vídeo foram incluídos. Inicialmente, as redes de dados e voz, tanto fixa como móvel apresentavam suas próprias estruturas de redes, como apresentado na figura 12.

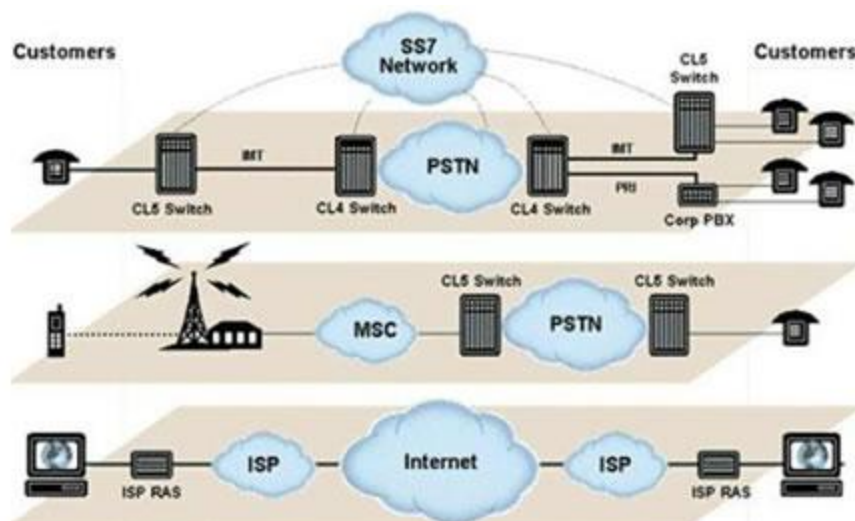


Figura 12 - Redes própria para telefonia fixa, móvel e dados [17]

Como mostra a figura 12, cada uma dessas redes era constituída de seus próprios protocolos e tecnologias para a transmissão de suas mensagens. Para que houvesse uma convergência entre as redes, era necessário a definição de um padrão que comportasse a transmissão de outros serviços.

Com o avanço das redes de computadores nas décadas de 80 e 90 e sua popularização, a transmissão de voz sobre dados tornou-se uma alternativa rentável e promissora. As tecnologias dessas redes permitiriam o transporte de informações e dados através do empacotamento desses.

Nesse processo escolhido para a transmissão, a voz, para operar na rede de dados, necessitava de alguns processamentos. Era preciso digitalizá-la, comprimi-la e empacotá-la. Esse último processo é feito por CODECS, responsáveis por criar esses pacotes de dados para trafegar na rede. Após essa etapa, esses pacotes sofrem o processo de multiplexação, fazendo com que a voz seja transmitida na forma de dados juntamente com os dados da rede.

Para todo esse processo de transmissão de voz via rede de dados, surgiram várias tecnologias interessantes que poderiam trazer grandes vantagens quando implantadas. Elas são conhecidas como tecnologias de comutação. Três delas merecem destaque: Voz sobre IP ou VoIP; Voz sobre Frame Relay ou VoFR; e Voz sobre ATM, ou VoATM.

3.4.1. Voz sobre Frame Relay [4] [8] [16]

O *Frame Relay* é uma tecnologia de comunicação que é utilizada para a transmissão de informação digital através de redes de dados, de forma barata e rápida. Esses dados são divididos em *frames* (quadros ou pacotes) que são utilizados para a transferência da informação

de um ponto de origem a outro de destino, obedecendo as regras e formatos estabelecidos. Essa tecnologia foi padronizada pelo ITU e é muito eficaz para o transporte de voz através de WANs.

O Frame Relay é considerado muito confiável, uma vez que opera na camada 1 (transporte) e principalmente na 2 (enlace) do modelo OSI, utilizando redes locais com um serviço orientado a conexão, tudo com baixo retardo e sem controle de erro nos nós. Ele proporciona orientação à conexão em sua camada de trabalho (modelo OSI - Camada de Enlace de Dados ou camada 2).

O funcionamento do Frame Relay baseia-se no estabelecimento de circuitos virtuais (VC – *Virtual Circuits*). Para a explicação do funcionamento operacional do *Frame Relay*, há alguns elementos principais a serem considerados:

- FRADs (*Frame Relay Access Device*);
- PVCs (*Permanent Virtual Circuit*);
- DLCIs (*Data Link Connection Identifier*);
- CIR (*Committed Information Rate*).

Os FRADs são os dispositivos de acesso à rede *Frame Relay*. Eles são os responsáveis pela entrega de pacotes à nuvem FR, no formato determinado pelo protocolo. Eles também recebem os pacotes, repassando-os ao usuário final. São eles quem escolhem os caminhos na rede para a entrega dos pacotes. Esses caminhos correspondem aos PVCs, circuitos permanentes virtuais que são programados para interconectá-los remotamente.

Cada FRAD tem é identificado por um endereço. Assim, os identificadores desses endereços em cada pacote são chamados de DLCI [16].

Existe ainda o CIR, que é a largura mínima de banda que a rede irá garantir. Isso quer dizer que ele é responsável por quantizar o mínimo permitido para o transporte de pacotes entre os FRADs e é contratado juntamente com a operadora que fornece a rede.

A rede com tecnologia Frame Relay normalmente é representada por uma nuvem, já que ela não é uma conexão física entre dois pontos distintos. Essa conexão é feita através do PVC configurado com o CIR. A alocação de banda física na rede é feita pacote a pacote no caso de transmissão de dados, diferentemente da tecnologia TDM que aloca uma banda fixa na rede indiferente de tráfego ou não de dados.

Basicamente, o funcionamento de uma rede *Frame Relay* começa com a chegada do pacote aos FRADs. Lá, a informação é convertida em dados e empacotada e define-se a rede e o caminho a ser utilizado (PVC). Definidos essas questões, o FRAD é responsável também pelo envio do pacote.

Cada pacote lançado na rede carrega as informações de origem e destino, através do DLCI. Através dele é feito o processo de encaminhamento e entrega do pacote. Entre os FRADs de origem e de destino existe um acordo que determina quantos frames no mínimo podem ser enviados entre eles, definidos pelo CIR. Assim, o pacote é entregue ao seu

destino. A figura 13 apresenta o esquema de funcionamento da rede *Frame Relay* e as posições de cada um dos seus principais elementos.

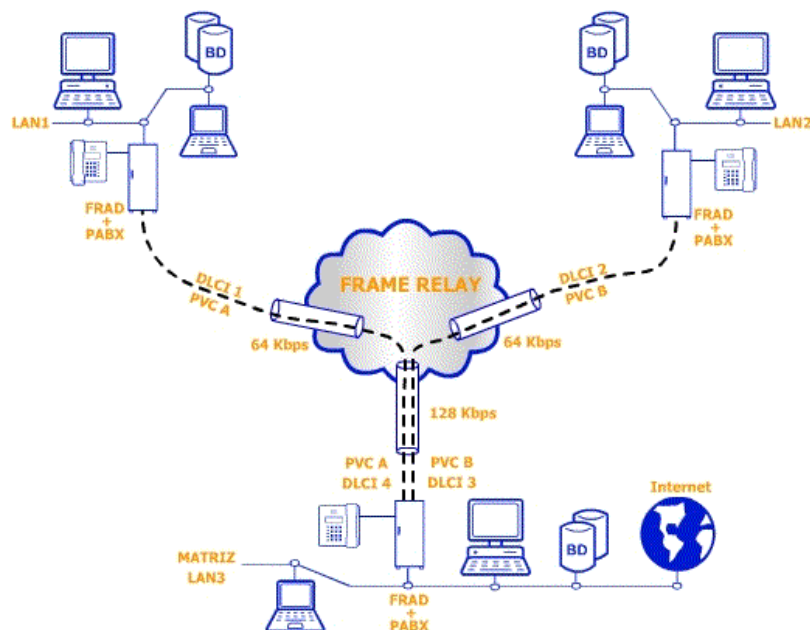


Figura 13 - Rede com Tecnologia Frame Relay [16]

Para redes locais remotamente interconectadas, mais conhecidas como WANs, o *Frame Relay* é a técnica altamente recomendada. Ela é flexível e eficiente para o transporte de dados, conectando várias localidades com um excelente custo/benefício. No entanto, com a popularização do IP e do ATM, além do aumento da utilização de VPN (*Virtual Private Network*) e outros serviços de acesso dedicados com o DSL (*Digital Subscriber Line*) acabaram por reduzir a utilização do Frame Relay, que aos poucos foi perdendo espaço.

3.4.2. ATM [4] [8] [16]

O ATM, do inglês *Asynchronous Transfer Mode* ou Modo de Transferência Assíncrona, é um protocolo padronizado pelo ITU. Foi considerado por especialistas como o protocolo "ideal", uma vez que foi projetado para implementação com alta qualidade de serviço e ambiente multimídia e multi-serviços na comunicação. No entanto, apesar desse rótulo de "ideal", seu custo de implementação foi considerado muito alto, dificultando assim sua popularização. O ATM surgiu no início da década de 90 e foi projetado como um protocolo de comunicação de alta velocidade que independe de qualquer topologia de rede específica. Usa uma tecnologia de comutação de células com alta velocidade de transmissão que pode tratar tanto dados como vídeo e áudio em tempo real.

A estrutura de uma rede ATM consiste basicamente de equipamentos terminais que são conectados a centrais ATMs. Nessas centrais, são encapsulados os dados em pacotes, denominados células, em um tamanho fixo de 53 bytes, sendo o cabeçalho (ou *header*) composto de 5 bytes e 48 bytes de dados (no chamado *payload*). Na figura 14 é apresentada essa célula. No entanto, nem sempre esse cabeçalho conterá informações de endereçamento e de camada superior ou mesmo informações de controle de pacote. A transmissão é escalonável, ou seja, as células podem ser transportadas em redes LAN, WAN ou mesmo MAN. A banda em redes ATM chega a ser superiores a 155 Mbps. Ela depende principalmente do tipo de cabeamento utilizado, podendo ser cabos de cobre ou até mesmo fibras ópticas.

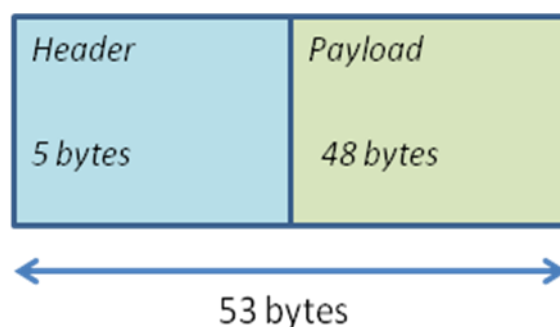


Figura 14 - Célula ATM

Por possuírem um tamanho relativo pequeno (53 bytes), essas células são mais facilmente gerenciadas pelos equipamentos, gerando maior velocidade de tráfego na rede ATM.

Assim como o *Frame Relay*, o ATM se baseia na criação de circuitos virtuais para o transporte de dados. Para o caso do transporte de voz, o sinal é digitalizado e em seguida transformado em pacotes no formato de células, para assim poder ser transportado na rede ATM. Apesar da ênfase dada a outros serviços que requerem altas velocidades, o transporte de voz também encontrou bons benefícios ao utilizar essa infraestrutura, obtendo assim qualidade de serviço e melhor desempenho.

Para o transporte na rede ATM, desenvolveu-se alguns métodos para melhorar o desempenho da rede: os AAL, *ATM Adaptation Layer*. Esses são responsáveis pela alocação dinâmica ou estática da banda. Dependendo do tipo de mídia a ser transmitida e de suas necessidades de banda, pode-se utilizar o AAL1 a AAL6.

No caso da voz, são utilizados apenas o AAL1 ou o AAL2:

- Camada de aplicação AAL1: utilizado para voz não comprimida com taxa de bits constante (CBR), ou seja, já está definida pela operadora a alocação estática da banda de transmissão, mesmo quando não há dados a transmitir.

- Camada de aplicação AAL2: utilizado para voz comprimida com taxa de bits variáveis (VRB), ou seja, tem alocação dinâmica da banda de transmissão, otimizando assim a utilização da banda contratada, conforme a quantidade de pacotes a ser transmitida.

Apesar de teoricamente apresentar-se como o melhor meio para transporte de serviços, não somente a voz, em redes de dados, o alto custo de implantação de uma rede ATM impede que se essa se popularize mais e seja amplamente utilizado.

3.4.3. VoIP - Voz sobre IP [4] [8] [16]

A tecnologia Voz sobre IP, ou VoIP é a mais utilizada para o transporte de voz sobre redes de dados. Ela apresenta o maior crescimento, uma vez que usa o protocolo IP, utilizado pela grande maioria das redes em todo mundo [16]. VoIP é uma tecnologia que possibilita o tráfego da voz na forma de dados utilizando os serviços de rede do protocolo IP [16]. De outra forma, pode-se dizer que o VoIP coloca a transmissão de voz como mais um serviço suportado pela rede de dados (Figura 15).

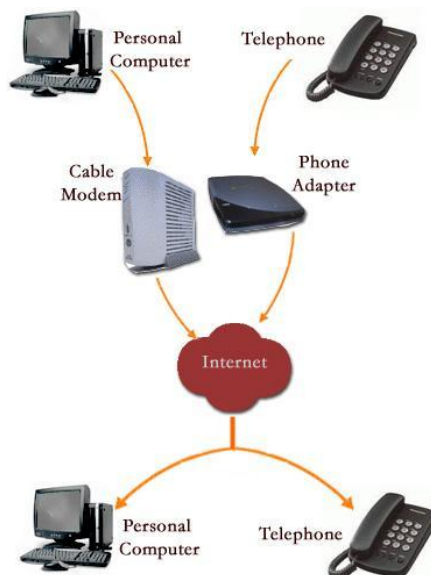


Figura 15 - Convergência de voz em dados via IP

O protocolo IP, do inglês *Internet Protocol*, é um protocolo de comunicação, utilizado para encaminhamento de dados. Juntamente com o TCP (*Transmission Control Protocol*), faz parte do grupo de protocolos utilizados para a comunicação de computadores em rede, tanto em uma rede local quanto em redes remotamente interconectadas. O protocolo IP não se limita apenas à Internet. Ele é um protocolo de comunicação entre computadores em rede, podendo ser utilizado por redes que não sejam ligadas à Internet, por exemplo, redes corporativas privadas.

Aqui, a diferença é que a Internet utiliza-se do protocolo TCP/IP para entregar suas informações através dos planos de endereçamento do protocolo IP.

Esse conjunto de protocolos TCP/IP define o conjunto de regras a serem seguidas para o gerenciamento e controle da transmissão de dados em uma rede. Isso faz com que todos os equipamentos da rede se comuniquem em uma mesma linguagem, evitando diferenças na comunicação entre os equipamentos, que tem fabricantes distintos. A definição de camadas para um conjunto de protocolos específico tem por objetivo separar os diversos tipos de informações que podem trafegar sobre este conjunto de protocolos, visando facilitar sua implantação e gerenciamento. O TCP/IP é responsável pelo gerenciamento de tráfego entre computadores de uma rede.

O VoIP baseia-se na arquitetura desse protocolo TCP/IP. Nessa estrutura, cada equipamento na rede deve ter um endereço único para ser identificado na rede, identificar uma conexão à inter-rede: o endereço IP (Figura 16).

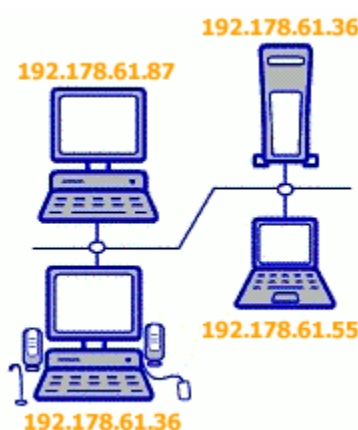


Figura 16 - Endereços IP [16]

É importante ressaltar que voz sobre IP significa voz sobre o protocolo IP, o que leva a uma interpretação errada, uma vez que muitas pessoas associam VoIP a “voz sobre internet” ou mesmo a algum sistema de tecnologia de telefones para a Internet [16]. Na arquitetura TCP/IP, a voz é um tipo de serviço ou aplicação que pode ser transportada através da rede, de um usuário que possui um endereço IP de origem para outro que possui um endereço IP no destino. Para o controle dessas chamadas, há dois protocolos principais utilizados pelo VoIP:

- H.323, padronizado pelo ITU (*International Telecommunication Union*); e
- SIP, *Session Initiation Protocol*, padronizado pelo IETF;

O H.323 apresenta um alto grau de complexidade enquanto o SIP apresenta um modelo bem mais simples. Cada um deles apresenta prós e contras, mas quando bem implementados são praticamente equivalentes, sem interferir na qualidade de serviço.

Na realidade, como dito anteriormente, a qualidade do VoIP depende diretamente da rede que o transporta. No Capítulo 4, esses protocolos serão detalhados com maior profundidade para justificar suas utilizações em redes convergentes.

3.5. Necessidades de uma Rede Convergente [4] [16]

Já foram apresentadas algumas das principais tecnologias utilizadas para a convergência das redes, principalmente no requisito da voz em redes de dados. No entanto para que essa técnica seja eficaz em sua tarefa, há a necessidade de que o usuário final esteja satisfeito com os serviços oferecidos. Por essa razão, a qualidade de serviço é uma medida que permite que a satisfação do cliente seja atingida, garantida pela estabilidade operacional da comunicação. No caso da comunicação de voz, existem três fatores principais que determinam a qualidade operacional:

- Confiabilidade do sistema
- Inteligibilidade da comunicação
- Boa relação custo/benefício

Atendendo esses três pontos principais, o nível de satisfação do cliente tende a aumentar. Segundo resolução do ITU, qualidade de serviço é o resultado da atuação coletiva de diversos fatores num sistema que determina a satisfação do usuário de um serviço de telecomunicações [16].

As redes de telefonia convencionais são conhecidas pela sua qualidade de serviço e confiabilidade do sistema, além de apresentar um excelente custo benefício. Essas tecnologias de convergência entre voz e dados, através do empacotamento de dados, tem evoluído a fim de obter a mesma qualidade de serviços das redes de telefonia tradicionais. E é exatamente por essa qualidade que o legado de telefonia ainda continua em operação: a substituição de toda rede seria uma medida inviável. Além do alto custo para a implantação, a rede ainda é muito segura e oferece confiança aos serviços já oferecidos por ela e garante o lucro às operadoras. Trata-se de um sistema já consolidado há muitos anos e por isso sua migração para a convergência tende a ser feita lentamente.

Também para um bom funcionamento de uma rede convergente, existem alguns parâmetros que devem ser fixados a fim de garantir uma boa qualidade da transmissão de voz em uma rede de dados. São eles:

- Codificação e decodificação de sinal de voz;
- Largura de banda e compressão de voz;
- Atraso e flutuações de atraso; e
- Congestionamento e perda de pacotes.

O ponto principal para a transmissão da voz é a necessidade de transmissão em tempo real. Isso para garantir a inteligibilidade e interatividade entre os usuários. Como o sinal de voz é analógico, para ser transformado em dados necessita que passe por um processo de digitalização.

Essa conversão digital consiste na transformação do sinal analógico em pulsos, em que o conteúdo binário (bits) contém as informações. Essa digitalização é chamada de codificação e o processo inverso de decodificação.

Esse processo de codificação é feito pelos CODECs(CODer-DECoder). Esses precisam ser exatos, pois qualquer alteração pode fazer com que o sinal transmitido se difere do original.

O processo mais comum para essa codificação é o PCM, ou Modulação por Codificação de Pulsos. O PCM converte um sinal analógico de voz em um sinal digital correspondente a uma taxa de 64 kbps [16].

3.5.1. Codificação do Sinal de Voz- Processo de Digitalização da Voz [4] [6] [16]

Para a digitalização do sinal analógico da voz, utiliza-se o PCM. Nele, existem três etapas fundamentais (Figura 17):

- Amostragem: em que o sinal é dividido em pequenas amostras, conforme uma frequência já pré-determinada;
- Quantização: em que o sinal amostrado sofre um refinamento;
- Codificação: transformação do sinal quantizado em binário.

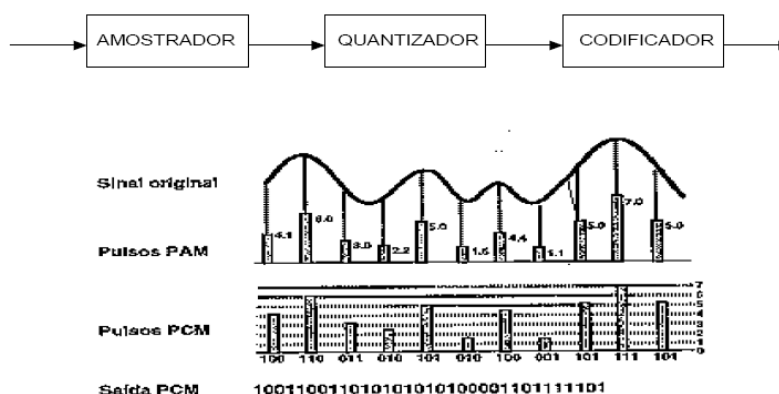


Figura 17 - Processo PCM [6]

Nesse tipo de modulação um sinal analógico em digital para que possa ser transmitido através de um meio físico com transmissão digital.

O processo consiste inicialmente em coletar amostras do sinal analógico, passo conhecido como amostragem. Essa amostra respeita o Teorema de Nyquist, em que o número

mínimo da amostra precisa ser duas vezes maior que a largura de banda do sinal que será digitalizado. Essa regra é utilizada para garantir que a reconstrução do sinal seja idêntica a original.

O segundo passo é chamado de quantização das amostras. Esse processo consiste no mapeamento do sinal, em que seus valores amostrados são “arredondados” a outros já pré-definidos. Em outras palavras, o ponto em questão é avaliado a partir de uma linha de nível: caso esteja acima dessa marcação, é aproximado para o valor correspondente superior; e estando abaixo, para o valor inferior. Esse processo é uma aproximação e portanto não corresponde aos valores reais do sinal. Por esse fator podem ocorrer erros. Por isso, quanto maior o número de níveis de quantização utilizados, menor será o erro causado pelo processo.

Para o PCM, há 256 níveis de quantização. Como ele utiliza a codificação binária, cada palavra-código tem 8 bits e cada bit pode assumir o valor 0 ou 1, resultando assim em $2^8 = 256$ níveis. Para o sinal de voz é o suficiente para apresentar menos erros.

Após o sinal ser quantizado, ele sofre o processo de codificação, sendo transformado em um sinal binário, considerando a sequência do sinal inicial. O resultado final é uma sequência composta de 0 e 1 que representa o sinal analógico. Esse novo sinal binário agora é capaz de ser transportado em uma rede de dados. Por todo esse processo, a utilização do PCM é muito vantajosa: Apresenta apenas dois níveis diferentes para o sinal modulado e reduz formas de ruídos que podem interferir sobre o sinal modulado, o que pode ocasionar dificuldades na decodificação de sua forma original.

3.5.2. Largura de Banda e Compressão [4] [16]

Ao considerar os inícios da convergência, o transporte de voz em redes de dados poderia causar um grande congestionamento na rede. Isso se dá devido ao tráfego de voz apresentar uma alta taxa, de 64 kbps. A fim de evitar esse tipo de problema, adotou-se a compressão dos dados antes da transmissão, otimizando a utilização da banda. Existem várias tecnologias responsáveis por fazer compressão. Uma das mais conhecidas é a G.729 que permite que o sinal de voz seja reduzido de 64kbps para 8kbps.: uma compressão de 8:1.

A técnica de compressão deve ocorrer logo no início de todo o processo, assim como o processo da descompressão deve ocorrer no fim. Isso garante a naturalidade e a inteligibilidade do sinal original. Essas técnicas são capazes de reduzir o custo de implantação de redes, principalmente das WANs.

3.5.3. Atrasos, Congestionamento e Perda de Pacotes [4] [16]

Outro fator importante a se considerar no transporte de voz em uma rede de dados é o atraso da comunicação. Como a voz necessita de tempo real, esses atrasos podem comprometer seriamente a qualidade do serviço.

Essas variações de atraso entre uma transmissão e outra são chamadas de *jitter*. É considerado variável, pois muda em tempo real, de acordo com as condições de tráfego e de congestionamento e ocorrem devido à ocupação ou a desocupação da rede na hora do envio de informações. Outro parâmetro que também pode afetar a qualidade do sinal é o atraso da comunicação que pode ser gerado no transporte da voz numa rede de dados.

Durante o processo de transmissão, caso a rede esteja desocupada para envio de um bloco de dados, o envio se dá imediatamente. Entretanto, no caso de ocupação da mesma, é necessário que haja a desalocação de banda suficiente para o transporte da informação. Esse tempo de espera causa um atraso na transmissão. Normalmente percebe-se esse atraso quando há uma ruptura na fala, ou a audição de uma voz artificial. Fora isso, com o congestionamento da rede, pode haver descarte e perdas de pacotes, fazendo com que o usuário final não receba a informação toda.

A fim de reduzir esse problema de perda de pacotes, desenvolveram-se algumas técnicas para evitar e diminuir esses atrasos. Uma delas, chamada de priorização de pacotes faz com que os pacotes de voz tenham prioridade na transmissão em relação aos outros. Isso faz com que a voz sempre seja transmitida antes, dando assim uma maior garantia da entrega correta dos pacotes e evitando assim atrasos.

3.5.4. Qualidade de serviço [4] [6]

Apesar da utilização de transporte de voz em redes de dados, o parâmetro principal que garante o sucesso da implementação, como dito na seção 3.5 é a qualidade do serviço. Como o ITU define, qualidade de serviço é o resultado da atuação coletiva de diversos fatores num sistema, que determina a satisfação do usuário de um serviço em questão. Com QoS pressupõe-se a entrega constante, previsível e satisfatória de um serviço, processo ou informação. Esta funcionalidade visa definir mecanismos de qualidade em que elementos de uma rede receberão parâmetros específicos de tráfego, para garantir [2]:

- A largura de banda de transmissão necessária para uma aplicação específica;

- Atrasos fixos e flutuações mínimas de atraso (*jitter* mínimo).

Sendo assim, pode-se considerar que é a partir do QoS que são definidas as funções e os procedimentos no sistema.

3.6. QoS na rede IP

Garantir Qualidade de Serviço para tráfego de sinais de voz em redes IP talvez seja o maior desafio para a integração total das redes de dados e de telefonia. A rede IP, apesar de ter sido escolhida juntamente com os protocolos TCP/IP para servir como a principal rede de convergência de voz e dados, apresenta uma característica que pode dificultar esse processo. Essa característica é conhecida como ‘melhor esforço’, ou ainda ‘serviço de datagramas não confiável’.

O ‘melhor esforço’ é conhecido por tratar os pacotes a serem enviados da mesma maneira. Não há nenhuma priorização ou garantia da entrega dos pacotes. Caso muitos pacotes necessitem ser enviados, congestionando a rede, há a perda de alguns pacotes, que simplesmente são descartados com o propósito de liberar a rede. Como todos os pacotes são igualmente tratados, não é possível controlar quais serão perdidos. Essa característica é muito prejudicial para o transporte de voz, vídeo ou ainda alguma outra aplicação que necessita de tempo real.

A rede IP é definida por parâmetros mínimos de desempenho que uma rede deve fornecer. Por isso, problemas como a largura de faixa, latência (atrasos) e *jitter* também podem ser comuns. Para garantir a entrega e melhorar o QoS da rede, especialistas buscaram desenvolver novas tecnologias que pudessem otimizar o desempenho da rede. Entre os principais focos na rede, encontram-se a alocação de recursos na transmissão ou ainda na otimização.

No caso da alocação de recursos de transmissão, existem duas tecnologias que merecem destaque: Serviços Integrados e Serviços Diferenciados.

Já no caso de garantir QoS com foco de otimização, o MPLS (*Multi-protocol Label Switching*) é considerada uma das principais técnicas. A seguir serão apresentadas breves descrições de cada uma dessas técnicas.

3.6.1. IntServ - Arquitetura de Serviços Integrados [4] [16]

Serviços Integrados procuram garantir os recursos reservando um banda de aplicação para cada aplicação utilizada. Para a garantia do recurso, é feito primeiramente uma reserva prévia de banda para a transmissão da rede. Para isso, a aplicação a ser transmitida deve especificar todos seus requisitos de banda. Com essas informações a rede utiliza um protocolo de roteamento para poder estabelecer um caminho para a transmissão. Em cada ponto do caminho, verifica-se se há recursos suficientes disponíveis para nova reserva. Completado esse processo, a reserva do caminho é feita e a aplicação pode iniciar a transmissão e o envio de pacotes pelo caminho, que está exclusivo para essa função. Essa técnica é ideal para aplicações de que necessitam de transporte em tempo real e VoIP, que exigem qualidade de serviço e não toleram atrasos. Nessa arquitetura, os roteadores devem ser capazes de reservar recursos a fim de prover QoS para fluxo de pacotes específicos.

Um dos protocolos mais utilizados nessa função é o RSVP, *Resource Reservation Protocol*, desenvolvido pelo IETF. Ele é utilizado para controle e preparação das reservas de QoS solicitadas pela aplicação. O RSVP é executado com base no IP e no UDP (*User Datagram Protocol*), que é um protocolo de transporte não muito confiável.

3.6.2. DiffServ - Arquitetura de Serviços Diferenciados [4] [16]

Já os Serviços Diferenciados trabalham com a combinação de controle nas periferias da rede para poder provisionar os recursos da banda e priorizar os pacotes mais importantes. Nele, os dados a serem enviados são divididos em pequeno número de classes de encaminhamento no pacote IP. Para cada uma dessas classes, a quantidade de tráfego que os usuários têm direito de colocar na rede pode ser diferenciada e limitada a partir de um ponto específico da rede. Nesse processo não existe um estabelecimento prévio de reserva de banda. A alocação das classes de encaminhamento é especificada através de um acordo entre o usuário da banda de transmissão e a operadora que fornece o serviço. Assim é uma solução mais simples que a dos Serviços Integrados.

A vantagem dessa tecnologia é a simplificação e flexibilização da rede, uma vez que ameniza as dificuldades de escalabilidade já que não há necessidade de verificar sempre se há banda suficiente alocada na rede. A banda já está garantida a partir da periferia de uma rede. Assim é possível dar preferência a um determinado tipo de dado.

Nessa técnica os pacotes são classificados, marcados e processados com base no conteúdo do rótulo DSCP (*Differentiated Service Code Point*).

3.6.3. Protocolo MPLS (MultiProtocol Label Switching) [4] [16]

Já a tecnologia mais conhecida para otimizar os níveis de desempenho de uma rede IP é o MPLS (*Multiprotocol Label Switching*), que organiza os recursos na rede para garantir maior probabilidade de entrega da informação ao menor custo. O MPLS tem como técnica o *label switching* ou rótulo de encaminhamento, em que o encaminhamento de pacotes é baseado em rótulos (*labels*) adicionados aos pacotes a serem transportados. Todo encaminhamento ao longo do backbone da rede é feito com base nesse rótulo. O MPLS é responsável por estabelecer as regras para definir o melhor caminho para a entrega e quais as prioridades de serviço que esses pacotes devem ter. Da mesma forma que o Frame Relay, o MPLS estabelece rotas prévias para a transmissão de pacotes, como pode ser inferido da figura 19:

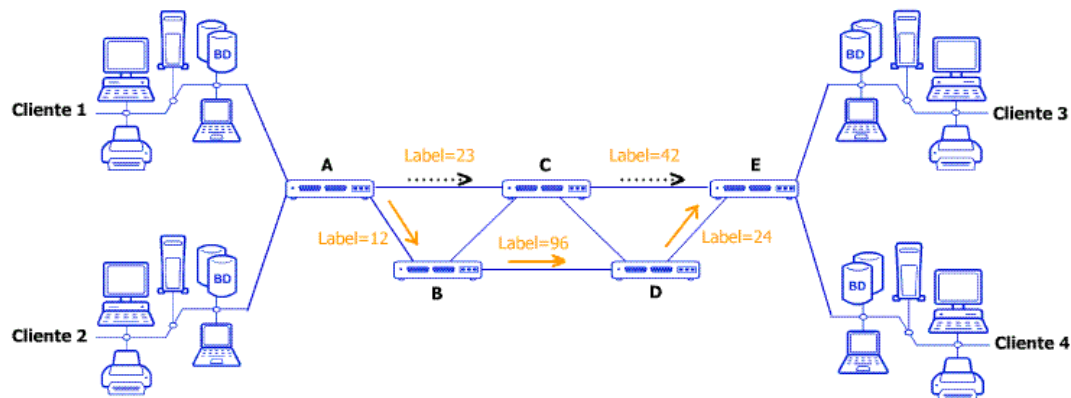


Figura 18 - Rede com MPLS [16]

Apesar de se mostrar extremamente eficaz no papel de entrega, o propósito do MPLS não é substituir o roteamento IP. Ele pretende melhorar os serviços providos em redes IP através de um roteamento pré-definido, oferecendo escopo para gerenciamento de tráfego e garantia de QoS.

Para implementar QoS, o MPLS também incorpora conceitos e ferramentas como os Serviços Integrados e os Serviços Diferenciados. Ele provê, por exemplo,

reserva de banda a ser também especificada no rótulo, e pacotes podem ser "marcados" para indicar sua prioridade. Todas estas ferramentas fazem do MPLS um mecanismo ideal para implementação de capacidades de gerenciamento de tráfego em redes IP [16].

3.6.4. Alternativas para QoS na rede IP [16]

Ainda como alternativas para a entrega em tempo real de serviços que necessitam dessa característica, foram criados os chamados protocolos de tempo real. Eles são protocolos desenvolvidos para utilização em aplicativos em tempo real, tais como videoconferência, transmissão de áudio ou telefonia sobre redes de dados, uma vez que estas aplicações não toleram atrasos na transmissão/recepção, que, no entanto, podem compensar uma eventual perda de pacotes.

Protocolos de tempo real operam, normalmente, sobre o protocolo UDP, pois o cabeçalho (*overhead*) da conexão TCP e o seu mecanismo de retransmissão e correção de erro são inúteis para aplicações em tempo real. Por isso, esses protocolos, isoladamente, “não garantem Qualidade de Serviço”, pois não proporcionam nenhum mecanismo de confiabilidade na entrega dos pacotes. No entanto, eles aumentam a eficiência da rede para esse tipo de aplicação, contribuindo para uma melhoria de desempenho nos parâmetros exigidos de QoS.

Para os aplicativos que necessitam de tempo real, como a voz e vídeo, os protocolos mais importantes são:

- RTP (*Real Time Protocol* - Protocolo de Transporte em Tempo Real), que é responsável pelo transporte dos pacotes de dados em Tempo Real;
- RTCP (*Real Time Control Protocol* - Protocolo de Controle em Tempo Real), que é responsável pelo controle e monitoração da sessão estabelecida no protocolo RTP.

Para o caso das NGN e redes com arquitetura IMS, esses protocolos são essenciais. Nos próximos capítulos serão estudados em que contexto eles se encontram.

3.7. Evolução das redes convergentes [15] [16] [17]

Nos dias atuais existem planos para convergência incluir no processo, além de serviços de telecomunicações, a tecnologia da informação (TI). Já é possível, por exemplo, o acesso de redes sociais em celulares ou mesmo transações bancárias.

Observa-se que a convergência de redes é a melhor opção para fornecimento de novos serviços e redução de custos no mercado de telecomunicações. É uma alternativa viável, no entanto há algumas observações importantes que devem ser observadas para garantir o bom funcionamento dessa rede. Outra questão importante é o oferecimento do serviço em qualquer ponto da rede, ou seja, deve ser uma rede com mobilidade.

A rede mais utilizada para a convergência atualmente é a rede IP, que utiliza a comutação de pacotes para envio das informações. Por ser o modelo mais atual de convergência, a utilização da rede IP merece destaque especial. O conceito adotado pelas operadoras é a *Next Generation Networks*, ou Redes de Nova Geração, ou ainda simplesmente NGN, que se baseia na rede IP. Essas já estão sendo implantadas e serão abordadas no próximo capítulo.

4. NGN - Next Generation Networks [4][6][17][18]

Com a globalização, a tecnologia se popularizou e diminuiu as distâncias entre a vida profissional e pessoal de cada indivíduo. Assim, as operadoras se depararam com a necessidade de alcançar o cliente com uma solução completa, em que ele possa escolher entre os diversos serviços os que mais lhe agradam. A solução encontrada pelas operadoras foi o conceito de NGN (*Next Generation Networks* ou Redes da Nova Geração). Com ele torna-se possível a migração da rede tradicional de serviços em uma rede convergente baseada no protocolo IP, mantendo a oferta de serviços já prestados e abrindo a possibilidade de novas opções de serviços.

4.1. Características

Com a popularização dos serviços de dados, como a Internet, o sistema de voz começou a entrar em decadência, devido a novos serviços que surgiram, como o VoIP. As operadoras fixas, preocupadas com esse cenário, precisavam de uma alternativa para manter seus antigos clientes e conquistar novos. A solução encontrada foi a migração para uma infraestrutura única que pudesse integrar os serviços, convergindo numa rede única, a NGN. Esta se baseia no oferecimento de serviço *triple play*, ou seja, é capaz de oferecer voz, vídeo e dados. Sua transmissão se dá através do encapsulamento das informações e envio através do protocolo IP. Essas redes de nova geração tem como principais características os seguintes aspectos [17]:

- Modelo de arquitetura funcional padronizado e aberto com suporte as redes legadas.
- Qualidade de serviço fim-a-fim.
- Plataformas de serviço com aprovisionamento e controle para múltiplas redes.
- Segurança adequando protocolos e API's (*Application Programming Interfaces*).
- Mobilidade generalizada, garantindo a interconectividade e o *roaming* ao usuário.

No entanto, para um bom desempenho, existem algumas condições básicas que a NGN deve satisfazer. São elas [18]:

- Comum acordo entre todos os envolvidos no processo que são os prestadores de serviços, prestadores de conectividade e prestadores de conteúdo.
- Existência de interoperabilidade internacional entre diferentes redes para o tráfego de dados e serviços prestados.
- Facilidade para incorporar novas tecnologias e serviços nas redes.

Também há algumas necessidades específicas para alguns pontos específicos das redes que a NGN precisa garantir. São eles [18]:

a) Para Usuários Finais:

- Serviços móveis com ampla cobertura (internacional);
- Alta rentabilidade das redes;
- Simplicidade na operação de serviços;
- Qualidade de serviços (QoS) negociáveis e constante em SLA's (*Service Level Agreement*) globais;
- Parcerias eficientes para bilhetagem de serviços entre as redes.

b) Para Provedores de Serviços:

- Serviço de criação de plataformas abertas e rápidas;
- Eficiente capacidade de informação dos usuários sobre os serviços disponíveis;
- Garantia de QoS das rede no que se refere a: transmissão, atraso, variação de atraso, perda, prazo de reparo, segurança, bilhetagem e disponibilidade;
- Capacidade de adaptar os serviços prestados de acordo com a variação de QoS ou limitação de dispositivos e/ou tecnologias de forma a atender o mercado no momento adequado.

c) Para Provedores de Redes

- Real convergência para elevação da produtividade e da eficiência e para a redução dos custos.

Por todos esses requisitos, o processo de implantação de uma rede NGN é bem complexo e exigente. Por essa razão, ele é feito de modo gradual. Nessa implantação há outros aspectos que precisam ser considerados no projeto [17].

Primeiramente, há necessidade de mudanças nos sistemas de chaveamento e roteamento dos provedores de serviços de redes. As interfaces de redes fornecidas pelos provedores de redes também devem ser bem definidas e internacionalmente acordadas. A rede também deverá ter um excelente controle de seus recursos, na forma de gerências escalonadas por: QoS, SLA (*Service Level Agreement*) e CRM (*Customer Relationship Management*). Tudo isso para permitir o desenvolvimento de novos serviços que possam operar de maneira eficiente e independente.

Também devido à incerteza dos serviços a serem oferecidos no futuro, há uma grande dificuldade de uma ação proativa nas redes. Medidas estão sendo tomadas como o desenvolvimento de infraestrutura de comunicação programável. A idéia básica por trás deste conceito é de desenvolvimento de uma plataforma e uma interface aberta correspondente que poderá ser usada para programar a rede quando se implementa um novo serviço de rede. São esforços para padronização das plataformas e aplicações de operação dos dispositivos de redes.

4.2. Padronização

Vários órgãos em conjunto estão voltados para padronizar as NGN. Eles procuraram se associar para buscar uma alternativa comum, que pudesse interoperar com as redes já existentes a partir da padronização de protocolos, interfaces e arquiteturas comuns e assim obter maior flexibilidade. Os principais órgãos envolvidos na padronização da NGN são o TISPAN (*Telecoms & Internet Converged Services & Protocols for Advanced Networks*), um grupo de trabalho do ETSI (*European Telecommunications Standards Institute*); o FGNGN (*Focus Group on NGN*); e o NGN-GSI (*NGN-Global Standards Initiative*) ambos os grupos são pertencentes ao ITU-T (*International Telecommunication Union Telecommunication Standardization Sector*) [17].

A primeira versão de padronização NGN foi feita pela TISPAN em dezembro de 2005, que estabeleceu as primeiras especificações para a implantação da NGN. Entre as principais resoluções estavam a substituição da PSTN, fornecendo todos os serviços que a mesma já fornecia e a introdução de serviços multimídia, novos serviços baseados no protocolo IP e a arquitetura das NGN, assim como seus subsistemas e o modo como operam [6] [17].

A segunda versão, chamada de REL-2, foi desenvolvida, levando em consideração outros aspectos, como [17]:

- Análise de exigências do FMC (*Fixed-Mobile Convergence*) e do FMCA (*Fixed-Mobile Convergence Alliance*).
- Análise das capacidades da rede em suportar os serviços de IPTV (*IP Television*) com ATIS IIF (*Automatic Terminal Information Service – IPTV Interoperability Forum*) e DVB (*Digital Video Broadcasting*).
- Integração do IPTV com a NGN *Services and Capabilities* utilizando IMS. Suporte para os serviços de Business e interoperabilidade com redes de empresas.

Nessa versão já houve um sincronismo entre o TISPAN e o 3GPP a fim de verificar o alinhamento dos padrões para agora a interface móvel também. Outras versões do NGN já foram desenvolvidas, incluindo novas funcionalidades e características para esse padrão.

4.3. Arquitetura

O grande diferencial da arquitetura NGN é a divisão de todo conjunto em camadas, com interfaces abertas e padronizadas. Normalmente, a rede é dividida em três camadas principais:

- Camada de Aplicação e Serviços,

- Camada de Controle,
- Camada de Acesso e Transporte ou Infraestrutura de dados.

Com essa arquitetura é possível o oferecimento de serviços multimídia em uma rede de dados integrada com as redes existentes a partir dessa organização em camadas, com interfaces abertas e padronizadas (Figura 19). Mas o principal ponto se encontra na separação da sinalização e controle da camada de aplicação e serviços, o que permitiu serviços serem oferecidos e também desenvolvidos de maneira independente.

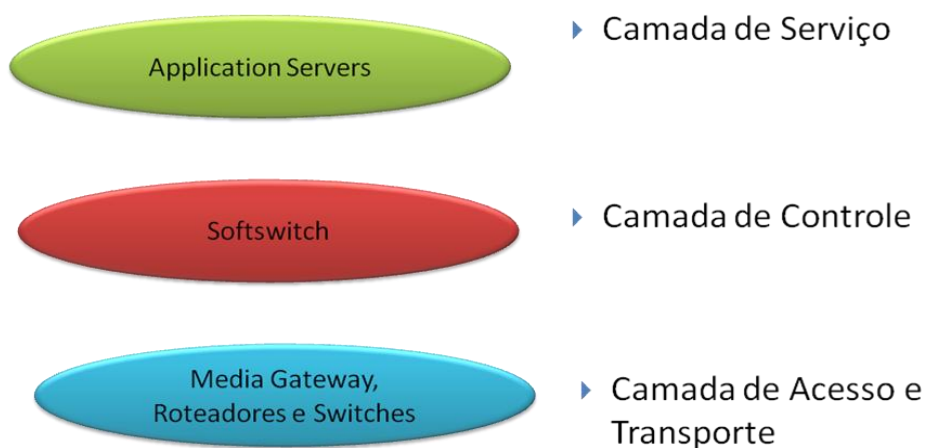


Figura 19 - Camadas de uma rede NGN

A figura 20 apresenta a topologia de rede de uma rede com arquitetura NGN. Nela pode-se comprovar as rede de acesso se comunicando com a rede IP através dos *Media Gateways* e o controle dos *Media Servers (Softswitch)*, que serão apresentados na seção 4.4.

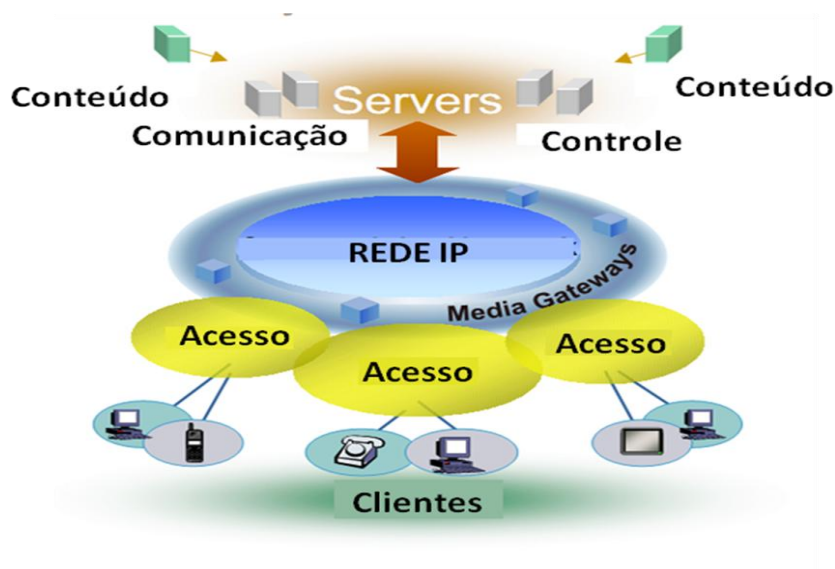


Figura 20 - Topologia de uma rede NGN

4.3.1. Camada de transporte

A camada de transporte ou infraestrutura de dados é a responsável pelo transporte dos dados. Ela que possui QoS, tanto no caso da rede IP para as NGN ou uma rede ATM. Nela também se encontra toda a infraestrutura (*backbone*) da rede de dados e redes de acesso. Nessa camada que se encontram os switches, roteadores e media gateways, responsáveis pela codificação de mídia entre redes diferentes. Também é nessa camada que os CODECs atuam sobre a informação a ser tratada.

As principais funções dessa camada são:

- Processamento da voz – fazem a compressão/ descompressão e o empacotamento.
- Transporte da sinalização para a camada de controle

Na camada de transporte também estão presentes os servidores de mídia (*Media Servers*) que são responsáveis por fazer todo o processamento da mídia, a gravação, reprodução de mensagens, conversão *text-to-speech*, reconhecimento de fala,...

Para a comunicação com a camada de controle são utilizados protocolos entre os elementos. Um dos principais utilizados é o MGCP (*Media Gateway Control Protocol*) ou posteriormente o H.248 pelos *Media Gateways*, como será visto no item 4.5.

4.3.2. Camada de Controle

A camada de controle de chamadas é considerada o ponto estratégico da arquitetura da NGN. Ela é responsável pelo estabelecimento, supervisão, liberação e tarifação das chamadas que trafegam na rede IP. É através dela que é permitido o acesso a novos serviços e aos serviços das redes legadas. É o caso, por exemplo, do oferecimento de serviços TDM em redes NGN.

As principais funções da camada são:

- Sinalização
- Controle de Chamadas
- Acesso a Camada de Serviços

É também na camada de controle que se localiza o *Softswitch*, considerado a inteligência da rede NGN. Eles são responsáveis pelo controle dos Gateways na camada de transporte.

4.3.3. Camada de Aplicação e Serviços

A camada de aplicação e serviços contém todos os serviços que podem ser oferecidos ao usuário conectados a rede. Ela é constituída de servidores ou controladores de serviços, que contem toda a lógica e inteligência para prestação dos serviços e as respectivas bases de dados dos usuários.

A camada de serviços é implementada pela utilização de workstations de alto desempenho e servidores de bases de dados padrão de mercado. Estes servidores são responsáveis por gerenciar os servidores de mídia que estão localizados na camada de transporte. Ela utiliza plataformas de hardware como Windows e Linux, que podem ser utilizados por vários serviços.

Os modos mais comuns para oferecimento dos serviços são:

- Pelo modelo da Rede Inteligente INAP;
- Pelo protocolo H.323 ou SIP.

Atualmente, entre os serviços mais comuns oferecidos pelas operadoras estão *voice-mail*, serviços pré-pagos, *unified messaging*, *voice browser* entre outros.

4.4. Equipamentos da rede NGN

Como visto na arquitetura, a rede NGN é composta por vários equipamentos na sua rede que fazem a manipulação e encaminhamento dos pacotes. Alguns dele merecem um destaque especial. A seguir serão detalhados alguns dos principais elementos das redes NGN.

4.4.1. Media Gateways

Gateways são entidades responsáveis pela interligação das redes e tradução de protocolos. Por essas funções são consideradas entidades intermediárias, que se localizam entre redes distintas. Um roteador ou mesmo um *firewall* podem ser *gateways*, já que servem de intermédio entre o usuário e a rede. Os *Media Gateways* (MGW) têm como missão a conversão de diferentes tipos de fluxo de mídia que vêm de diferentes redes. Isso quer dizer que eles são responsáveis pela interconexão entre as redes legadas e atuais. Sendo assim é possível, por exemplo, a conversão entre a rede telefônica tradicional (TDM) e a rede de dados em que serão enviadas as informações e vice-versa. Além da conversão, os media gateways fazem também a manipulação das mídias e outras atividades como a compressão, cancelamento do de eco e envio de detenção de tons. O gateway não possui nenhuma inteligência, ou seja, é controlado por outro elemento, o *Softswitch*. Esse controle se dá via protocolos de controle, como o MCGP ou o MEGACO/H.248 ou mesmo o SIP. Essa falta de inteligência é feita de maneira proposital, uma vez que facilita a gerência e evolução da rede. Caso esses elementos tivessem inteligência própria para decidir o que fazer, necessitariam de constantes atualizações para cada novo

serviço que aparecesse, tornando o processo inviável. Na figura 20 observa-se onde se localizam os MGW, assim como a sua função de tradução de protocolos entre redes distintas.

Como dito anteriormente eles também são responsáveis por fazer a troca entre o mundo TDM das telefonias fixas tradicionais e o mundo IP. O *Media Gateway* tem algumas características principais [21]:

- Interfaces diretas entre PSTN e IP/ATM
- Conversão em tempo real sobre TDM para voz sobre IP
- Protocolos de controle para uma inteligência de serviços centralizada e separação dos serviços do transporte
- Integração densa para uma baixa ocupação de espaço e consumo.

O símbolo do *Media Gateway* normalmente é retratado como o bloco apresentado na figura 21.



Figura 21 - Representação de um bloco Media Gateway [6]

4.4.2. *Softswitch*

O *Softswitch* é considerado o cérebro da rede NGN. Ele que tem toda a inteligência e controla os elementos da rede, como os gateways. Ele é responsável pelo controle de chamadas e sessões e implementa as facilidades e serviços suplementares ofertados [6]. O *Softswitch* é um *call agent* e faz parte da camada de controle das redes NGN. Call agents são responsáveis pela função de cobrança, sinalização, roteamento e ligação. Assim, o *Softswitch* tem autonomia de controlar vários media gateways dentro de uma rede TCP/IP. Além disso, ele é o principal responsável pelo controle de chamadas. Para esse controle, ele precisa trabalhar em contato com outros elementos da rede. Com isso ele é capaz de manipular a troca de sinalizações, podendo elas serem de redes de pacotes, como o H.323, SIP, MEGACO e MGCP; ou mesmo rede comutada, como SS7 e ISUP. Eles ainda são conhecidos como *Media Servers*, e são ilustrados na rede NGN na figura 20.

Para cada serviço, o *Softswitch* deve entender a sua lógica própria e traduzi-la em comandos que cada elemento da rede possa entender e realizar seu papel. No caso da voz, ele é responsável pela manipulação da sinalização e controle de chamadas juntamente com o gatekeeper. Essa divisão entre sinalização e controle e manipulação da mídia pelos gateways

permite uma maior flexibilidade para a evolução das redes. Como principais características de um *Softswitch*, podem-se destacar [21]:

- Interface com os protocolos de sinalização, como ISUP, INAP, H.323/SIP, MGCP/H.248, entre outros
- Separação do controle de chamadas/ serviços do transporte de dados;
- Mediação de alta velocidade entre diferentes sistemas de sinalização no mundo de voz;
- Inteligência de serviço em aplicações de serviços de voz;
- Inteligência centralizada para a rápida introdução de novos serviços convergentes;
- Interface aberta para dispositivos/ aplicativos comerciais ou de outros fabricantes;
- Confiabilidade e Segurança na tarifação, mediação de desempenho e controle dos recursos.

A figura 22 apresenta um bloco de um *Softswitch*, normalmente utilizado em topologias de rede:



Figura 22 - Representação de um Softswitch [6]

Tanto a parte de controle como a de sinalização é feita através dos protocolos utilizados. Na NGN são vários os protocolos que podem ser utilizados, desde aqueles utilizados para aplicações em tempo real (RTP) ou mesmo protocolos de sinalização e controle (SIP). A seguir serão descritos os principais protocolos utilizados para o controle das sessões.

4.5. Protocolos de controle da rede

A seguir serão discutidos alguns dos principais protocolos controle da rede utilizados para a NGN, principalmente em relação ao transporte de voz.

Entre os diferentes padrões utilizados, o SIP merece um destaque especial, uma vez que atua não só na rede NGN, mas é o principal protocolo utilizado no controle de sessão da rede IMS.

4.5.1. MGCP - *Media Gateway Control Protocol* [6]

O MGCP ou *Media Gateway Control Protocol* é um protocolo de *Gateway* de Mídia, definido pela IETF, utilizado para transporte de voz através da rede IP, por exemplo. O MGCP é projeto com uma arquitetura distribuída, separando o tratamento de operações de controle, sinalização e mídia e tem como principal objetivo o controle dos *Media Gateways* na rede IP através do agente externo *Media Gateway Controller (Call Agent/Softswitch)*. No entanto, já está caindo em desuso, sendo substituído pelo H.248.

4.5.2. Protocolo Megaco/ H.248 [6] [22]

O protocolo MEGACO é também um protocolo de controle para o *Media Gateway*. Foi desenvolvido pelo IETF e ITU-T, definido através da recomendação H.248. Ele permite a sinalização para redes NGN. Assim como o MGCP, sua principal função é permitir que os elementos da rede troquem informações de controle e gerenciamento de serviços.

O Megaco/H.248 adota o modelo de separação física do plano de controle (*Softswitch*), do plano de transporte (*Media Gateway*), conforme apresentado na figura 19. O *Softswitch* é responsável por trocar as sinalizações e mensagens com as outras redes e protocolos, converter as mensagens para os comandos do Megaco/H.248 e encaminhar na rede IP para os *Media Gateways*,

O MG pode estar localizado fisicamente distante do *Softswitch*, recebendo os comandos Megaco/H.248 para criar e deletar as entidades do protocolo. Ele recebe a mídia de diferentes tipos de rede e faz a conversão para a rede IP.

Apesar de existirem muitos protocolos para os sistemas de comunicação, o Megaco/H.248 se fixou como opção compatível e complementar ao MGCP, obtendo assim uma maior vantagem em relação a outros protocolos, e assim vem se tornando o seu sucessor.

4.5.3. H.323 [6] [22]

O protocolo H.323 especifica os sistemas de comunicação multimídia em redes de pacotes que não provem qualidade de serviço (QoS). Ele não se limita apenas a rede IP, mas pode agir também em outras redes de pacotes como IPX (*Internet Packet Exchange*), redes metropolitanas (MANs), redes de longa distância (WAN) e conexões discadas.

Ele também estabelece padrões para codificação e decodificação para dados de voz e vídeos, além de outras mídias. Isso garante a interoperabilidade entre equipamentos de diferentes fabricantes, já que todos conversarão no mesmo padrão (H.323).

O protocolo H.323 foi desenvolvido pelo ITU-T. Sua primeira versão apareceu em 1996, sendo adotado inicialmente somente por empresas. Em 1998, surgiram as duas versões posteriores, conhecidas por *Tunneling* e *Fast Connect*, que trouxeram maior agilidade e leveza. O

H.323 se contrasta ao MGCP em suas concepções. Enquanto o primeiro tem participação ativa no processamento da chamada, no outro, todo o processamento da chamada é feito no *Media Gateway Control*.

Uma grande vantagem desse padrão é que ele é completamente independente aos aspectos relacionados à rede. Isso quer dizer que há uma livre escolha das tecnologias de enlace que serão utilizadas, com *Ethernet* ou mesmo *Token Ring*. Também não há restrições a topologia da rede, ou seja, pode ser uma ligação ponto a ponto ou terem vários segmentos de redes interconectados.

O protocolo H.323 é mais voltado para o suporte de transporte de voz, ou seja, suas regras foram feitas visando principalmente o áudio. No entanto, para outros casos, como vídeos ou dados, o transporte é feito seguindo as especificações estabelecidas pelo padrão.

A arquitetura H.323 se baseia nas recomendações do ITU-T conforme as especificações:

- H.323 - Especifica a arquitetura e procedimentos gerais.
- H.225 - Especifica a sinalização de chamadas e o protocolo de registro, admissão e status
- H.245 - Especifica o protocolo e os procedimentos para controle de capacidades de terminais e controle de canais lógicos.

Existem também algumas recomendações e padrões que são usados juntamente com o H.323. Alguns são responsáveis por outras especificações como padrões de codificação e compressão de vídeos e compartilhamento de dados.

Na figura 23 são apresentados os principais padrões utilizados para o padrão H.323:

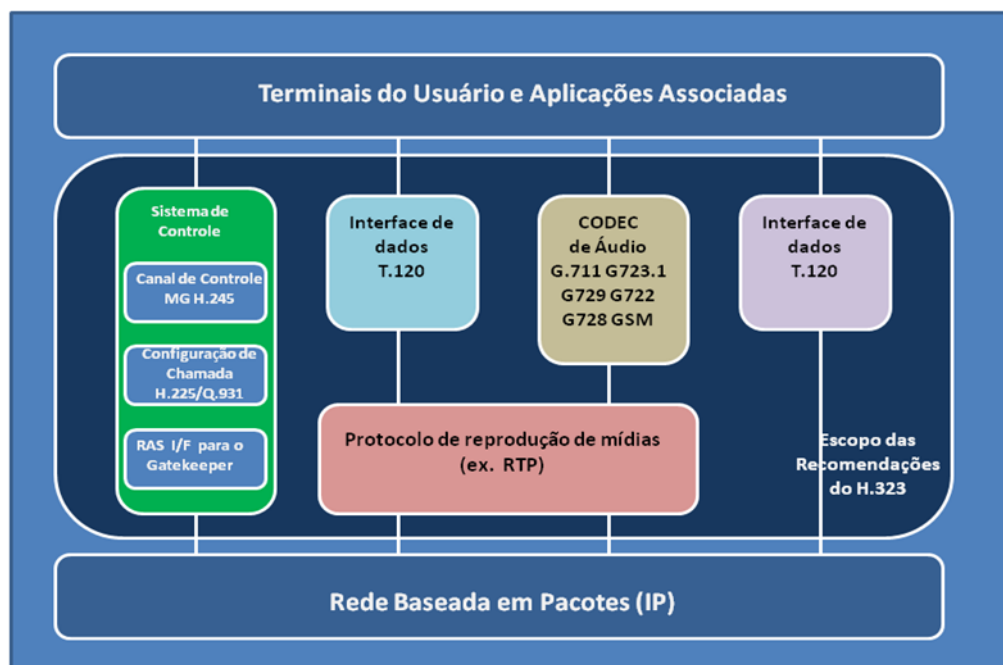


Figura 23 - Componentes do Protocolo H.323

Apesar de ser recente, o cenário do protocolo H.323 já está se tornando obsoleto. Em 1999, o grupo de trabalho *Multiparty Multimedia Session Control* (MMUSIC) do IETF aprovou o padrão SIP (*Session Initiation Protocol*) através do RFC 2543, como alternativa para o padrão H.323. E desde então as redes vem substituindo o H.323 pelo SIP pelas vantagens que este apresenta.

4.5.4. Protocolo SIP [8] [6] [22]

O SIP, do inglês *Session Initiation Protocol*, é um protocolo aprovado pelo IETF em março de 1999, sendo considerado o substituto do protocolo H.323. Foi criado devido à necessidade de incluir e excluir dinamicamente participantes de uma sessão multicast, coisa que o antigo padrão não permitia. O SIP é o principal protocolo do IMS, que é o objeto principal desse estudo.

O H.323 era considerado um protocolo que tinha solução “tudo em um”. No entanto, não se mostrou flexível o bastante para o fornecimento de outros serviços que não fossem a voz. Com isso o desenvolvimento do SIP foi muito bem recebido e foi considerado um grande marco, assim como foi o protocolo HTTP (*Hypertext Transfer Protocol*), que foi responsável pela inovação das páginas da WEB que podiam agora ter links clicáveis conectados com texto, vídeo, áudio e etc. Assim como esse protocolo, o SIP consegue integrar vários conteúdos e sessões de administração. Ele foi rapidamente adotado, principalmente em serviços de comunicação integrada, em especial os que utilizavam presença.

O SIP é responsável por estabelecer, modificar e finalizar sessões multimídias (conferências) e chamadas telefônicas via Protocolo IP. Esse processo se dá devido ao fornecimento dos meios para endereçamento e localização dos membros. A grande vantagem desse padrão é que ele independe do tipo de dado a ser enviado, ou seja, seus procedimentos não dependem do tipo da mídia ou da aplicação da chamada.

Em uma rede de telecomunicações existem dois tipos diferentes de tráfego. O primeiro se refere ao controle e sinalização, que é utilizado para estabelecer gerenciar e terminar as sessões de comunicação. O outro é o tráfego de dados (podendo ser voz, por exemplo). No caso das redes SIP, os caminhos dos tráfegos de sinalização e de dados são independentes. Essa separação é responsável por um gerenciamento mais eficiente, e também facilita a adaptação a possíveis mudanças funcionais que possam ocorrer. Resumindo: o SIP é flexível e aberto para permitir novos serviços desenvolvidos por qualquer desenvolvedor. Por essa característica, ele tem maior vantagem sobre outros protocolos.

A sinalização SIP suporta os seguintes perfis de gerenciamento da sessão multimídia [22]:

- *User Location* (Localização do Usuário) – permite o acesso telefônico ou a outras aplicações características de localizações remotas
- *User Availability* (Disponibilidade do Usuário) – analisa a disponibilidade do usuário entrar em uma sessão de comunicação.
- *User Capabilities* (Capacidade do Usuário) - Determina a mídia e os parâmetros da mídia que serão usados na sessão
- *Session Setup* (Configuração da Sessão)– Estabelece os parâmetros para sessões ponto-a-ponto e multicast.
- *Session Management* (Gerenciamento da Sessão)– permite a transferência e a finalização da sessão; a modificação dos parâmetros da sessão; a e chamada de nova sessão de serviços

Isso quer dizer que o SIP só é responsável pela localização, disponibilidade e capacidade do usuário além da configuração e gerenciamento da sessão. Ele não é capaz de determinar quais serviços estão sendo enviados e recebidos, assim como não afeta como é passada a informação. Isso quer dizer que não importa o modo como é feita a comunicação, seja ela por rádio, cabos ou satélites.

4.5.4.1. Comparações entre o SIP e HTTP [10] [22]

Como visto na seção 4.3.4, O SIP funciona de modo similar ao SMTP (*Simple Mail Transfer Protocol*) e ao HTTP. Enquanto o HTTP fornece a integração de conteúdo (como

texto, áudio, vídeo e links para outras páginas), o SIP consegue integrar diferentes mídias dentro de uma mesma sessão. Ele é considerado um protocolo cliente/servidor ou requisição/resposta, similar ao HTTP quanto à sintaxe e semântica. Em outras palavras, são feitas requisições pela entidade cliente e então enviadas ao servidor. O Servidor processa a mensagem e então envia uma resposta para o cliente. Ele também leva os controles de aplicação para terminal, eliminando a necessidade de centrais de comutação. Ele também é responsável pelo envio de mensagens e códigos, como é o caso do “*error code - 404: Address not found*”. No entanto, diferente do HTTP, o SIP é um protocolo *peer-to-peer*. No HTTP, o servidor web não origina as requisições, enquanto no SIP qualquer agente usuário pode enviar pedidos para iniciar ou modificar a sessão. Além disso, com o SIP é possível gerar respostas de múltipla escolha (com a possibilidade de vários destinos) a partir de uma única requisição [10].

Outra diferença marcante nas arquiteturas é que os serviços HTTP se hospedam em servidores HTTP, responsável por gerar a resposta aos pedidos feitos, enquanto no SIP seus servidores (Servidor Proxy, Servidor de Registro e Servidor de Redirecionamento) conseguem fornecer serviços de ‘roteamento inteligente’.

4.5.4.2. Principais características do protocolo SIP

A seguir serão apresentadas algumas das principais características que o protocolo SIP apresenta em sua composição e que são necessárias para compreender sua ação dentro da rede.

4.5.4.2.1. SIP Server (Servidores SIP) [6]

Em uma rede SIP, com funções específicas, existem três tipos de servidores:

- Servidor de Registro – responsável por receber as atualizações sobre a localização corrente de cada usuário
- Servidor Proxy – responsável por receber as requisições e enviá-las a outros servidores, conhecidos como servidores *next-hop*.
- Servidor de Redirecionamento – responsável por receber requisições e determinar um servidor *next-hop*. Entretanto, ao contrário dele próprio reenviar à requisição, ele retorna o endereço do servidor *next-hop* para o cliente.

4.5.4.2.2. Identificação SIP [22]

O SIP utiliza o URI (*Uniform Resource Identifier*) para identificar os participantes e pesquisas na sessão de comunicação. O SIP URI é semelhante também ao URL (Uniform

Resource Locator) também utilizado pelo HTTP. Em geral, sua forma é feita seguindo o parâmetro:

- ***sip:user:password@host:port;uri-parameters?headers***

Para cada serviço, existe um usuário sip criado. Seguem a seguir alguns casos [22]:

- a. Chamada a um telephone PSTN (*Public Switched Telephone Network*)

- ***sip:1-222-444-9876@pstnnetwork.com;user=phone;***

Parâmetros SIP podem ser usados para fornecer informações adicionais:

- ***user=phone*** → utilizado para indicar que a chamada é direcionada para um numero telefônico

- b. Chamada para Internet

- ***sip:123-4567@itso.com;user=phone; context=privnet;***

- ***Context=privnet*** → indica uma rede IP privada, como por exemplo, a rede de uma empresa, além de ***user=phone*** indicar que se trata de uma chamada para um número telefônico.

- c. Chamada a um computador pessoal

- ***sip:joasilva@itso.com***

Caso a aplicação demande segurança, adota-se a seguinte configuração:

- ***sips: joasilva@itso.com, semelhante ao HTTPS.***

4.5.4.2.3. Relação entre SIP e outros protocolos

Apesar das vantagens do SIP, é importante ressaltar que ele não age sozinho. Ele é apenas um protocolo de aplicação e necessita de outros protocolos trabalhando em conjunto para garantir suas vantagens. No caso do transporte de voz sobre IP (VoIP), por exemplo, o SDP (*Session Description Protocol*) é quem define o formato para a descrição das características e parâmetros para a sessão multimídia; RTP (*Real-Time Transport Protocol*) define o formato do pacote padronizado para entrega de áudio e vídeo através da Internet; e o Real-time *Transport Control Protocol* (RTCP) fornece controle fora da banda com as informações sobre qualidade do fluxo de dados RTP.

Assim, como apresentado na figura 24, o protocolo SIP trabalha em conjunto com outros protocolos.

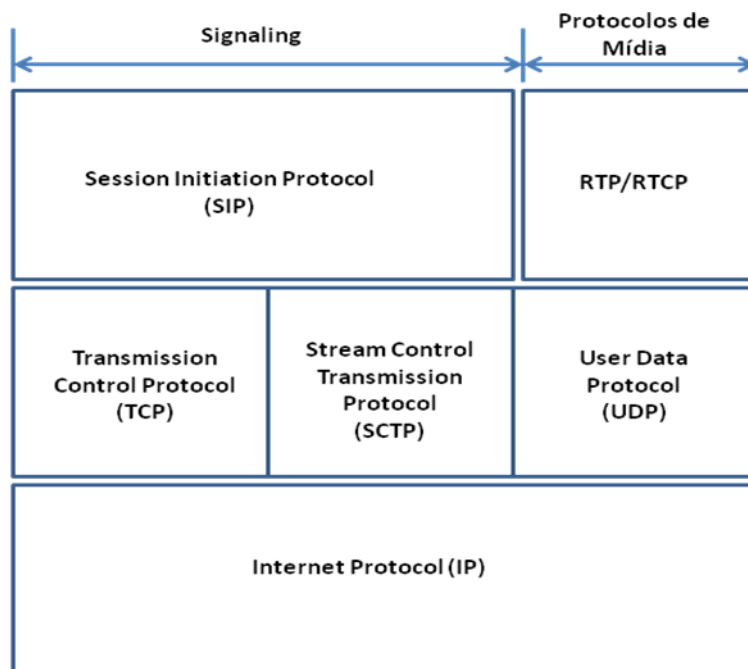


Figura 24 - Protocolos utilizados em conjunto com o SIP [22]

4.5.4.2.4. Mensagens SIP

Na sinalização SIP, todo o estabelecimento, modificação ou encerramento das sessões é feito através da troca de mensagens. No protocolo SIP as mensagens que serão enviadas consistem na linha inicial que especifica o método. Na figura 25 é apresentada a estrutura das mensagens SIP. A linha inicial (*start-line*) pode ser uma linha de requisição ou linha de status. As mensagens de requisição utilizam essa linha para indicar o tipo de solicitação a ser feito. As mensagens de resposta que indicam se a solicitação foi aceita ou não se encontram na linha de status. As mensagens de cabeçalho (*message-headers*) são os campos com nomes/valores.

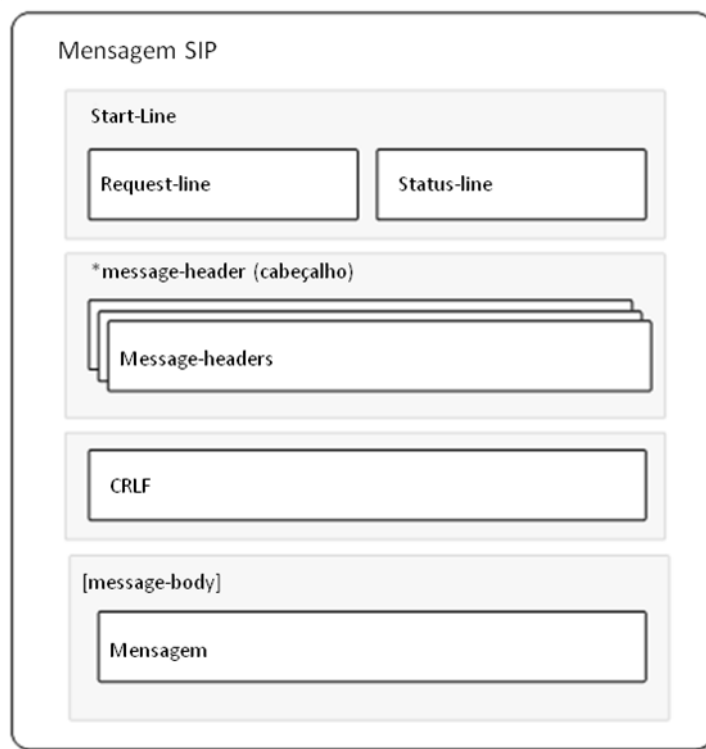


Figura 25 - Estrutura de mensagens SIP [22]

Existem dois tipos de mensagens SIP:

- *Request* (Requisição): que contém os métodos que indicam a ação solicitada
- *Response* (Resposta): que se baseia na estrutura HTTP, podendo haver seis tipos de resposta diferentes.

4.5.4.2.4.1. Mensagens SIP *Request*

Na mensagem SIP *Request* existe a linha de requisição. Seu formato é ilustrado pela figura 26.

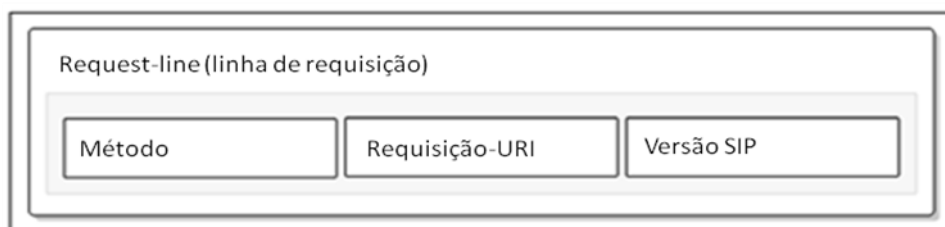


Figura 26 - Mensagem SIP Request [22]

O campo método (*method*) indica qual será o método a ser adotado. Segundo RFC 3261 há seis métodos principais:

- *Invite*: Convida um usuário para participar de uma chamada estabelecendo uma conexão;

- *Bye*: Termina uma conexão;
- *Options*: Informações sobre capacidades suportadas;
- *ACK*: Indica que um “*invite*” foi aceito;
- *Register*: Informa a localização do usuário ao servidor.
- *Options*: Consulta o servidor quanto a sua capacidade

No campo requisição URI (*request-URI*) indica o usuário ou serviço que pedido está sendo requisitado. Nele contem o SIP-URI ou SIPS.

Por fim no campo versão SIP (*version SIP*) indica qual a versão do protocolo SIP está sendo utilizada.

4.5.4.2.4.2. Mensagens SIP Response

O recebimento de uma requisição SIP é feito por servidor *Proxy* que formula uma resposta e o envia de volta com uma mensagem. A mensagem de resposta é colocada no campo linha de status, como apresentado na figura 27:

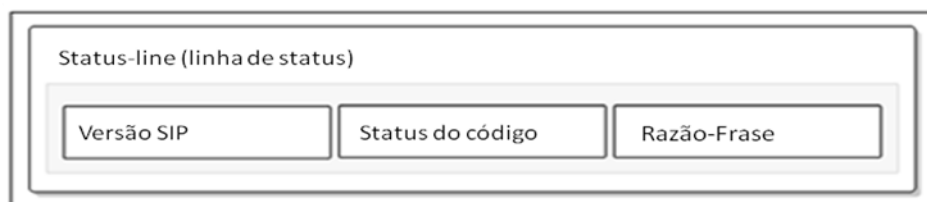


Figura 27 - Mensagem SIP Resposta [22]

No campo versão-SIP, assim como nas mensagens de requisição, indica o a versão do protocolo SIP utilizado.

O campo status do código (*status-code*) é composto de três dígitos que representam o resultado da solicitação. O primeiro dígito indica a classe da resposta. Segundo o desenvolvimento do SIP/2.0 existem 6 classes possíveis, que podem ser:

- 1XX: *Progress*;
- 2XX: *Successful Request*;
- 3XX: *Redirection*;
- 4XX: *Incorrect Request*;
- 5XX: *Server Failure*;
- 6XX: *Global Failure*.

Por fim o campo razão-frase (*reason-phrase*) trata de uma breve descrição textual do código de status. Sempre que o status é destinado ao processamento da máquina, esse campo é preenchido para mostrar o que é mostrado ao *user agent*.

Assim, devido ao modelo e características que o SIP apresenta, rapidamente ele se tornou a melhor opção para controle de sinalização. Tendo apenas cinco funções e utilizando uma abordagem por camadas, tem flexibilidade e abertura que permite que ele seja utilizado tanto por novas como antigas aplicações. Rapidamente o SIP ganhou espaço na arquitetura das comunicações multimídia no mundo IP. Por todas essas razões, também se tornou o principal protocolo utilizado pelo IMS, que será visto no próximo capítulo.

5. IP MULTIMEDIA SUBSYSTEM – IMS

O IMS, do inglês *IP Multimedia Subsystem*, é uma arquitetura de referência que visa a entrega de serviços de comunicação de voz, vídeo e dados através de uma rede IP. Ela é descrita como subsistema (*subsystem*), uma vez que faz parte de uma rede completa, necessitando de outros componentes para funcionar como um sistema multimídia. O IMS é considerado uma arquitetura de referência porque seus princípios agregam práticas bem sucedidas de outras redes. Além disso, existem contribuições que ainda estão sendo agregadas a fim de resolver problemas que não foram bem definidos nas primeiras definições. O IMS não é um novo conceito de redes para a transmissão de serviços e sim, uma particularidade, uma solução desenvolvida a fim de melhorar a qualidade dos serviços e gerar receita às operadoras.

A seguir são apresentadas as principais características e motivos para que o IMS seja apontado como alternativa viável para o futuro das redes de serviços.

5.1. Justificativas para o IMS

Como visto no primeiro e segundo capítulo, nos últimos anos a Internet passou por crescimento exponencial. Esse crescimento se deu principalmente à sua capacidade de fornecer serviços e entretenimento aos usuários. Serviços como email e WWW foram possíveis graças à utilização de protocolos abertos que estão disponíveis para que qualquer desenvolvedor possa criar seu próprio aplicativo. Para isso basta conhecer os protocolos utilizados [23].

As comunicações móveis, principalmente o celular, também se popularizaram muito nos últimos anos. Além de chamadas telefônicas, existem muitos outros serviços como o SMS, caixa postal e, com o advento das NGN (redes convergentes), o celular também foi capaz de oferecer serviços de dados. O que possibilitou esse processo foi o GPRS (*General Packet Radio System* ou Serviço de Rádio de Pacote Geral), uma tecnologia que permite o aumento das taxas de transferência de dados nas redes GSM, sendo o transporte de dados feito a partir da comutação de pacotes [23]. Essa nova tecnologia permitiu às operadoras oferecerem a Internet móvel em uma boa velocidade (taxa de transmissão teórica de 2,4Mbps) e com custo razoável, uma vez que a cobrança é feita em função da quantidade de pacotes de dados que são transmitidos, não pela duração da conexão à rede. Assim, além da internet móvel, outros serviços surgiram com o advento do GPRS, como o acesso WAP (*Wireless Application Protocol* ou Protocolo para Aplicações sem Fio) e acesso às redes corporativas VPN.

No entanto, apesar de se mostrarem eficazes, esses serviços não foram capazes de atrair consumidores suficientes para justificar os altos investimentos nessas redes de pacotes para celular. A solução então encontrada foi a convergência entre as redes de telefonia móvel com a Internet. Essa alternativa ficou conhecida como 3G, ou terceira geração. Esta tem maior capacidade da rede, devido a uma melhor eficiência espectral. Por isso é possível a transmissão de dados além da telefonia de voz a longas distâncias, a altas velocidades. Isso possibilitou o oferecimento de novos serviços.

O IMS surgiu para complementar esse cenário. Ele é uma promessa que visa tornar o paradigma da internet móvel possível e de maior eficácia para o usuário. O conceito de IMS representa o elemento chave na arquitetura 3G para tornar possível o acesso universal a todos os serviços que a Internet dispõe, via celular. Deste modo, visa facilitar, por exemplo, a navegação em *web-pages*, leitura de e-mails, visualização de filmes e participação de videoconferências usando dispositivos 3G.

As redes 3G já possuem nativamente a comutação de pacotes em sua estrutura, tornando a comunicação de dados mais rápida e eficiente, comparada com dispositivos 2G, que fazem uso da comutação de circuitos. Este é o motivo pelo qual os dispositivos 3G possuem taxas de transmissão maiores e possuem maior largura de banda para acessar a Internet. Desta maneira, os usuários podem navegar na web, ler e-mails, fazer download de arquivos e fazer virtualmente tudo o que seriam capazes através de uma conexão de banda larga de suas casas. Um bom exemplo é o caso de um usuário que pode instalar VoIP no celular e assim realizar uma chamada, não pagando uma ligação, mas sim transferência de dados. No entanto, esse tipo de serviço já está disponível de maneira eficiente a usuários que utilizam essa tecnologia. Assim, pode-se perguntar qual seria o motivo para a concentração de esforços e implantação de IMS. A resposta a se questão se resume a três fatores [23]:

- Qualidade de Serviço (QoS)
- Tarificação
- Integração de serviços variados.

Um dos principais problemas da transmissão por pacotes é garantir o fornecimento de serviços multimídia em tempo real. Não existe garantia de que a rede oferece largura de banda suficiente a um usuário para uma conexão em particular ou prever eventuais atrasos na entrega dos pacotes. Assim, uma das principais metas do IMS é ser o responsável pela sincronização dos enlaces, provendo assim QoS.

Outro ponto chave é o modo como um usuário é tarifado. Em uma conexão 3G ele é tarifado de acordo com a quantidade de dados transmitidos, não importando que sejam estes voz, vídeo ou mensagens instantâneas. Não existe nenhum sistema que faça a distinção entre o conteúdo dos bytes transmitidos. Se as operadoras forem capazes de diferenciar o conteúdo dos bytes, a cobrança poderá ser feita de uma maneira mais benéfica para o usuário. Um bom

exemplo é o caso novamente da chamada VoIP que pode ser cobrada por sua duração e não pela quantidade de bytes envolvidos; ou ainda uma mensagem instantânea que pode ser tarifada por sua unidade, não dependendo assim do seu tamanho. No entanto é importante ressaltar que IMS não define um padrão de como as operadoras irão tarifar tais serviços. Pelo contrário, deixa livre para as operadoras tarifarem da maneira que melhor lhe convirem. O IMS apenas fornece informação sobre os serviços requisitados pelos usuários deixando a tarifação a cargo da operadora.

E por fim, com a padronização das interfaces de serviços, serviços de terceiros podem ser implementados sem maiores dificuldades, podendo ser compatíveis com serviços já existentes ou que venham a surgir. Isso é a integração de serviços variados que podem também ser utilizados pelo usuário tanto em sua rede de origem como quando está em *roaming*.

Assim, para atender essas principais premissas, o IMS faz uso do protocolo e da rede IP. No entanto, ele também é capaz de se comunicar com outras redes, como a TDM. Com IMS é possível estabelecer sessões diferentes ao mesmo tempo, como durante uma videoconferência; saber o status de disponibilidade de outros usuários; ler e-mails ou enviar arquivos. Como todos os serviços que o IMS provê utilizam pacotes, a comunicação é mais eficiente. Também o ambiente criado por esse novo padrão faz com que qualquer serviço possa acessar qualquer aspecto da sessão estabelecida. Isto permite que aplicações mais ricas e interessantes possam ser desenvolvidas, ao contrário do uso do ambiente que cada serviço é diferente e eles não se interconectam. E como IMS não depende do mundo TDM, a interconexão com outros dispositivos, como laptops conectados à internet usando algum software de videoconferência, se tornará algo trivial.

Portanto, o grande trunfo do IMS é prover serviços de Internet (incluindo serviços multimídia em tempo real) com QoS e por um preço aceitável e justo.

5.2. Premissas IMS [23] [25] [26] [28] [31]

O IMS foi arquitetado para entregar serviços multimídia IP de valor agregado para os usuários finais. Para que sua implementação seja eficiente, há alguns pontos-chaves que esse padrão deve obrigatoriamente suportar:

- Estabelecimento de sessões multimídia IP;
- Um mecanismo para negociar QoS;
- Interconexão com a Internet e a PSTN;
- *Roaming*;
- Controle rigoroso dos serviços entregues ao usuário final;
- Criação de serviços de forma mais ágil, não dependendo de padronizações;

- Acesso universal via diferentes tipos de redes de acesso (WLAN, ADSL, *Cable Modem*, 3G, CDMA, PSTN, etc.), podendo dar suporte a transmissão de várias tecnologias distintas.

No caso do estabelecimento de sessões multimídias IP, o IMS é necessário para a comunicação de áudio e vídeo, uma vez que este é um serviço de grande importância e popularidade entre os usuários.

Um mecanismo para negociação de QoS é fundamental uma vez que este é um componente chave no IMS. Portanto, com o IMS é permitido às operadoras um controle fino do QoS que um usuário recebe, de modo que a operadora possa diferenciar certos grupos de usuários de outros.

O IMS também deve ter suporte à conexão com a Internet e com isso o número de origens e destinos para sessões multimídia é potencialmente expandido. E também, como esperado, suporte com a PSTN e redes celulares também são suportados.

O suporte a *roaming* se torna inerente uma vez que as redes celulares 2G já o suportam.

Deve haver também um controle rigoroso sobre os serviços entregues aos usuários, devido às políticas impostas. Elas podem ser de dois tipos:

- Políticas gerais: aplicáveis a todos os usuários de uma rede. Por exemplo, a operadora que deseja restringir o uso de codecs que consomem muita largura de banda, como o G.711. Ao invés, elas querem promover o uso de codecs que usam menos banda, como o AMR (*Adaptive Multi-Rate*).
- Políticas individuais: um conjunto de políticas que são personalizadas para cada usuário. Um usuário pode ter acesso a serviços IMS, menos aos que utilizem vídeos, por exemplo. Em outras palavras, este tipo de política depende dos termos contratuais de cada usuário.

Os serviços no IMS não precisam de padronização, algo que no passado causou grandes problemas. Cada serviço tinha uma padronização ou era proprietário, e muitas vezes não eram compatíveis entre si ou simplesmente não funcionavam quando em *roaming*. Diante desse contexto, o IMS busca reduzir o tempo necessário para se introduzir um novo serviço, ganhando tempo evitando a padronização do serviço e o teste de interoperabilidade. O IMS é responsável por padronizar as capacidades dos serviços ao invés dos serviços em si.

O IMS, com sua premissa de ter acesso universal, deve permitir outros acessos além do GPRS. Como o IMS se insere apenas em uma rede IP, independe das camadas mais baixas e consequentemente do tipo de acesso.

A Figura 28 apresenta as principais redes que o IMS irá agregar em sua aplicação. Também apresenta as principais características de cada uma que ira herdar em sua composição. Na rede celular, por exemplo, irá agregar sua mobilidade e o *roaming*:

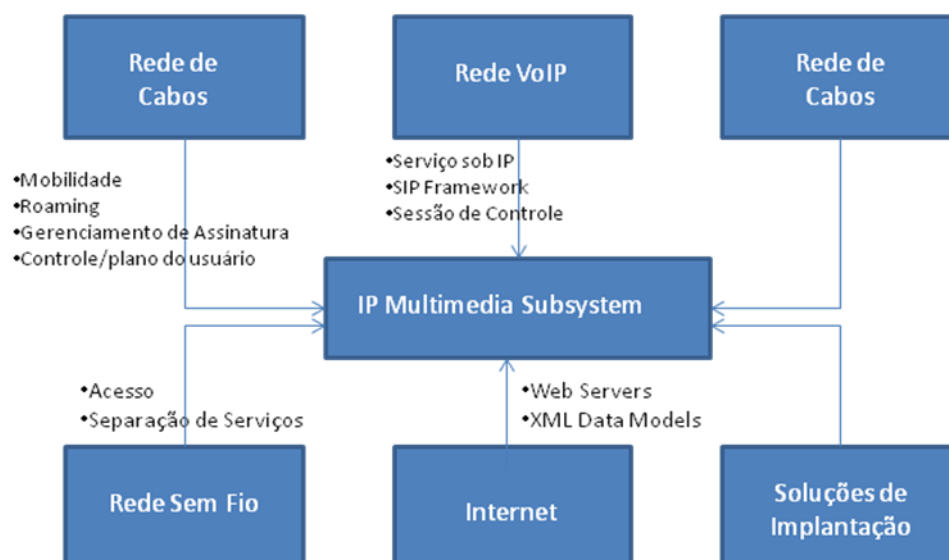


Figura 28 - Redes conectadas com IMS

Na Figura 29 são apresentadas as várias tecnologias que podem convergir no IMS: CDMA, redes IP, redes de cabos, PSTN, WiFi e WiMAX.

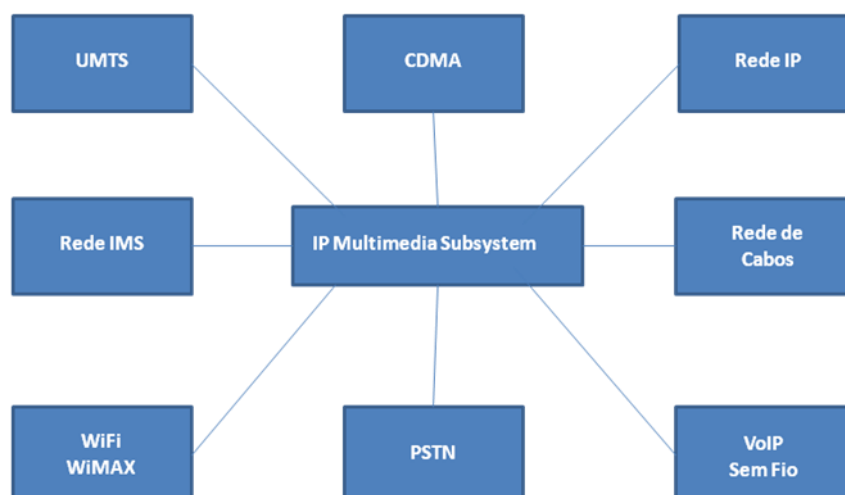


Figura 29 - Tecnologias convergentes no IMS

Nas redes NGN o tipo de serviço oferecido é o triple play, que combina voz, vídeo e dados em um mesmo canal. O Sistema *Quad-play* agrega além das outras três características, a capacidade de suportar redes sem fio (*wireless*). O IMS é considerado uma plataforma única, que será capaz de oferecer serviços completos, inclusive redes sem fios com a vantagem da mobilidade e da possibilidade da combinação dos serviços. Isso quer dizer que o IMS visa o oferecimento de serviços *Quad-play* e assim poderá permitir também uma nova geração de serviços com mais conteúdo, como presença e entretenimento. Essa nova arquitetura permite que serviços cruzem por múltiplas redes de acesso, provendo assim um melhor controle do QoS do início ao fim, não necessitando que cada rede distinta provenha seu próprio método para

garantir QoS. Assim, pode-se dizer que o IMS é capaz de abrir novas formas para estruturar o negócio, revolucionando o modelo de negócio das operadoras, tanto para atacado quanto para varejo. [25]. Esse modelo é apresentado na figura 30:

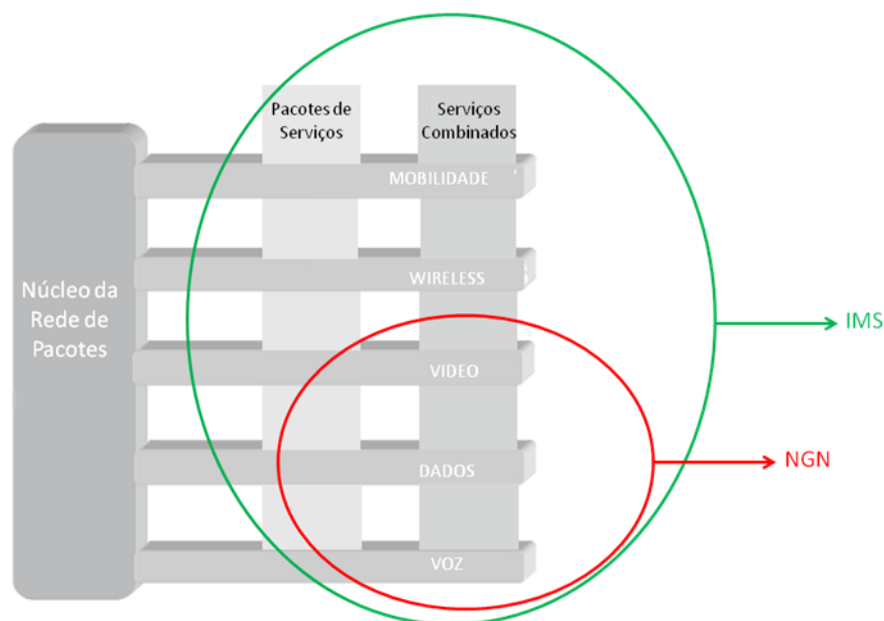


Figura 30 - Sistema NGN e IMS [27]

5.3. Exemplos de Aplicação

Com a implantação do IMS, muitas são as aplicações possíveis que irão melhorar seu QoS, além novos serviços que podem ser desenvolvidos. A seguir, estão listados algumas aplicações suportadas pelo IMS [27]:

- Suporte à TV → Exemplo: mensagens na TV ou ligações via TV;
- TV interativa → Exemplo: assistir um programa na TV, com interatividade; convidar outros usuários para assistir o mesmo programa (via chat), fazer conferência (voz/vídeo), *instant messaging*, etc.;
- Transferência de conteúdo entre dispositivos → Exemplo: assistir vídeo no celular (comprar, baixar) e transferir para a TV o que foi comprado ao chegar em casa;
- Compra de conteúdos (com pedido de autorização)
- Presença → Exemplo: utilização de um livro de endereços ativos (com status e localização, *buddies* e grupos (em que apenas os amigos podem ver o status, com prévia autorização do usuário), voz e vídeo-chamadas (todas as opções disponíveis para conexão podem ser mostradas));
- Videoconferência;

- Receber ligações no desktop → Exemplo: recebimento de ligações no desktop e depois transferi-la para o celular;
- Regras para encaminhamentos → Exemplo: criar regras para encaminhamento de mensagens quando em reunião, almoço, praticando esportes, etc.)
- Push-to-talk over celular (PoC) → sistema de celular via rádio

Na figura 31 é apresentado um exemplo de aplicação no IMS. Trata-se de uma representação de uma lista de contatos com várias aplicações em que é possível se comunicar com eles. Nesse caso, pode-se escolher, conforme o usuário qual serviço o originador pode utilizar, podendo ser mensagens instantâneas, email, vídeo ou mesmo uma chamada VoIP. Todos esses serviços podem estar disponíveis para usuários em uma rede IMS, conforme sua escolha de aplicativos [23].

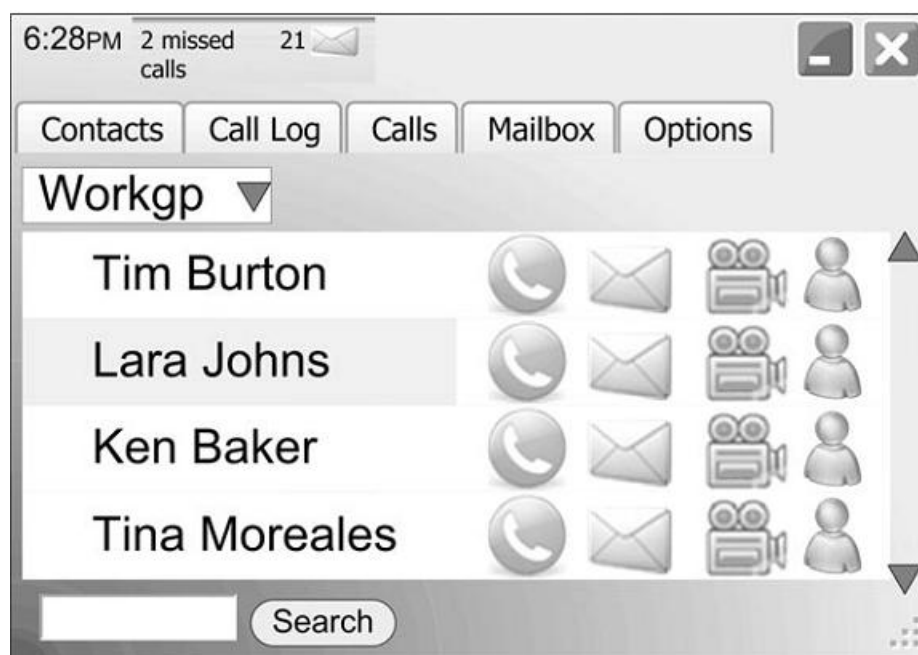


Figura 31 - Lista de contatos de um usuário com IMS [27]

Já na figura 32 é apresentado o IMS nas aplicações IPTV. Na figura, observa-se que o usuário pode assistir um programa comprado pela televisão e assisti-lo pelo celular ou mesmo receber chamadas telefônicas em seu televisor.



Figura 32 - Exemplo Aplicação IMS [27]

5.4. Padronização IMS

Como visto anteriormente, a tendência de convergência das redes de telecomunicações buscando sempre maior qualidade de serviço em uma rede IP foi fundamental para o desenvolvimento do IMS.

5.4.1. Órgãos responsáveis

Entre as principais organizações mundiais para padronização do IMS estão:

- IETF: organização que une projetistas, operadoras, fornecedores e institutos de pesquisa que trabalham em conjunto para desenvolver a arquitetura, os protocolos e a operação da internet pública. É aberto a qualquer indivíduo interessado em colaborar. Desenvolveu a maioria dos protocolos usados hoje na internet. É a “fábrica” de protocolos utilizados com o IP.
- 3GPP :Nasceu em 1998 como um acordo de colaboração entre vários grupos de padronização regionais, conhecido como *organizational partners*. Sua tarefa original era desenvolver especificações e relatórios técnicos que pudessem ser aplicados em todo o globo pra o sistema 3G GSM.
- 3GPP2 :Nasceu com o objetivo de envolver as redes celulares da América do Norte e da Ásia, baseadas nos padrões de acesso via rádio ANSI/TIA/EIA-41 e CDMA2000, num sistema de terceira geração. Assim como o 3GPP, no 3GPP2 os seus membros são também conhecidos como *organizational partners*.

- Colaboração IETF-3GPP/3GPP2: Ao invés de criarem novos protocolos do zero, 3GPP/3GPP2 estabeleceram uma colaboração com o IETF de modo a certificarem que os protocolos desenvolvidos por este atendessem as suas necessidades, uma vez que o IMS usaria protocolos IP.

5.4.2. Evolução da Padronização do IMS [20] [28]

Para a padronização das redes 3G, o ITU-T adotou um padrão chamado IMT-2000. Esse foi resultado da colaboração entre vários grupos padronizadores que buscavam prover o acesso aos serviços de telecomunicações via rádio, incluindo satélites e redes terrestres. O início da padronização do IMS se deu a partir de dois grupos que compõem o IMT-2000: o 3GPP (*Third Generation Partnership Project*) e o 3GPP2 (*Third Generation Partnership Project 2*). Esses ainda são dois dos principais colaboradores na padronização do IMS, apesar de ambos trabalharem em diferentes projetos de IMS, são bastante similares, como por exemplo a utilização da rede IP (padronizado pela IETF). No entanto existem também outros órgãos padronizadores que auxiliam no processo de criar um padrão único.

Em 1999, foi definida a primeira versão do padrão IMS pelo 3GPP, com o objetivo principal de prover sistemas de comunicação móvel via rede IP. Ela apareceu como parte da padronização do sistema 3G dentro das redes UMTS. Nele também se definiu o protocolo SIP como principal protocolo de sinalização. Também seria capaz de suportar o GSM e prover GPRS.

O 3GPP2 adicionou ao IMS o suporte a tecnologia CDMA. Em seguida, o 3GPP foi responsável pela adição da interconexão com WLAN, interoperabilidade entre o IMS e diferentes redes com conectividade IP, grupos de identificação de rotas, serviços de presença, reconhecimento de voz e serviços de voz (PoC).

Em seguida surgiram esforços para conectar as redes fixas e NGN ao IMS. Nesse ponto, outros órgãos ajudaram a criar padrões a serem utilizados. Em 2004 o ITU-T criou o *NGN Focus Group* (NGN-FG), cujo objetivo é desenvolver a rede NGN para acesso fixo ao IMS. Paralelamente, em 2004, o ETSI, na Europa, criou o TISPAN (*Telecoms & Internet Converged Services & Protocols for Advanced Network*) cujo objetivo é padronizar a rede NGN de acesso fixo baseada no IMS. Na América do Norte, também em 2004, o ATIS criou o *NGN Focus Group* para estudar a aplicabilidade da NGN e do IMS para as redes fixas da América do Norte. Os três grupos trabalharam em conjunto para manter sincronizadas suas definições. Adicionalmente, estes grupos trouxeram novos requisitos ao 3GPP/3GPP2 para o suporte de acesso fixo de banda larga ao IMS (Release 7).

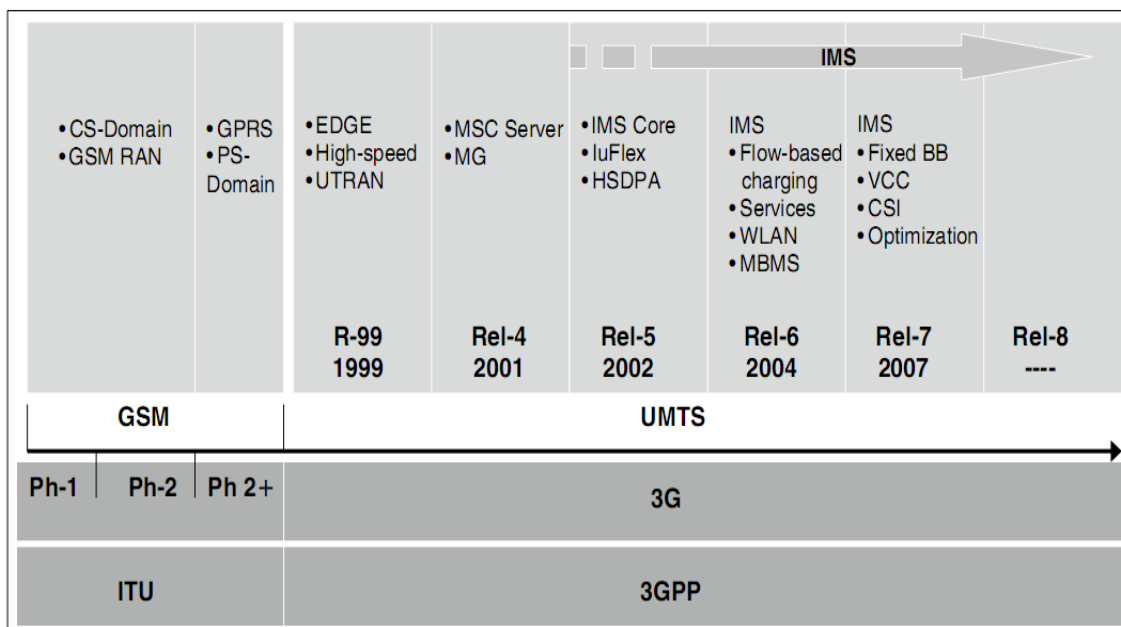


Figura 33 - Release apresentados ao longo dos anos para o IMS

A figura 33 apresenta a evolução da padronização conforme as novas atualizações iam surgindo. No entanto, a figura já está desatualizada. Já foram lançadas as três novas versões de atualização: Release 8, 9 e 10 que são responsáveis por novas adições de suporte como:

- Suporte à tecnologia *Long Term Evolution* (LTE também conhecida como 4G) e *System Architecture Evolution* (SAE), além de centralização de serviços IMS.
- Suporte de chamadas de emergência sobre GPRS e EPS, além de reforço a serviços centralizados e continuados.
- Suporte a melhorias para *Single Radio Voice Call Continuity* (SRVCC) e sessões de emergência IMS. Esse foi o Release-10, que foi lançado recentemente.

E ainda sim continuam fazendo melhorias para a arquitetura IMS a fim de torná-lo um padrão altamente confiável e assim fazê-lo ganhar mercado.

5.4.3. Visão geral da padronização do IMS.

Ao contrário de outros padrões, por exemplo, o GSM que teve maior parte dos seus protocolos criados especificamente para o seu uso, o IMS se baseia em protocolos IP que já estão desenvolvidos. Assim, é possível destacar os principais que foram escolhidos para a aplicação do IMS.

Para o controle das sessões, havia três candidatos: BICC (evolução do ISUP e faz a separação entres os planos de sinalização e de enlace, além de suportar IP, SS7 ou ATM); o H.323 (desenvolvido do zero para suportar IP); e SIP (protocolo para o estabelecimento e gerenciamento de sessões multimídia sobre IP, e segue a filosofia do modelo cliente-servidor).

O que se mostrou mais eficaz foi o SIP, sendo este o escolhido. O seu funcionamento é descrito no Capítulo 4.

Para a tarefa de AAA (*authentication, authorization and accounting*) escolheu-se o Diameter, que é uma evolução do RADIUS, um protocolo muito utilizado para AAA da Internet. O Diameter consiste de um protocolo base que é complementado por Aplicações Diameter. Tais aplicações personalizações ou extensões deste protocolo para se encaixarem numa aplicação em particular em um dado ambiente. O Diameter é usado em diversas interfaces no IMS, apesar de nem todas as interfaces usarem as mesmas aplicações. Por exemplo, existe uma aplicação Diameter para interagir com SIP durante o setup das sessões e outra que contabiliza o controle de créditos do usuário.

Outros protocolos que são encontrados no IMS, como o COPS (*Common Open Policy Service*), usado para especificar cliente/servidor para apoio na política de controle sobre QoS e protocolos de sinalização; H.248, usado nos nós de sinalização para controlar nós no plano de enlace; e RTP e RTCP destinados ao transporte de multimídia em tempo real.

Na figura 34, apresenta-se uma comparação da composição dos protocolos utilizados pelo IMS e o modelo OSI, em que camada eles se inserem. Percebe-se que os protocolos utilizados pelo SIP correspondem às três últimas camadas do modelo OSI: sessão, apresentação e aplicação.

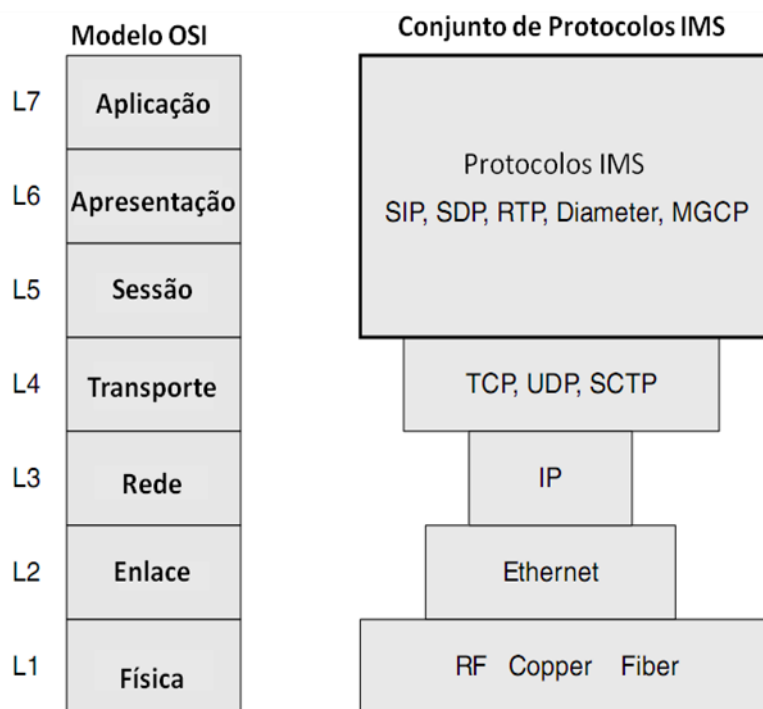


Figura 34 - Comparação - Modelo OSI e Protocolos IMS [27]

5.5. Estrutura / Arquitetura IMS

A arquitetura do IMS é projetada para oferecer novos serviços com eficiência. Assim ele é capaz de combinar qualidade e interoperabilidade das redes de telecomunicações com as vantagens da rede IP. A arquitetura é uma coleção de funções interligadas por interfaces padronizadas, uma vez que o 3GPP não padroniza os nós e sim suas funções. Logo, podem existir nós que implementam mais de uma função como também funções que podem ser implementadas por mais de um nó.

Para entender a separação de uma rede que contém a arquitetura IMS, toma-se como base a figura 35. Primeiramente, existem os equipamentos que os usuários finais utilizam: telefones celulares, computadores ou telefones fixos representados na camada mais baixa da figura. Para a troca de informação entre esses equipamentos, utiliza-se uma rede de acesso básica que os conecta. Esta pode ser uma rede de telefonia fixa, uma rede 3G ou até mesmo uma rede WiMAX. Todas essas redes se conectam em uma rede convergente, que no caso do IMS é a IP. Nesse ponto, todo o controle de sessão e controle de serviço será feito segundo o que a arquitetura IMS apresenta. Essa parte é responsável por acessar os serviços e aplicações que os usuários desejam utilizar. A seguir será estudado um pouco dessa arquitetura e seus componentes para esclarecer como funciona todo o processo do IMS.

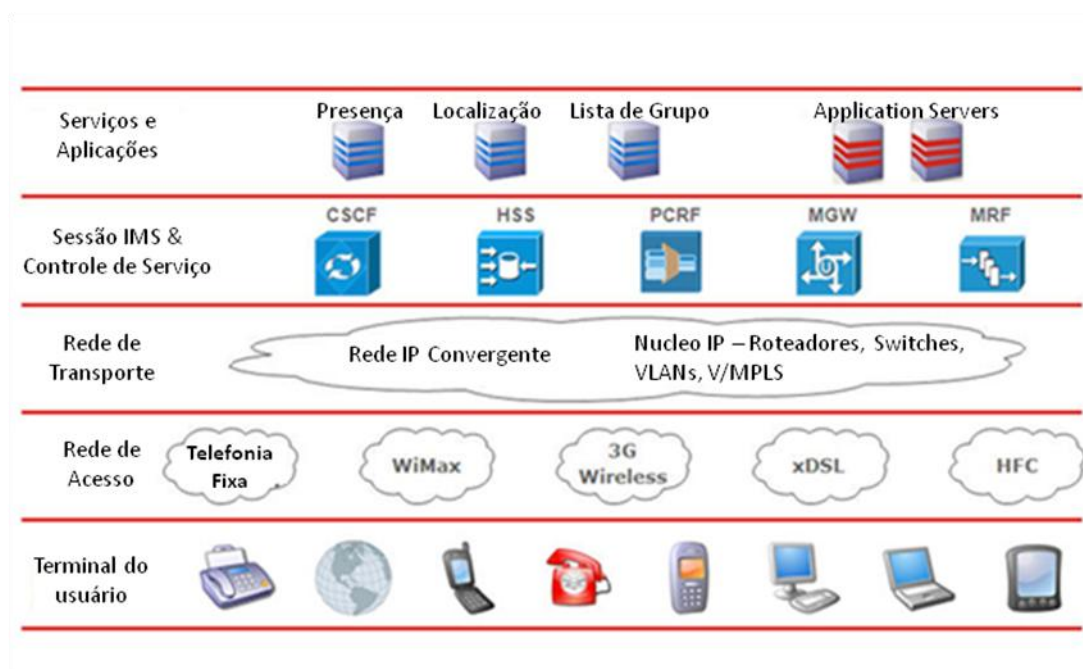


Figura 35 - Separação em camadas dos componentes em uma rede IMS [24]

5.5.1. Estrutura

A arquitetura IMS pode ser dividida em três planos (ou camadas), de forma semelhante ao modelo apresentado pelo NGN (Figura 36). A primeira camada é a camada do transporte, em que se localizam todos os equipamentos dos usuários ligados a rede, a própria rede acesso e rede IP, com todos seus respectivos equipamentos responsáveis pelo encaminhamento das mensagens.

A segunda camada é conhecida como a camada de controle. Ela faz todo o controle das sessões, assim como toda a parte de operação, gerenciamento, aprovisionamento e cobrança. Ela também é responsável por controlar o acesso a terceira e última camada, a de aplicação e serviços.

Na camada de aplicação estão todos os aplicativos e serviços que as operadoras da rede oferecem aos seus usuários.

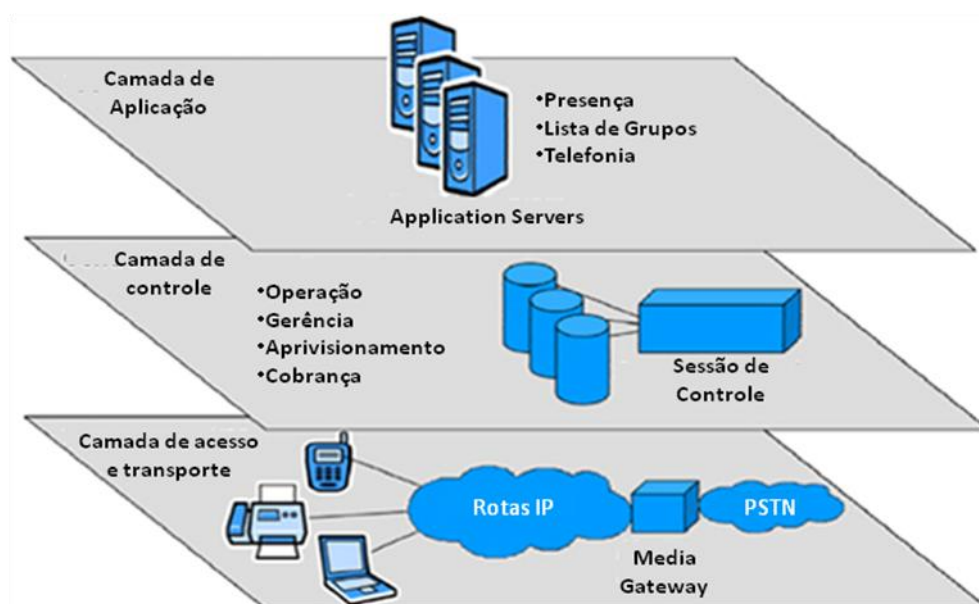


Figura 36 - Modelo em Camadas do IMS

5.5.2. Interfaces

Para a arquitetura IMS foram criadas diversas interfaces para a comunicação. Cada uma delas tem um papel específico para conversar entre si a partir dos protocolos definidos para a arquitetura. A tabela 1 apresenta algumas das principais interfaces utilizadas, assim como a comunicação que ela faz, seu protocolo e sua principal função.

Tabela 1 - Interfaces do IMS [27]

INTERFACE	ENTRE	PROTOCOLO	FUNÇÃO
Rx	P-CSCF ↔ GGSN	Diameter	A interface Rx é uma interface para aplicação em políticas e regras de tarifação para fluxo de IP em GGSN. A interface Ty corresponde a interface 3GPP2
Mp	MRFC ↔ MRFP	Megaco/H.248	A Mp é uma interface do H.248 para controlar recursos de mídia em um MRFP através de MRFC
Mn	MGCF ↔ IMS-MGW	Megaco/H.248	A Mn é uma interface do H.248 para controle de recursos de mídia em um IMS-MGW através de MGCF
Mb	MRFP ↔ UE IMS-MGW ↔ UE	RTP	A Mb é uma interface IP entre a MRFP ou a IMS-MGW e a UE, levando o pacote IP correspondente ao fluxo de mídia
Ut	UE ↔ AS	XCAP	A Ut é uma interface OMA definida para clientes de aplicação UE para gerenciar dados relacionados a serviços com os servidores de aplicação
Sr	MRFC ↔ AS	XCAP	A interface Sr é utilizada pelo MRFC para obter informação sobre a mídia tais como <i>scripts</i> do AS

Já na figura 37 apresenta-se onde se inserem essas interfaces, além de algumas outras não demonstradas na Tabela 1.

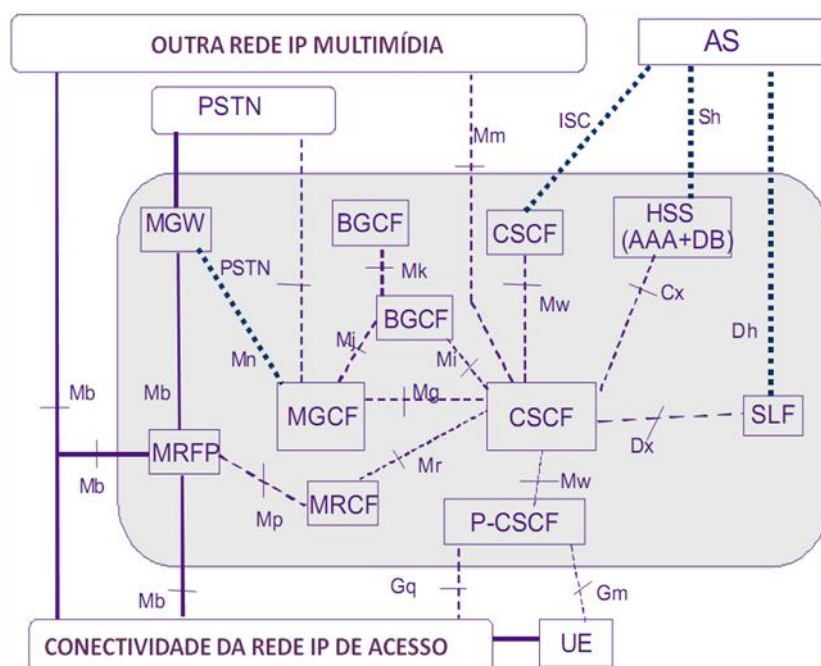


Figura 37 - Localização das Interfaces na rede IMS [30]

Para compreender onde essas interfaces se encaixam, a seguir serão descritos os principais componentes da arquitetura IMS.

5.5.3. O núcleo IMS [23] [25] [26]

Toda a inteligência da arquitetura IMS é conhecida como Núcleo.

No caso do IMS, o núcleo da rede é composto pelas seguintes entidades:

- banco de dados de usuários (HSSs e SLFs);
- *SIP servers*, conhecidos coletivamente como Call Session Control Function (CSCFs);
- *Application Servers* (ASs);
- Media Resource Functions (MRFs), divididos em MRFC (controladores) e MRFP (processadores);
- *Breakout Gateway Control Functions* (BGCFs);
- PSTN gateways, divididos em *Signaling Gateway* (SGW), *Media Gateway Control function* (MGCF) e *Media Gateway* (MGW).

A seguir serão descritas cada uma dessas entidades.

5.5.3.1. Bases de Dados: HSS e SLF

O *Home Subscriber Server* (HSS) é o responsável pelo armazenamento central de informações referentes ao usuário. Tecnicamente, HSS é uma evolução do HLR (*Home Location Register*), das redes de telefonia móvel. Nesse banco estão armazenados todos os dados da conta do cliente necessários para gerenciar as sessões multimídia. Entre as principais informações contidas estão: os serviços que o usuário está autorizado a acessar; as diversas identidades; as redes permitidas para *roaming*; e localização do usuário.

Caso exista um número muito grande de usuários, pode existir mais de um HSS. Nesses casos, existe também um elemento chamado *Subscription Locator Function* (SLF), cuja função é o mapeamento dos usuários com seus respectivos HSS. Um consulta ao SLF recebe o endereço de usuário como entrada e retorna em qual HSS estão as informações daquele usuário como saída.

O HSS é responsável por várias tarefas: identidade do usuário, informações de registro, parâmetros de acesso, informações de segurança e localização, informações do perfil do usuário (inclui serviços em que o usuário está inscrito) e informações relativas a qual S-CSCF está alocado para servir o usuário.

Tanto HSS como SLF usam Diameter com uma aplicação IMS específica. Uma das funções críticas do HSS é prover chaves de criptografia e autenticação para cada assinante. Um das grandes vantagens do HSS é a centralização dos dados. Além disso, o HSS é capaz de gerenciar múltiplas identidades para um mesmo assinante.

A figura 38 representa as funções principais do HSS. A função identidade é responsável pelo endereço do usuário e informação da numeração do mesmo [27].

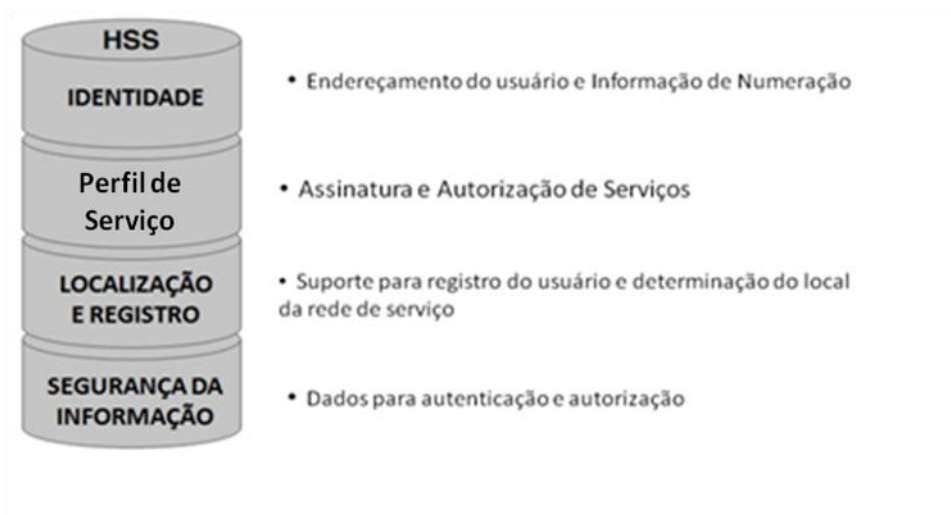


Figura 38 - Pilha HSS [27]

A função perfil de serviços trata da autorização e do cadastro que os usuários tem acesso.

A função registro e localização são responsáveis por dar suporte aos registros dos usuários e determina sua localização do seu servidor de rede.

A última função do HSS é a segurança da informação em que ele é responsável pela autenticação e autorização dos dados a serem acessados.

5.5.3.2. Controle das sessões – *Call/Session Control Function* – CSCF

O controle das sessões no IMS é feito através dos CSCF – *Call/Session Control Function*. Eles são os elementos de sinalização na arquitetura. Esses CSCF são servidores SIP que processam os pacotes de sinalização SIP, controlando as sessões, serviços e alocando recursos, ou seja, controlam o estabelecimento, o monitoramento e a finalização das sessões. Eles são, em conjunto, responsáveis pelo controle central do core IMS, provendo o *set up*, o estabelecimento, modificação e o término de sessões multimídia.

- Podem ser de três tipos, com funções distintas:

- *Proxy-CSCF* (P-CSCF);
- *Interrogating-CSCF* (I-CSCF);
- *Serving-CSCF* (S-CSCF).

5.5.3.3. *Proxy-CSCF* (P-CSCF)

O Proxy- CSCF é o ponto de entrada (na borda) do domínio IMS (do ponto de vista de sinalização). Ele funciona como um servidor Proxy para um TU (Terminal do Usuário) que tem como objetivo requisitar serviços no IMS. Esse se conecta ao P-CSCF antes de fazer os registros na rede ou iniciar uma sessão SIP, podendo estar localizado na rede de origem (home network) ou na rede visitada (*roaming*).

Do ponto de vista SIP, o P-CSCF é um SIP *proxy server* de entrada e saída, ou seja, todo os pedidos iniciados pelo terminal ou destinados a ele atravessam este nó. O processo começa com o encaminhamento pelo P-CSCF das requisições e respostas SIP na direção correta. Um P-CSCF é alocado ao terminal IMS durante seu registro na rede e não muda enquanto durar o acesso. Algumas funções são relacionadas à segurança, então primeiramente, é estabelecido um número de associações IPsec, um protocolo de segurança, no sentido do terminal, oferecendo proteções de integridade dos dados.

Uma vez que o usuário foi autenticado, o P-CSCF garante ao restante dos nós na rede a identidade do usuário. Desta maneira, não há a preocupação dos outros nós em autenticar o usuário, porque todos eles confiam no P-CSCF.

Adicionalmente, o P-CSCF também verifica a correção das requisições SIP para evitar mensagens SIP fora do padrão. Como SIP é um protocolo baseado em texto, suas mensagens podem ser grandes. Por conta disso, o P-CSCF também inclui um compressor e descompressor dessas mensagens.

Opcionalmente, pode existir um *Policy Decision Function* (PDF), que pode ou não estar integrado com o P-CSCF ou implementado com um nó separado. Esse PDF é responsável por autorizar os recursos de enlace e gerenciar QoS.

Paralelamente, o P-CSCF também gera CDRs (*Clock and Data Recovery*) de todas as sessões que passam por ele. Ele também inclui cabeçalhos nas mensagens de requisição e resposta, antes de enviá-las ao próximo CSCF. Tais cabeçalhos não são compartilhados com outras redes, evitando com que outras operadoras aprendam sobre a rede da concorrente e seus serviços.

O P-CSCF também pode ser usado para separar acessos de qualquer dispositivo. Entretanto, o P-CSCF não mantém a função de autenticador, trabalho do S-CSCF que será visto a seguir. Ele é crítico para prevenir acessos não autorizados à rede e pode estar localizado tanto na rede visitada como na rede de origem. No caso da rede ser baseada em GPRS, o P-CSCF

estará sempre na mesma rede que o GGSN (*Gateway GPRS Support Node*). Assim, resumidamente é possível selecionar algumas das principais funções do P-CSCF:

- Encaminhamento de requisições/respostas na direção correta;
- Alocação a um TU toda vez que houver o registro na rede, e não trocar enquanto durar o acesso;
- Possuir atribuições de segurança: IPsec no sentido do TU (oferece integridade de dados);
- Após ser autenticado, garantir a TODOS os nós da rede a identidade do usuário (os outros nós não precisam se preocupar em autenticar um usuário porque todos os nós confiam no P-CSCF);
- Verificar se as requisições SIP estão corretas, isto é, dentro do padrão;
- Fazer a compressão/descompressão de mensagens SIP.

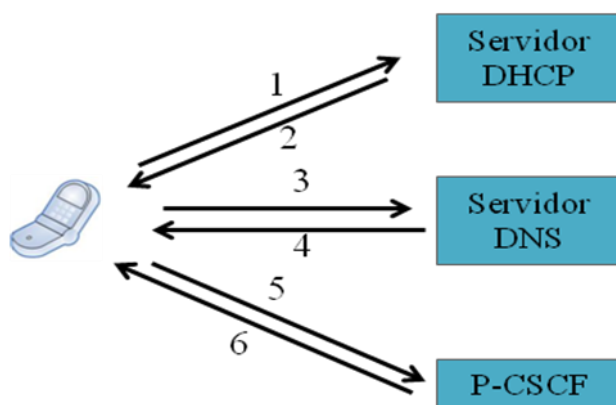


Figura 39 - Acesso ao Servido P-CSCF

Na figura 39 é apresentado o procedimento de localização do P-CSCF. Ele pode ser resumido em 6 etapas, descritas na figura, que são apresentadas a seguir:

- 1- consulta o DHCP (*Dynamic Host Configuration Protocol*)
- 2- recebe um endereço IP, IP do DNS e o P-CSCF FQDN (*Fully Qualified Domain Name*).
- 3- consulta o DNS (*Domain Name System*) sobre o FQDN do P-CSCF
- 4- recebe o endereço IP do P-CSCF.
- 5- registro do UE no P-CSCF, via SIP.
- 6- Autorização ou não do pedido.



Figura 40 - Resposta ao Pedido de Acesso

A figura 40 apresenta as etapas 5 e 6 do processo anterior. Essa parte necessita de um destaque especial uma vez que é neste momento em que ocorre o IPsec security associations (SAs) para 4 portas entre o UE e o P-CSCF, garantindo que toda a sinalização subsequente passe pelas portas protegidas. O Protocolo de Segurança IP, também conhecido como IPsec é um método padrão que fornece privacidade de usuário, o que aumenta a confiabilidade das informações inseridas pelo usuário para alguma localidade na Internet, garantindo que o conteúdo chegue ao seu destino e a autenticidade das informações.

Assim, o P-CSCF é considerado a porta de entrada na rede IMS, pois é ele que permite ou não o acesso.

5.5.3.4. Interrogating – CSCF (I-CSCF)

O *Interrogating- CSCF* ou apenas I-CSCF funciona como um servidor *proxy* SIP no domínio IMS e se localiza sempre na borda de um domínio administrativo. Durante o registro IMS, o I-CSCF consulta o HSS para selecionar o *Serving-CSCF* que irá servir o TU.

Esse servidor age como o ponto de entrada para terminar as requisições SIP, roteando os pedidos de sessões para o *Serving-CSCF*. É semelhante a um *gateway* (para outras redes IMS) da rede, determinando se um acesso é permitido ou não, e assim encaminhando as mensagens SIP para outras operadoras. Por este motivo, pode ser usado para esconder detalhes da rede (*THIG function – Topology Hiding Inter-Network Gateway*), criptografando as mensagens SIP que contém informações sensíveis da rede (como número de servidores, nomes DNS ou suas capacidades).

Os endereços do I-CSCF estão listados nos servidores de DNS do domínio. Quando uma mensagem SIP segue os procedimentos para encontrar o próximo nó, o servidor obtém o endereço do I-CSCF. Este determina se um acesso é permitido ou não de outras redes e assim encaminha mensagens SIP para a operadora. Por este motivo, o I-CSCF pode ser usado para esconder de outras operadoras, detalhes da rede, criptografando partes das mensagens que podem contém informações importantes sobre a rede. Além disso, protege o S-CSCF e o HSS

de acessos não autorizados. Para fins de escalabilidade e redundância, a rede pode conter mais de um I-CSCF. Eles normalmente se encontram na rede de origem, mas em casos especiais podem estar na rede visitada (*roaming*).

É o I-CSCF que determina qual S-CSCF será usado de acordo com as capacidades do momento ou as políticas de cada serviço. Duas estratégias podem ser usadas para a escolha do S-CSCF: a primeira diz respeito à atribuição de um S-CSCF para cada tipo de serviço e a outra o S-CSCF é atribuído geograficamente. Essas informações estão armazenadas no HSS para referência.

Quando um usuário é registrado na rede, um S-CSCF é atribuído por um I-CSCF, que mantém tal informação (juntamente com informações de roteamento) enquanto durar o tempo do registro. Enquanto o usuário permanecer na mesma área de serviço tal registro é mantido. Por exemplo, numa rede celular o registro é mantido enquanto o usuário permanecer na mesma célula (ou seja, não mudou de endereço) ou numa rede fixa, e o registro mantém-se ativo enquanto o endereço IP for o mesmo. No entanto, existe um tempo limite de vida para cada registro, que quando atingido, serão expirados tanto no S-CSCF quanto no HSS.

O I-CSCF pode ser considerado como o *firewall* para a rede IMS, roteando entre outras operadoras e prevenindo outras redes de acessarem sua rede.

Assim, enquanto o P-CSCF é considerado como um *gateway* para redes não-IMS, o I-CSCF trabalha como um *gateway* para as redes com arquitetura IMS. O P-CSCF funciona como o *gateway* para redes não-IMS e *access point* para a rede IMS enquanto que o I-CSCF funciona como um *gateway* entre redes IMS.

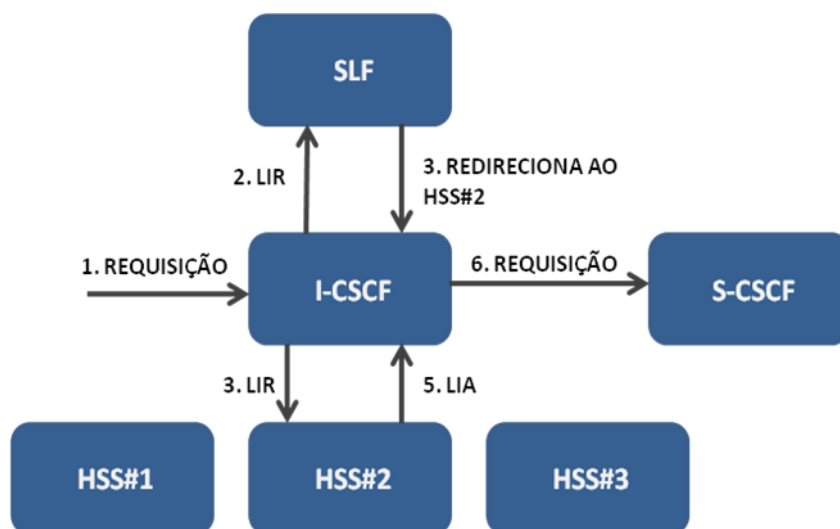


Figura 41 - Informações trocadas pelo I-CSCF

A Figura 41 representa o procedimento de consulta pelo I-CSCF. Esse recebe o convite e manda um requisição de informação de localização (*Location-Information-Request -LIR*) para descobrir em qual HSS se encontra o usuário (2). O SLF redireciona o I-CSCF para o HSS desejado (3), que e novamente questiona o banco de dados escolhido (4). Este retorna uma mensagem LIA (*Location-Information-Answer*) indicando qual o S-CSCF deve ser utilizado (5). Assim, o I-CSCF manda um convite para o S-CSCF controlar a sessão (6).

5.5.3.5. *Serving* - CSCF (S-CSCF) [26]

O Serving-CSCF ou S-CSCF (Figura 41) também é um servidor SIP que realiza o controle de sessões. É considerado o nó central no plano de sinalização e funciona também como um SIP registrar, mantendo uma interação entre a localização do usuário e o seu endereço SIP (Identidade Pública do Usuário). Ele é responsável por controlar todos os aspectos dos serviços dos usuários, mantendo um status de cada sessão que o usuário inicia, além do status do registro do usuário para outras aplicações e mantém o controle sobre estes serviços enquanto durar o registro na rede. Portanto, quando o status de um usuário muda dentro da rede, todas as entidades participantes da sessão envolvida são notificadas. Por isso, assim como o I-CSCF, o S-CSCF mantém uma interface Diameter com o HSS pelos seguintes motivos:

- Baixar vetores de autenticação do usuário que tenta acessar a rede;
- Baixar o perfil do usuário, que inclui o perfil de serviços, que pode fazer com que a mensagem SIP seja encaminhada através de um ou mais *Application Servers*;
- Informar o HSS que é o S-CSCF alocado ao usuário para aquela sessão.

Outro ponto importante é que todas as sinalizações SIP que o terminal IMS envia ou recebe passam pelo S-CSCF alocado, que funciona como roteador. Além disso, o S-CSCF também é responsável por aplicar as políticas contratuais, como no caso de um usuário que não está autorizado a estabelecer um determinado tipo de sessão. O S-CSCF também é responsável por manter um controle total das sessões e ser capaz também de originar ou terminar sessões em nome de outra entidade da rede. Ele também coordena com o *Media Resource Function* (MRF) o envio de anúncios/tom para quem está originando uma sessão. Assim como todas as entidades CSCF, ele é capaz de gerar informações de bilhetagem.

Assim, pode-se dizer resumidamente que o S-CSCF salva sobre um dispositivo registrado na rede:

- Endereço do HSS;
- Perfil do usuário;
- Endereço do P-CSCF de entrada na rede;
- Domínio do P-CSCF, no caso do dispositivo ter se conectado através de outra rede;

- Identidade Pública do Usuário;
- Identidade Particular do Usuário;
- Endereço IP do dispositivo (TU).

O S-CSCF está sempre localizado em sua rede de origem (*home network*).

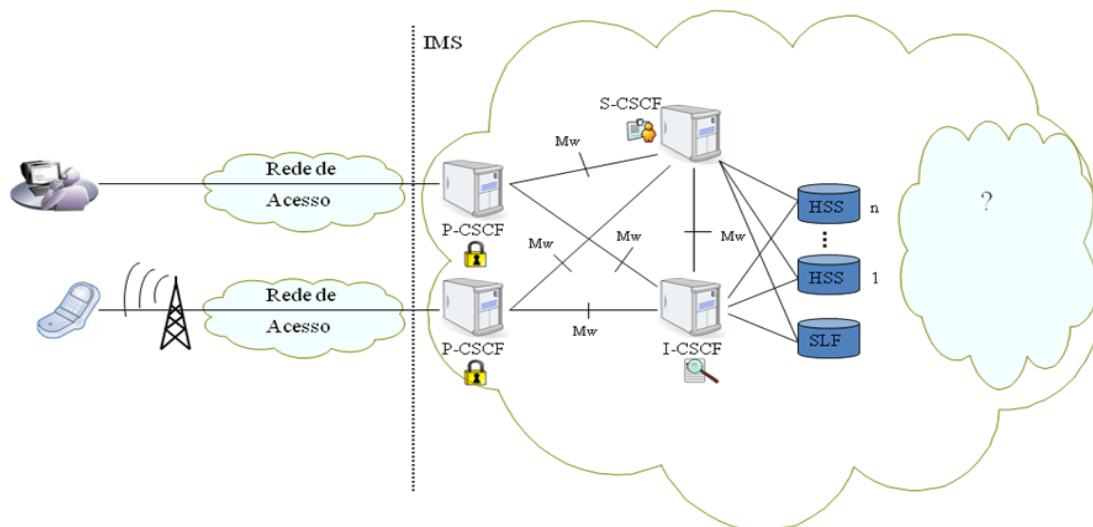


Figura 42 - Controle feito pelos CSCFs

A figura 42 apresenta o controle de uma sessão feita pelos elementos P-CSCF, I-CSCF, S-CSCF, HSS e SLF. Nela são apresentados todos os passos até obter o servidor do usuário, e a figura 43 apresenta a sessão de controle IMS completa.

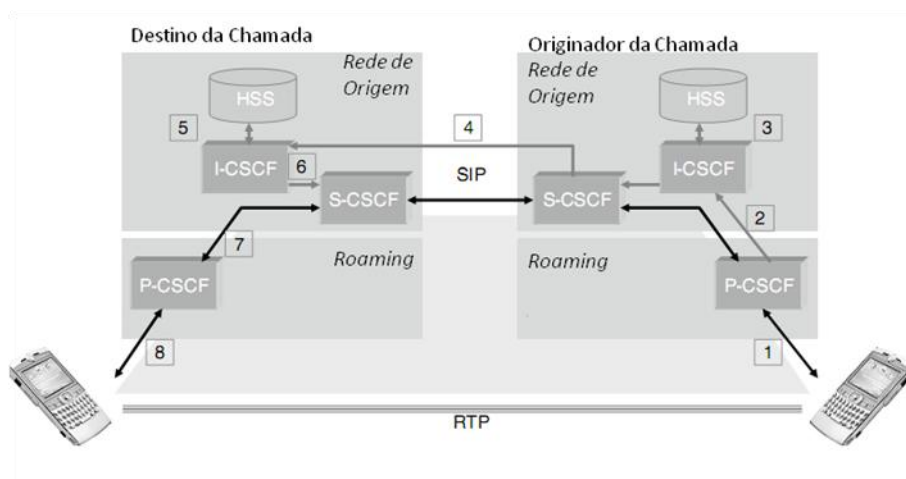


Figura 43 - Sessão Completa de Controle IMS [27]

Como exemplo, numa simples chamada de voz, como representada na Figura 43, ocorrem os seguintes passos:

- 1 – o TU solicita uma sessão SIP através P-CSCF

- 2 – o P-CSCF, através o I-CSCF faz uma requisição de consulta a base de dados do usuário (HSS) em sua rede local
- 3 – após consulta, o HSS irá autenticar TU no P-CSCF, liberando ou não o acesso.
- 4 – a solicitação SIP é enviada ao S-CSCF que irá buscar o terminal destino na outra rede IMS.
- 5- o I-CSCF da rede destino faz a localização do terminal desejado através de uma consulta ao HSS.
- 6 - o HSS devolve a localização para o S-CSCF da rede destino
- 7 - o S-CSCF da rede destino envia solicitação de acesso ao P-CSCF que controla o terminal procurado.
- 8 - o P-CSCF completa a sessão SIP com o Terminal Destino, estabelecendo a sessão via RTP.

5.5.3.6. Plano de Aplicação - *Application Servers* – AS [23] [25] [31]

O *Application Server* (AS) é a entidade SIP que contém e executa os serviços (não faz parte do core da rede). Os ASs podem operar em um dos seguintes modos, dependendo do serviço:

- SIP *proxy mode*;
- SIP *redirect mode*;
- SIP UA (*user agent*) mode (capaz de criar um diálogo SIP);
- SIP B2BUA (*Back-to-back User Agent*) mode (concatenação de dois user agents); usado para esconder informações de ambos UA, como uma sessão sem identificação.

O AS se comunica com o S-CSCF via protocolo SIP. O S-CSCF é o responsável por acionar o AS para providenciar o serviço requerido pelo usuário. É importante ressaltar que os ASs não são entidades IMS puras, e sim funções no topo do IMS. No entanto, são sempre descritas como parte das funções IMS por fornecerem serviços multimídia de valor agregado. Eles podem estar na rede de origem ou mesmo em redes de terceiros (prestadores de serviços). Como visto no item 5.5.1, a arquitetura de serviços no IMS separa-se entre os elementos de controle e lógica dos serviços, de modo que novos serviços podem ser criados sem a necessidade de upgrade dos elementos de controle das sessões. Entre as principais funções do AS, pode-se citar:

- Possibilidade de processar e impactar uma sessão SIP recebida do IMS;
- Capacidade de originar requisições SIP;
- Habilidade para enviar informações de bilhetagem.

Existem também três tipos de AS. Cada um deles tem uma função específica, podendo ser [23]:

- SIP AS (*Application Server*): é o tipo “nativo” de AS, que contém e executa os serviços IMS via protocolo SIP. Espera-se que os serviços futuros sejam todos baseados neste tipo.
- OSA-SCS (*Open Service Access-Service Capability Server*): é um AS que contém uma interface para um Servidor OSA/parlay (API – *Application programming Interface* – para a rede de telefonia). Do lado IMS esse nó age como um AS (interagindo com o S-CSCF via SIP) e do outro lado, como uma interface entre o OSA AS e o OSA API.
- IM-SSF (*IMS Service Switching Function*): emula uma função SSF para o acesso de um *Service Control Point* (SCP) da Rede Inteligente (RI). Deve ser capaz de prover translação entre SIP e protocolos da RI, como CAMEL (*Customized Application for Mobile Network Enhanced Logic*). O IM-SSF permite que um gsmSCF (*GSM Service Control Function*) para controle de uma sessão IMS. Do lado IMS, é um AS trabalhando com o S-CSCF via SIP.

Todos os três tipos se comportam como um AS SIP na direção da rede IMS. Eles podem conter ou não uma interface para o HSS. O SIP-AS e o OSA-SCS na direção do HSS usam Diameter e são usados para baixar ou carregar dados relacionados a usuários armazenados no HSS. Já o IM-SSF usa MAP na interface com o HSS. Se o AS estiver na rede de terceiros, não possui interface com o HSS.

A figura 44 representa a arquitetura IMS com os ASs:

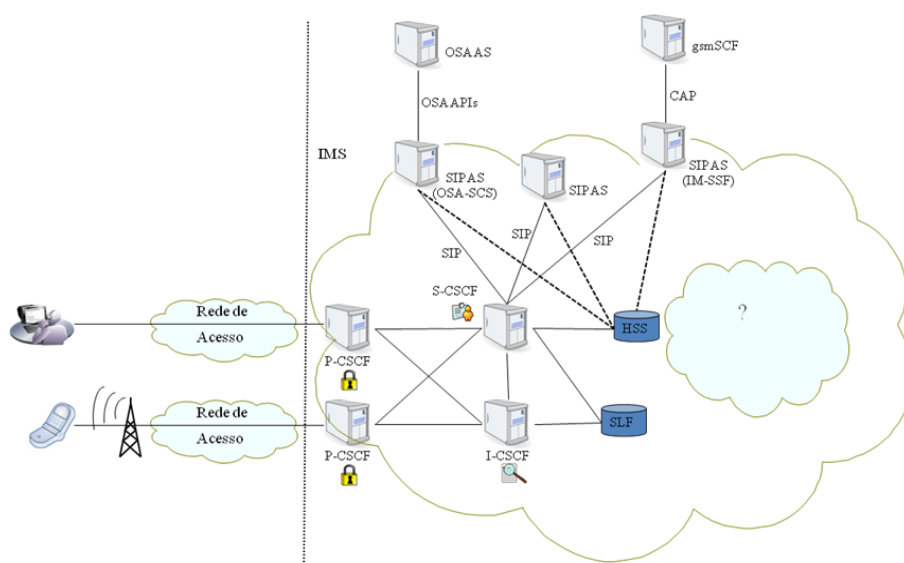


Figura 44 - Localização do AS na rede IMS

A figura 45 apresenta o modelo de uma consulta feita pelo S-CSCF no AS a fim de prover um serviço. Utilizando a interface ISC, o Proxy é capaz de receber uma requisição SIP e comunicar-se com o AS a partir de uma interface SIP e obter o serviço desejado.

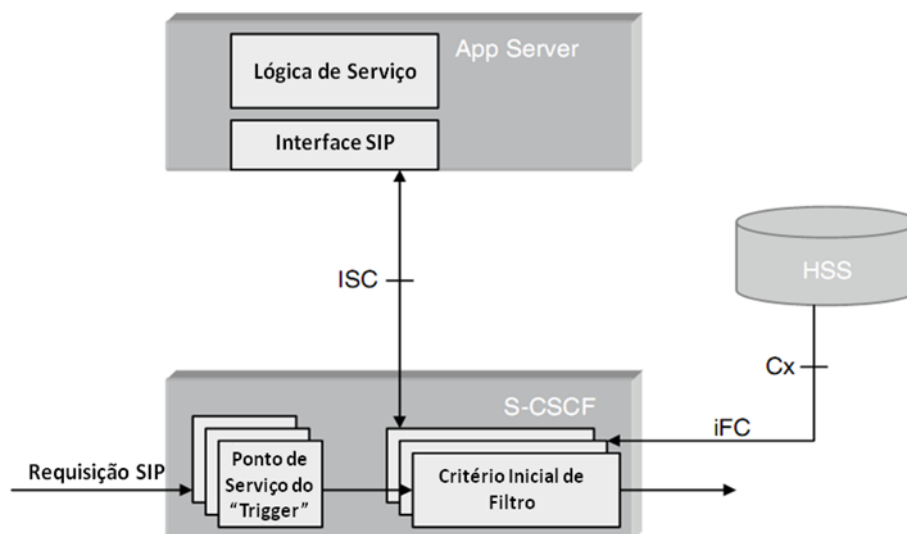


Figura 45 - Pedido de serviço ao AS pelo S-CSCF [27]

5.5.3.7. Media Resource Function (MRF) [23]

É o provedor de mídia na rede interna, com capacidade de tocar anúncios, misturar *streams* de mídia (no caso de sessões em grupo/conferências), fazer a transcodificação entre codecs diferentes, obter estatísticas e fazer qualquer outro tipo de análise de mídia. É dividido em:

- MRFC (*Media Resource Function Controller*), no plano de sinalização;
- MRFP (*Media Resource Function Processor/Plataform*), no plano de rede.

O MRFC é responsável pelo controle dos recursos de mídia no MRFP. Ele interpreta as informações vindas do AS via S-CSCF (via SIP) e assim pode controlar o MRFP. Esse controle se dá via protocolo H.248.

Já o MRFP implementa todas as funções relacionadas com mídia. Ele gerencia vários tipos diferentes, sempre sobre o controle do MRFC. Ele pode, por exemplo, misturar os fluxos de mídia, podendo fazer uma conferência; originar fluxos de mídia (anúncios) processar esses fluxos de mídia (transcodificação de áudio, análises); etc.

O MRFP também é capaz de prover tons e de suportar DTMF (*Dual-Tone Multi-Frequency*). Sempre com a ocorrência de um evento, ele notifica o MRFC sobre o fato. Assim, pode-se destacar alguns exemplos de mídias controladas pelo conjunto MRF:

- Geração/detecção de tom;

- Conferência;
- *Text-to-speech*;
- *Automated speech recognition* (ASR);
- Facsimile;
- Anúncios

Os MRFs devem também estar sempre na rede local da operadora (*home network*).

A Figura 46 apresenta os principais componentes do MRF, enquanto a Figura 47 revela em que posição na rede IP os componentes MRF se localizam.



Figura 46 - Componentes do MRF [27]

5.5.4. Versões diferentes de IP [23] [31]

No caso das redes IP serem de versões diferentes, ou seja, IPv4 e IPv6, o IMS também desenvolveu entidades que possam fazer essa conversão. Os principais responsáveis por esse trabalho são IMS-ALG (*Application Layer Gateway*) e o TrGW (*Transition Gateway*).

Ambos são as entidades responsáveis pela interconexão entre redes IPv4 e IPv6. O IMS-ALG funciona como um SIP B2BUA, mantendo duas pernas de sinalização independentes: uma na direção do terminal IMS e outra na direção da outra rede, onde cada perna fala uma versão diferente do protocolo IP. Ele também trabalha com o I-CSCF para o tráfego que chega e com o S-CSCF para o tráfego de saída.

Já o TrGW é efetivamente um NAT-PT/NATPT-PT (*Network Address Port Translator-Protocol Translator*). Ele também faz a tradução (no nível de transporte) entre endereços IPv4/IPv6 e portas, além de fazer também a translação entre protocolos. O TrGW é configurado com uma gama de endereços IPv4 que são dinamicamente alocados para uma dada sessão.

Assim, o *backbone* completo da arquitetura IMS é representado na figura 49:

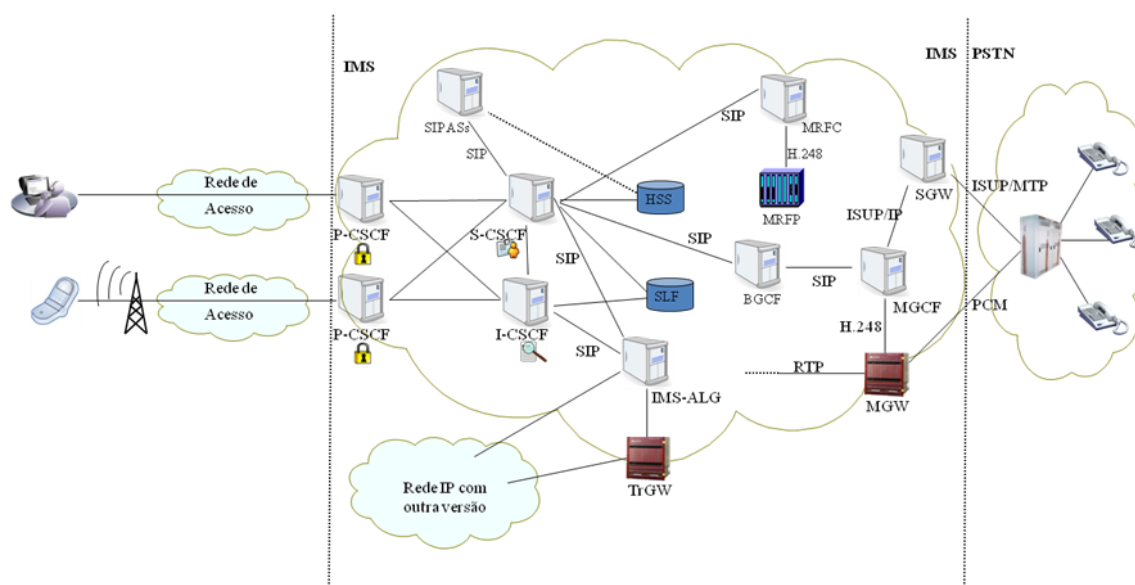


Figura 49 - Conversão do IMS com diferentes versões do IP

De forma geral, a constituição da arquitetura IMS é descrita por todas essas entidades. No entanto, para entender como se dá o funcionamento da arquitetura, necessita-se compreender como os usuários são identificados. Assim, nos próximos itens serão apresentados os perfis e modos de cobrança diferenciados que o IMS apresenta.

5.6. Identidade do usuário no IMS

Como já visto anteriormente, o IMS se baseia no protocolo SIP para controle de suas sessões. Assim, as identidades dos usuários também utilizam essa técnica, podendo assim assumir duas formas diferentes:

- Identidade Pública
- Identidade Privada

5.6.1. A identidade pública [23]

A cada usuário IMS é alocado uma ou mais identidades públicas, na forma de SIP URI ou TEL URI, e são usadas como as informações de contato nos cartões de visita. As mensagens de sinalização SIP no IMS são baseadas nas identidades públicas. Em comparação com o GSM, é o mesmo que o MSISDN (*Mobile Subscriber ISDN Number*). O usuário então pode assumir essas duas identidades públicas [23]:

- SIP URI (*Uniform Resource Indicator*)
 - sip: *nome.sobrenome@operadora.com* ou
 - Sip: +1-212-555-0293@operadora.com; *user=phone*
- TEL URI (necessário para ligações IMS<->PSTN)
 - Tel: +55-19-3213-1990

Pela lógica, as operadoras devem implementar pelo menos um SIP URI e um TEL URI por usuário. Mais de uma identidade pública pode ser necessária, como no caso de identidades pessoais e comerciais, sendo então mais que necessário essa diferenciação. No IMS é possível registrar várias identidades públicas em apenas uma mensagem, economizando assim tempo e largura de banda.

5.6.2. Identidade Privada

As identidades privadas não estão no formato de SIP URIs ou TEL URIs. Ao invés, tomam o formato NAI (*Network Access Identifier*): *usuário@operadora.com*. Estas identidades não são usadas para o roteamento de requisições SIP. Elas são utilizadas exclusivamente para fins de identificação da assinatura e autenticação. Comparando com o GSM, é equivalente ao IMSI (*International Mobile Subscriber identifier*). Este tipo de identidade não precisa necessariamente ser conhecida pelo usuário, uma vez que se encontra gravada em um smart card, da mesma maneira que o IMSI é armazenado em um SIM (*Subscriber Identity Module*)

card. A figura 50 apresenta o modelo de identificação do IMS, tanto sua identidade pública quanto privada.

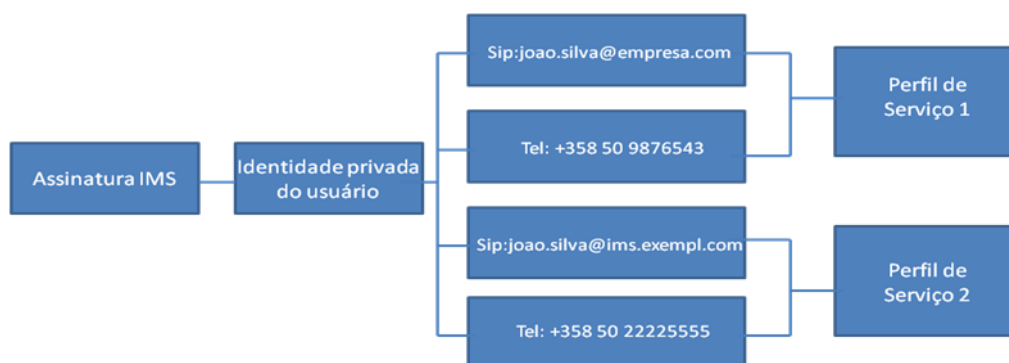


Figura 50 - Identidades Públicas e Privadas do Usuário [27]

Já a figura 51 apresenta a evolução do *smart card* ao longo das tecnologias, até chegar ao IMS SIM.

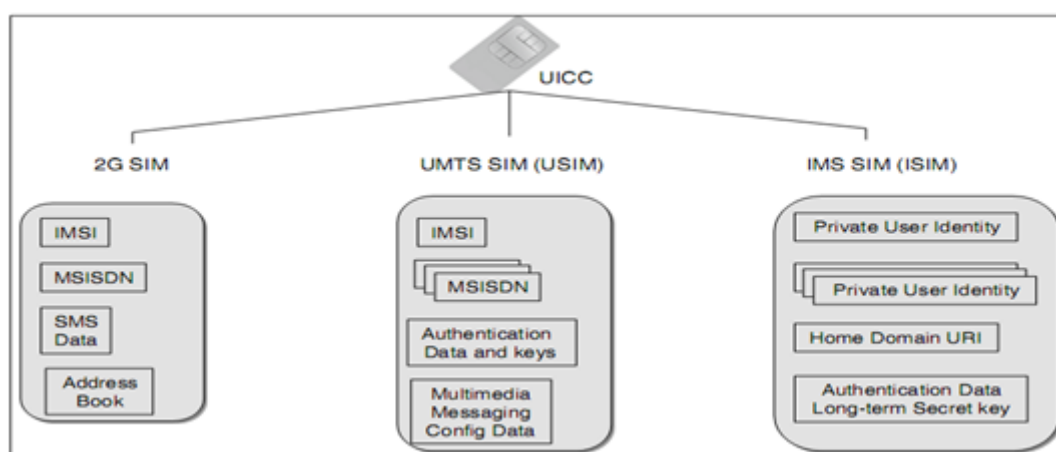


Figura 51 - Evolução dos SIM [27]

5.7. Cobranças no IMS (*Charging*) [23]

Como visto nas motivações e razões para a utilização do IMS, a cobrança diferenciada dos serviços utilizados pelo usuário era um dos principais fatores. Procurando diferenciar os serviços, nessa arquitetura existem dois tipos diferentes de cobrança:

- *Online*
- *Offline*

Outra vantagem é que os dois tipos de cobrança podem estar presentes na mesma sessão. Por exemplo: *offline* para voz e online para vídeo.

5.7.1. Cobrança Online

A cobrança *online* interage em tempo real com a conta do usuário, controlando e monitorando as cobranças relacionadas ao uso de serviços.

Nesse caso, o S-CSCF, o AS e o MRFC estão habilitados para essa cobrança.

Para o procedimento, tanto o AS quanto o MRFC utilizam a interface Ro, enquanto que o S-CSCF usa o *IMS Service Control* (ISC) para se comunicar com o *Online Charging System* (OCS).

5.7.2. Cobrança Offline

A cobrança *offline* é o processo no qual a informação de cobrança é coletada, principalmente, após a finalização da sessão, sem a atualização do sistema de cobrança em tempo real. É o modelo em que o usuário recebe a conta no final do mês. Esse caso tem uma configuração centralizadora, em que todos os nós que geram informações de cobrança estão ligados ao *Charging Collector Function* (CCF), via interface Rf (*Diameter*). O CCF compila todas as informações recebidas de todos os nós e gera os CDRs, que serão enviados para o *Billing System*. Se o usuário não estiver em *roaming*, apenas um CCF é envolvido na bilhetagem. O CGF (*Charging Gateway Function*) provê um mecanismo de para transferir informações de cobrança do SGSN e GGSN para o *Billing System*. Uma diferença em relação ao CCF é que o CGF recebe CDRs válidos.

5.8. Caso de uma sessão IMS

Concluídas as principais entidades e sistemas da arquitetura IMS, pode-se utilizá-los em uma aplicação possível. Para compreender o funcionamento de uma rede com IMS, a seguir é apresentado um caso prático [32].

A figura 52 representa um caso de uma sessão de um usuário no mundo IMS para outro também no mesmo mundo. Suponha que seja uma vídeo-chamada em que o usuário de origem está em *roaming*.

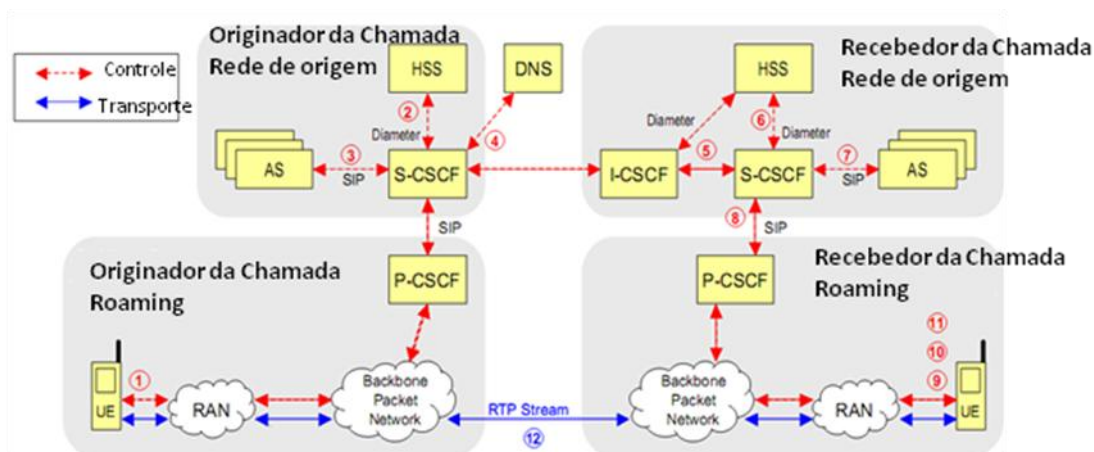


Figura 52 - Caso de uma sessão IMS [32].

1. O UE (*User Equipment*) solicita uma chamada SIP através da RAN (*Radio Access Network*), que é a rede de acesso e a rede de pacotes (NGN). Essa requisição chega ao P-CSCF, que irá autorizar ou não o acesso do usuário
2. O HSS é consultado a partir do pedido do P-CSCF ao S-CSCF com o perfil do usuário para saber quais serviços ele pode ter acesso.
3. Consulta ao *Application Server* se tem contratado o serviço solicitado. No caso, uma vídeo-chamada.
4. Autorizado o usuário, ele consulta o DNS para resolver o endereço da rede de destino e enviar a solicitação SIP para ela.
5. Encontrada a rede destino, o I-CSCF recebe a solicitação, consulta o HSS para obter o S-CSCF do terminal procurado
6. Encontrado o S-CSCF correspondente consulta o HSS para saber se o terminal procura pode receber o serviço, no caso a vídeo- chamada
7. O S-CSCF consulta o AS para verificar a disponibilidade do serviço ao terminal procurado
8. O S-CSCF encaminha a solicitação SIP ao P-CSCF que controla o terminal de destino.
9. Encaminha o pedido através do *backbone* (NGN) e RAN até o usuário de destino
10. Toque e estabelecimento de chamada
11. Atendimento e estabelecimento da vídeo-chamada
12. Através da RAN e do *backbone*, ativação da sessão e controle da vídeo-chamada via RTP.

Basicamente esse é o funcionamento de uma sessão que se utiliza da arquitetura IMS para prover serviços. As operadoras estão bem otimistas quanto ao sucesso do IMS em suas redes. No Capítulo 6 se apresenta um pouco do ambiente real da convergência das redes e a adoção do IMS.

6. Situação Atual e Futuro da NGN e do IMS

Tendo em vista que os capítulos 3, 4 e 5 apresentaram os principais conceitos de redes convergentes, NGN e IMS, respectivamente, a seguir serão apresentados alguns fatores e fatos reais que indicam o sucesso e/ou fracasso dessas implantações.

6.1. Redes convergentes e NGN [33]

Nos últimos anos observou-se uma grande mudança no cenário corporativo das telecomunicações. Grandes empresas começaram a comprar outras, *join ventures* se formaram entre outras ações interessantes. Entre as mais significativas, pode-se citar [33]:

- Compra do Skype pelo eBay por US\$ 4,1 bilhões em 2005 e recentemente sua compra pela Microsoft em 2011
- Compra da Marconi pela Ericsson em 2005 por £\$ 1,2 A fusão da *Lucent* e da *Alcatel* para fornecimento conjunto de soluções em 2006.
- Criação de uma *joint-venture* entre a *Nokia* e a *Siemens* para redes móveis e fixas em 2006. criam uma *joint venture* que cobre tanto as redes fixas como as redes móveis em Junho de 2006.
- A aquisição do *Youtube* pelo *Google* por US\$ 1,65 bilhões em 2006 O Google adquire YouTube por US\$ 1,65 bilhões em Outubro de 2006.
- O anúncio pela *Juniper* e pela *NEC* de planos para soluções FMC (*Fixed-Mobile Communications*) comuns em 2006.

Essas mudanças se deram exatamente por essas empresas buscarem alternativas para o oferecimento de serviços diferenciados em suas redes. E isso não foi um fato isolado de operadoras, mas também das fornecedoras de equipamentos. Assim, a convergência de redes pela NGN já é um fato que vem ocorrendo nesses últimos anos.

Nesse conceito de convergência, novamente outros fatos interessantes ocorreram [33]:

- A *Orange* (Reino Unido) e a *Wanadoo* testam soluções convergentes.
- A *France Telecom* e a *KPN* (Holanda) realizam trials de convergência fixo-móvel com usuários.
- A TIM (Itália) lança o *Video Sharing* baseado em IP chamado.

Assim, o conceito de NGN vai sendo contextualizado no mercado. Ele é capaz de proporcionar um caminho para a migração da rede tradicional para uma rede baseada em IP mantendo os serviços existentes, sendo capaz de oferecer todos os tipos de informação, serviços e mídias. Com o conceito NGN é possível:

- Redução dos custos de implantação e operação (CAPEX e OPEX);
- O aumento do ROI (*Return On Investment*);
- A provisão de uma base comum para serviços fixos, móveis e corporativos;
- Permissão de uma rápida criação e distribuição de serviços inovadores e convergentes sob demanda;
- A possibilidade de dimensionamento flexível e de fácil escalabilidade e centralização do controle da rede para bilhetagem e O&M;
- a atração de serviços providos por terceiros sem que se perca o controle da rede;
- A garantia o uso de interfaces abertas e de diferentes fornecedores, permitindo a escolha de melhores equipamentos.

O processo de migração vem sendo feito aos poucos. As operadoras ainda obtêm muita receita vinda das tradicionais redes legadas, que são extremamente confiáveis, e substituí-las todas de uma vez seria uma alternativa inviável. No entanto, a arquitetura NGN não é perfeita e por isso mesmo há esforços coletivos tentando sempre encontrar alternativas e soluções. Isso se dá também devido a altíssima velocidade que as telecomunicações evoluem. Entre uma dessas idéias, uma das mais promissoras foi o conceito de IMS para a rede IP. Entre as principais promessas dessa estrutura estavam a melhoria do QoS, integração de serviços integrados e tarifação diferenciada desses serviços.

6.2. IMS

A Plataforma IMS apresenta um modelo completamente já padronizado para o controle de sessões, fazendo com que os serviços desenvolvidos não dependam de nenhum tipo de padronização. Ele foi desenvolvido para usuários 3G, que já se utilizam da rede NGN,

Grandes empresas fabricantes de equipamentos de telecomunicações já adotaram o IMS como um padrão a ser explorado no futuro. Empresas como *Nokia Siemens*, *Ericsson* e *Alcatel-Lucent* já tem em seu escopo de produtos muitos que já utilizam o modelo do IMS. Em 2008, a venda de produtos que comportam IMS cresceu 74% [35]. O mercado mundial também espera um investimento de até \$ 2,5 bilhões de dólares até 2013.

Muitos estudiosos vêem o IMS como o modelo do futuro a ser implantado nas atuais redes convergentes. Ainda mais com a popularização do 3G, como mostra a tabela 2, onde são apresentadas as porcentagens totais de celulares 3G nos principais países que já contam com essa tecnologia, em setembro de 2010. [34]

Tabela 2 - Assinantes 3G por total de celulares

-	País	3G/Total de Celulares
1.º	Coreia do Norte	99,9%
2.º	Japão	94,6%
3.º	Coreia do Sul	71,7%
4.º	Australia	64,6%
5.º	Taiwan	58,1%
6.º	Suécia	55,4%
7.º	Cingapura	49,2%
8.º	Espanha	49,1%
9.º	Eslováquia	48,0%
10.º	Hong Kong	45,5%
11.º	França	40,0%
12.º	Reino Unido	38,1%
13.º	Estados Unidos	31,4%

Fonte: TeleGeography's GlobalComms Database - 2010

Além disso, o LTE, também conhecido como 4G, já está sendo desenvolvido de forma a ser compatível com a arquitetura pregada pelo IMS. Assim, segundo previsões, o modelo ainda poderá ser muito utilizado ao longo dos anos. No entanto, no mundo das telecomunicações a concorrência acirrada faz com que rapidamente alguns conceitos mudem e tecnologias se tornem obsoletas.

O conceito da arquitetura IMS não está pronto e com as constantes mudanças nas tecnologias das telecomunicações sempre há a necessidade de se alterar alguma coisa. O gráfico da figura 53 apresenta as dificuldades encontradas pelo IMS para se consolidar.

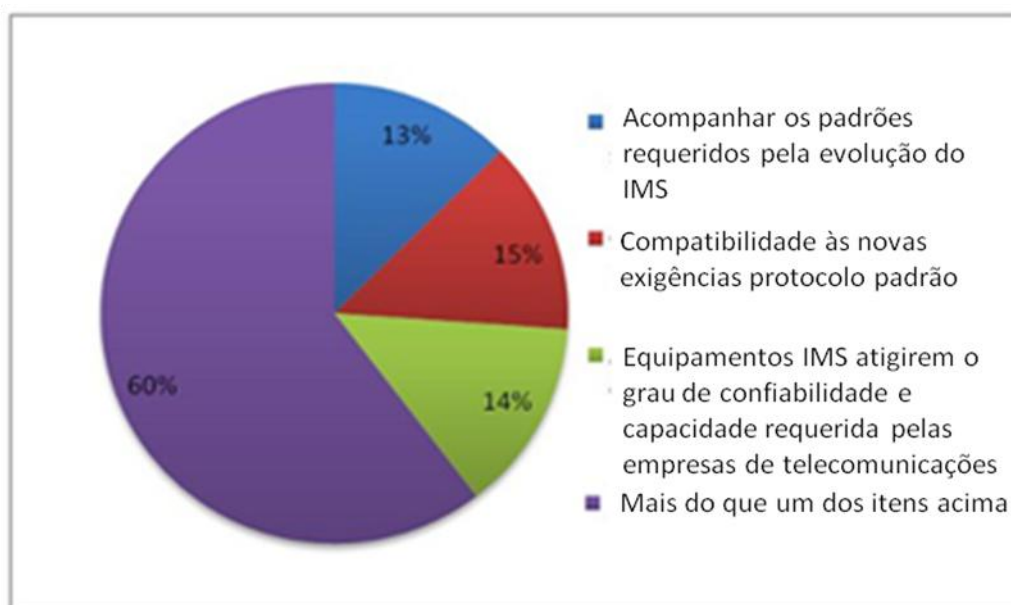


Figura 53 - Dificuldades do IMS [36]

Pelo gráfico apresentado na figura 53, observa-se que o IMS ainda precisa de muitas atualizações e adaptações para se fixar como um padrão mundial. A sua compatibilidade e confiabilidade ainda são questionáveis e por isso que os órgãos padronizadores responsáveis concentram tantos esforços conjuntos para tornarem o IMS eficaz e confiável.

Por, apesar de ser uma promessa que pretende resolver e simplificar os problemas enfrentados pelas operadoras, o futuro para o IMS ainda é incerto. No entanto, pela sua atual situação, ainda é o modelo que mais tem chances de se tornar o novo padrão, assim como as redes NGN.

7. Conclusão

A convergência de redes se tornou uma realidade inevitável no mercado de telecomunicações. As operadoras precisavam aumentar a quantidade de serviços prestados para clientes cada vez mais exigentes. Esse processo iniciou-se com a voz sendo transmitida por redes de dados e posteriormente, outros serviços foram introduzidos, como as mídias de vídeos.

Entre os vários modelos apresentados para a convergência de redes, o mais aceito foi da transmissão de outros serviços sobre a rede IP. Isso porque esta é a rede de maior abrangência no âmbito mundial, principalmente pela Internet se basear nos protocolos TCP/IP.

A rede adotada pelas operadoras para a convergência é conhecida como NGN, *Next Generation Networks*, baseada na rede IP. A grande inovação da NGN é a separação da operação em camadas, principalmente da sinalização e serviços e transporte. Isso possibilitou a independência de todas as umas dessas camadas, podendo fazer modificações em cada uma dela sem interferir nas outras. As redes NGN trabalham com equipamentos dentro das suas redes que são responsáveis por fazer conversões de uma tecnologia para outra. Esses são conhecidos com gateways, que fazem toda a conversão e manipulação de mídias. No entanto são estruturas sem inteligência própria, que necessitam de um controlador próprio para instruí-las a realizar suas tarefas.

Essa entidade controladora, conhecida como *Softswitch*, é considerada o cérebro de uma rede NGN. É responsável pelo controle de toda rede, pois controla todas as entidades que compõem a rede, realiza controles de chamadas. Também manipula todas as sinalizações, podendo atuar em redes comutadas (SS7) ou de pacotes também (SIP, H.248).

É exatamente nesse ponto que ocorre a separação entre a manipulação de mídia e transporte pelos gateways e a manipulação de sinalização e controle nas sessões pelo *Softswitch*. A NGN trabalha com um conjunto de protocolos que fazem com que a arquitetura seja eficiente. Entre os principais, estão o MGCP, H.248, H.323 e SIP. Tanto o MGCP quanto o H.248 são protocolos de controle dos *Media Gateways*, usados principalmente para interconexão das redes legadas como a TDM e a NGN. Protocolos como o H.323 e SIP também são responsáveis pelo controle, no entanto não se limitam apenas aos *Media Gateways* e sim a toda rede IP. Eles fazem o trabalho de estabelecer, modificar e finalizar sessões multimídias. No entanto, o H.323 se mostrou mais eficiente apenas em chamadas telefônicas. Por essa razão, o SIP acabou se tornando o principal protocolo para esse controle de sessões, sendo amplamente utilizado nas redes atuais. Também por esse motivo, se tornou o principal protocolo utilizado pelo IMS, que é considerado o futuro das arquiteturas da rede IP.

O *IP Multimedia Subsystem*, ou IMS é uma arquitetura de referência que visa à entrega de serviços de comunicação de voz, vídeo e dados através de uma rede IP. Ele pretende melhorar a qualidade dos serviços oferecidos pelas operadoras que adotaram o sistema de redes convergentes (NGN). Foi desenvolvido inicialmente para usuários que utilizam 3G. O IMS não tem como objetivo uma mudança na infraestrutura de rede e oferecimento de serviços, mas sim propor um novo modelo de arquitetura que tenha como premissa o desenvolvimento de três pontos básicos: qualidade de serviço; tarifação diferenciada; e integração de serviços variados.

A arquitetura IMS é uma padronização única, desenvolvida por vários dos principais órgãos responsáveis pela padronização das telecomunicações em conjunto. Baseando-se num modelo em separação em camadas, é possível a criação de novos serviços sem que precisem se adaptar a rede. O ponto chave é o núcleo da arquitetura. Este é composto de bancos de dados, *SIP Servers* que fazem toda a parte do controle das sessões; *Application Servers*; entre outras estruturas que foram descritas neste estudo. Todos esses elementos em conjunto colaboram em conjunto para a solução completa do IMS.

O futuro do IMS ainda é desconhecido devido as evoluções e mudanças constantes que ocorrem no campo das telecomunicações. Seu modelo de arquitetura atualmente está sendo muito bem aceito, sendo assim desenvolvidos muitos equipamentos com os seus conceitos. Com a popularização dos serviços 3G, os quais o IMS foi especialmente desenvolvido, o padrão provavelmente ainda durará por um bom tempo. Ainda mais com o LTE, também conhecido como 4G, já apresentar a solução com a presença do núcleo do IMS. No entanto, como a evolução da área das telecomunicações se dá num ritmo muito acelerado, pode surgir rapidamente outra solução melhor e de maior confiabilidade. Mesmo assim, para as operadoras estão essa nova arquitetura se apresenta como um modelo que poderá prover melhor qualidade no serviço e conseqüentemente maior lucro para elas.

Referências Bibliográficas

- [1] ASSOCIATES, H. I.; *Telecommunications A Beginner's Guide*; 1ª Ed; McGraw-Hill/Osborne; 2002
- [2] FARLEY, T.; *Telephone History Series*; Disponível em: [Http://www.telephonetribute.com/pdf/telephone_history_series_rev1.pdf](http://www.telephonetribute.com/pdf/telephone_history_series_rev1.pdf); Acesso em Maio 2011
- [3] <http://ericssonhistory.com/>; Acesso em maio 2011
- [4] TROPICO, Treinamento; *Tecnologia Voz sobre IP – VoIP*; 2009
- [5] <http://www.heritageimages.com/Preview/PreviewPage.aspx?id=2358703&pricing=true&licenseType=RM>; Acesso em Março 2011
- [6] TROPICO, Treinamento, *Telefonia para especialistas*, 2009
- [7] http://en.wikipedia.org/wiki/Harvard_Mark_I, Acessado em Maio de 2011
- [8] TANENBAUM, A. S., *Redes de Computadores*, 4ª ed, Elsevier, 2003.
- [9] http://pt.wikipedia.org/wiki/Modelo_OSI; Acesso em Março 2011
- [10] <http://pt.wikipedia.org/wiki/TCP/IP>; Acesso em Março 2011
- [11] RAMACHANDRA, R; *Evolution to 3G Mobile Communication*; Disponível em: <http://www.ias.ac.in/resonance/Sept2003/pdf/Sept2003p60-72.pdf>; Acesso em Maio 2011
- [12] http://kylerhorner.com/blog/index_files/category-history.php; Acesso em Maio 2011
- [13] *Wireless Communication Systems and Standard*; Disponível em: <http://www.wirelesscommunication.nl/reference/chaptr01/evolut.htm>; Acesso em Maio 2011
- [14] http://www.jofamericanscience.org/journals/am_sci/am0612/08_3205am0612_63_68.pdf

- [15] ALBERTI, A. M.; *Convergência Digital em Telecomunicações: Das Redes Especializadas à Internet do Futuro*, Inatel, 2009
- [16] *Curso: Básico de Redes Convergentes*; Disponível em http://www.portaldigitro.com.br/pt/suporte_ensinare.php; Acesso em Fevereiro 2011
- [17] FUNICELLI, V.B.; *NGN e IMS I: Redes Legadas e Redes Convergentes*, Teleco, 2007
- [18] FILHO, H. B.; SANCHEZ, W. P.; *NGN*, Teleco, 2004
- [19] <http://www.itu.int/ITU-T/ngn/fgngn/index.html>; Acesso em Maio de 2011
- [20] <http://www.imsforum.org/>; Acesso em Maio de 2011
- [21] GOLENIEWSKI, L.; *Telecommunications Essentials*, 1ª Ed.; Pearson, 2002
- [22] BAZOT, P.; HUBER, R.; JACKSON, C.; KAPPEL, J.; MARTIN, C.; SUBRAMANIAN, B. S.; SUR, A.; OGUEJIOFOR, E.; GEORGES, B., *Developing SIP and IP Multimedia Subsystem (IMS) Applications*, IBM Redbooks, 2007
- [23] CAMARILLO, G; GARCIA-MARTIN, M. A; *The 3G IP Multimedia Subsystem*; 2ª Ed; John Wiley & Sons Ltd; 2006
- [24] <http://www.tti-telecom.com/solutions/ims-service-assurance#>, Acesso Abril 2011
- [25] HANDA, A.; *System Engineering for IMS Networks*; 1ª Ed.; Newnes; 2009
- [26] RUSSEL, T.; *The IP Multimedia Subsystem*, 1ª Ed.; McGraw-Hill, 2008
- [27] <http://www.ericsson.com/campaign/ims>; Acesso em Maio de 2011
- [28] POIKSEKÄ, M.; MAYER, G.; KHARTABIL, H.; NIEMI, A.; *The IMS IP Multimedia Concepts and Services in the Mobile Domain*; 1ª Ed; John Wiley & Sons Ltd; 2004
- [29] Copeland, R; *Converging NGN Wireline and Mobile 3G Networks with IMS*; Auerbach Publications, 2009

[30]KNIGHT, D; *IMS based NGN Architecture and its application*; Disponível em http://www.itu.int/ITU-T/worksem/ngn/200604/presentation/s2_knight.pdf; Acesso Maio de 2011

[31]CHAKRABORTY, S.; FRANKKILA, T; PEISA, J.; SYNNERGREN, P.; *IMS Multimedia Telephony over Cellular Systems*; 1ª Ed.; John Wiley & Sons Ltd; 2007

[32] TOWLE, T.T; *IMS in Next Generation Networks*; Disponível em: <http://www.itu.int/ITU-T/worksem/ngntech/presentations/s1-towle.pdf> ; Acesso em Maio de 2011

[33]SULTAN, A; *Introduction to NGNs and Convergence*; Disponível em: Http://docbox.etsi.org/tispan/Open/Information/NGN_Presentations/Boswarthick_for_ETSI_ternal_training_NGN.ppt; Acesso em Maio de 2011

[34] <http://www.teleco.com.br/pais/3g.asp>; Acesso em Maio de 2011

[35] <http://www.fiercevoip.com/story/ims-equipment-sales-grow-74-2009/2009-06-08>; Acesso em Maio de 2011

[36]<http://blog.radvision.com/voipsurvivor/2010/06/07/lte-and-the-future-of-ims-a-post-webinar-post/>; Acesso em Maio de 2011