

CENTRO UNIVERSITÁRIO DO ESTADO DO PARÁ
ÁREA DE CIÊNCIAS EXATAS E TECNOLOGIA
BACHARELADO EM ENGENHARIA DA COMPUTAÇÃO

Rodrigo Formigosa Lima Pena de Souza

VIRTUALIZAÇÃO DO EVOLVED PACKET CORE

Belém

2017

CENTRO UNIVERSITÁRIO DO ESTADO DO PARÁ
ÁREA DE CIÊNCIAS EXATAS E TECNOLOGIA
BACHARELADO EM ENGENHARIA DA COMPUTAÇÃO

Rodrigo Formigosa Lima Pena de Souza

VIRTUALIZAÇÃO DO EVOLVED PACKET CORE

Trabalho de Curso na modalidade Monografia, apresentado como requisito parcial para obtenção do grau em Bacharelado em Engenharia da Computação do Centro Universitário do Estado do Pará – CESUPA, sob orientação da Professora Msc. Michelle Bitar Lelis dos Santos.

Belém

2017

Dados Internacionais de Catalogação-na-publicação (CIP)

Biblioteca do Cesupa, Belém - PA

Souza, Rodrigo Formigosa Lima Pena de.

Virtualização do Evolved Packet Core / Rodrigo Formigosa Lima Pena de Souza; orientação de Michelle Bitar Lelis dos Santos, 2017.

Trabalho de Conclusão de Curso (Bacharelado em Engenharia de Computação) - Centro Universitário do Estado do Pará, 2017.

1. Telefonia móvel. 2. Redes. 3. Virtualização. I. Santos, Michelle Bitar Lelis dos (orient.). II. Título.

CDD. 23º ed. 004.6

Rodrigo Formigosa Lima Pena de Souza

VIRTUALIZAÇÃO DO EVOLVED PACKET CORE

Trabalho de Curso apresentado na modalidade monografia, como requisito parcial para obtenção do grau em Bacharelado em Engenharia da Computação do Centro Universitário do Estado do Pará – CESUPA.

Data da Defesa: 13/12/2017

Banca Examinadora:

**Orientadora Profa. Msc. Michelle Bitar Lelis dos Santos –
CESUPA**

Prof. Esp. Itamar Jorge Vilhena De Brito – CESUPA

Prof. Msc. Paulo Tadeu Resque de Souza – UFPA

Belém

2017

Dedico este trabalho à minha família, principalmente minha mãe, à minha orientadora e meus amigos. Todos eles foram de extrema importância para a realização do mesmo.

Dedico, também, a todos que forneceram suporte para eu chegar a este passo da minha vida.

AGRADECIMENTOS

À minha família, em especial, minha mãe Noely Formigosa de Lima, que além de todo amor e carinho, participou ativamente me encorajando a fazer este trabalho. Ao meu irmão Jairo Felipe Formigosa que participou ativamente me auxiliando para conclusão deste trabalho.

À minha orientadora Michelle Bitar, que não possuo palavras para descrever o quanto me ajudou a realizar este trabalho do jeito que ele está presente. Ajudou-me na minha preparação para obstáculos muito maiores que irão vir em seguida. Mostrou esse possível tema pelo qual eu me apaixonei. Muito obrigado, Professora.

A todo o corpo de professores e funcionários do CESUPA, em especial, ao Professor Itamar Brito que nos mostra um mundo novo, que é a engenharia e seus valores, a todos os seus alunos com sua maneira única de falar. A todos os professores, desde o maternal até o corpo docente do CESUPA que me proporcionaram todo o conhecimento que possuo hoje. Sem vocês, nada disso seria possível.

A todos os meus amigos que me acompanharam até aqui. Não irei citar nomes específicos para não desmerecer nenhum, mas todos sabem o quanto foi difícil esse período da minha vida e o quanto eu precisei do apoio de cada um deles.

Rodrigo Formigosa Lima Pena de Souza

涙を流すたび人は強くなる 涙を流すたび思い出が
-矢口 真里

RESUMO

Esta monografia disserta sobre a virtualização do Evolved Packet Core (EPC), abordando a arquitetura do 4G e detalhando as funções de rede do EPC. O EPC é o núcleo da rede 4G e, sua virtualização consiste no desacoplamento do hardware e software utilizados nos equipamentos por datacenters padronizados que executam máquinas virtuais para a realização das funções da rede. A virtualização mostra-se de grande valor para as operadoras de telefonia móvel, proporcionando diversas vantagens técnicas e econômicas. Entre elas, uma rede mais flexível, que atenderá a demanda de mercado de forma mais rápida e econômica, menores custos de operação de equipamentos e a introdução de novos serviços de maneira mais simples. A virtualização do EPC é fundamental para as operadoras móveis, representando o primeiro momento da evolução da rede núcleo rumo ao 5G. Porém, a implementação dessa função de rede representa também uma série de desafios até que as operadoras possam usufruir plenamente todos os benefícios prometidos.

Palavras-chave: Evolved Packet Core. 4G. Long Term Evolution. Virtualização. EPC. vEPC. NFV.

ABSTRACT

This monograph discusses the Evolved Packet Core (EPC) virtualization, addressing the 4G architecture and detailing the EPC network functions. The EPC is the 4G core network, and its virtualization consists of decoupling the hardware and software used in the network nodes by implementing standard datacenters that run virtual machines to perform the network functions. Virtualization is of great value to mobile operators, providing many technical and financial advantages. Among them, a more flexible network, which will respond to market demand more quickly with low costs, lower equipment operation costs and the introduction of new services in a simpler way. EPC virtualization is critical for mobile operators, representing the first moment of the core network evolution towards 5G. However, the implementation of this network function also represents a series of challenges until operators can fully enjoy the promised benefits.

Keywords: Evolved Packet Core. 4G Technology. Long Term Evolution. Virtualization. EPC. vEPC. NFV.

LISTA DE FIGURAS

Figura 1 - A evolução dos assinantes móveis	22
Figura 2 - Representação do sistema celular	24
Figura 3 - Arquitetura GSM e GPRS	29
Figura 4 - Comparação com arquitetura das redes 3G e 4G	34
Figura 5 - Arquitetura básica para o LTE	36
Figura 6 - Mudança na arquitetura UTRAN e Evolved UTRAN	38
Figura 7 - Divisão entre plano de dados e plano de controle	38
Figura 8 - Adicionando os nós PCRF/OCS/OFCS à arquitetura básica do EPC	40
Figura 9 - Arquitetura EPS para VoLTE	42
Figura 10 - Parâmetros de QoS na rede LTE	44
Figura 11 - Classes de QoS	44
Figura 12 - Bearers e QCI	45
Figura 13 - Funções do HSS e suas interfaces	47
Figura 14 - MME e a comunicação com as entidades do EPS	50
Figura 15 - SGW e a comunicação com as entidades do EPS	55
Figura 16 - PDN GW e a comunicação com as entidades do EPS	58
Figura 17 - Inspeção de tráfego: SPI, DPI e análise heurística	60
Figura 18 - Visão geral da arquitetura do PCC	62
Figura 19 - Arquitetura EPS com as tecnologias padronizadas pelo 3GPP	64
Figura 20 - Caminho percorrido pelo pacote	66
Figura 21 - Volume de dados utilizados relacionando tipos de dispositivos e aplicações.	67
Figura 22 - Cenário de arquitetura centralizada	70
Figura 23 - Cenário de arquitetura totalmente distribuída	71
Figura 24 - Representação do pool de MME	71
Figura 25 - Cenário de arquitetura com somente os GWs distribuídos	73
Figura 26 - Arquitetura de um sistema tradicional	75
Figura 27 - Arquitetura de um sistema virtualizado	76
Figura 28 - Hypervisor agindo como interface entre os recursos de hardware e as VMs	77

Figura 29 - Arquitetura do NFV	78
Figura 30 - Arquitetura do VNF	80
Figura 31 - Instância de VNF.....	81
Figura 32 - Invalidando o compartilhamento de VNFC.....	81
Figura 33 - Network Slice e suas aplicações.....	82
Figura 34 - Virtualização do EPC	83
Figura 35 - Arquitetura do vEPC baseada na arquitetura do NFV.....	84
Figura 36 - Crescimento dos dados na rede e arquitetura clássica de implementação	85
Figura 37 - Eficiência dos custos de 2016 até 2020.....	86

LISTA DE TABELAS

Tabela 1 - Marcos históricos	25
Tabela 2 - Tecnologias padronizadas pelo 3GPP e suas taxas de transferência	35

LISTA DE SIGLAS

1G	Primeira Geração de Telefonia Móvel
2G	Segunda Geração de Telefonia Móvel
3G	Terceira Geração de Telefonia Móvel
3GPP	<i>Third Generation Partnership Project</i>
4G	Quarta Geração de Telefonia Móvel
AKA	<i>Authentication and Key Agreement</i>
AMBR	<i>Aggregate Maximum Bit Rate</i>
AMPS	<i>Advanced Mobile Phone Service</i>
Anatel	Agência Nacional de Telecomunicações
APN	<i>Access Point Name</i>
AT&T	<i>American Telephone and Telegraph</i>
CEPT	<i>Conference Europeenne des Postes et Telecommunications</i>
CSFB	<i>Circuit Switch Fallback</i>
DPI	<i>Deep Packet Inspection</i>
DSCP	<i>Differentiated Services Code Point</i>
EDGE	<i>Enhanced Data Rates for GSM Evolution</i>
EPC	<i>Evolved Packet Core</i>
EPS	<i>Evolved Packet System</i>
ERB	Estação Rádio Base
ETSI	<i>European Telecommunications Standards Institute</i>
FCC	<i>Federal Communications Commission</i>
GERAN	GSM EDGE Radio Access Network
GBR	<i>Guaranteed Bit Rate</i>
GGSN	<i>Gateway GPRS Support Node</i>
GPRS	<i>General Packet Radio Service</i>
GSM	<i>Global System for Mobile Communications</i>
GSMA	<i>GSM Association</i>
GTP	<i>GPRS Tunneling Protocol</i>
HLR	<i>Home Location Register</i>
HSDPA	<i>High Speed Downlink Packet Access</i>
HSPA	<i>High Speed Packet Data Access</i>

HSS	<i>Home Subscriber Server</i>
HSUPA	<i>High Speed Uplink Packet Access</i>
IMEI	<i>International Mobile Equipment Identify</i>
IMS	<i>IP Multimedia Subsystem</i>
IP	<i>Internet Protocol</i>
ISO	<i>International Organization for Standardization</i>
ITU	<i>International Telecommunication Union</i>
LTE	<i>Long Term Evolution</i>
MME	<i>Mobility Management Entity</i>
MMS	<i>Multimedia Message Service</i>
MSC	<i>Mobile Switching Center</i>
MTS	<i>Mobile Telephone System</i>
NAS	<i>Non Access Stratum</i>
NFV	<i>Network Functions Virtualization</i>
NMT-400	<i>Nordic Mobile Phone</i>
<i>non-GBR</i>	<i>non-Guaranteed Bit Bate</i>
NTT	<i>Nippon Telephone and Telegraph</i>
OCS	<i>Offline Charging System</i>
OFCS	<i>Online Charging System</i>
OSI	<i>Open Systems Interconnection</i>
PCEF	<i>Policy Charging Enforcement Function</i>
PCRF	<i>Policy and Charging Rules Function</i>
PDW GW	<i>Packet Data Network Gateway</i>
QCI	<i>QoS class of identifier</i>
QoS	<i>Quality of Service</i>
RNC	<i>Radio Network Controller</i>
SDN	<i>Software Defined Networking</i>
SGSN	<i>Serving GPRS Support Node</i>
SGW	<i>Serving Gateway</i>
SMS	<i>Short Message Service</i>
SPI	<i>Shallow Packet Inspection</i>
SPR	<i>Subscriber Profile Repository</i>
TA	<i>Tracking Area</i>
TACS	<i>Total Access Communication System</i>

TAU	<i>Tracking Area Update</i>
TDMA	<i>Time Division Multiple Access</i>
TI	Tecnologia da Informação
TR	<i>Technical Report</i>
TS	<i>Technical Specification</i>
UE	<i>User Equipment</i>
UMTS	<i>Universal Mobile Telephone Service</i>
UTRAN	<i>UMTS Terrestrial Radio Access Network</i>
vEPC	Virtualização do EPC
VLR	<i>Visitor Location Register</i>
VoLGA	<i>Voice over LTE via Generic Access</i>
VoLTE	<i>Voice over LTE</i>
W-CDMA	<i>Wide-Band Code Division Multiple Access</i>
NFV	<i>Network Functions Virtualization</i>
NFV ISG	<i>Network Functions Virtualization Industry Specification Group</i>
VM	<i>Virtual Machines</i>
VMM	<i>Virtual Machine Monitor</i>
VNF	<i>Virtual Network Functions</i>
NFVI	NFV Infrastructure
NFV M&O	<i>NFV Management and Orchestration</i>
OSS	<i>Operations Support Systems</i>
BSS	<i>Business Support Systems</i>
VNFC	<i>VNF Components</i>

SUMÁRIO

1. INTRODUÇÃO	17
1.1 JUSTIFICATIVA.....	19
1.2 OBJETIVO GERAL.....	20
1.3 OBJETIVOS ESPECÍFICOS.....	20
1.4 METODOLOGIA	20
1.5 ESTRUTURA DO TRABALHO	20
2. EVOLUÇÃO DA TELEFONIA MÓVEL	22
2.1 AVANÇOS TECNOLÓGICOS ANTERIORES À PRIMEIRA GERAÇÃO DE TELEFONIA MÓVEL	23
2.2. PRIMEIRA GERAÇÃO DE TELEFONIA MÓVEL (1G)	25
2.3. SEGUNDA GERAÇÃO DE TELEFONIA MÓVEL (2G).....	26
2.3.1 Avanço da Segunda Geração de Telefonia Móvel	28
2.4 TERCEIRA GERAÇÃO DE TELEFONIA MÓVEL (3G)	29
2.5 QUARTA GERAÇÃO DE TELEFONIA MÓVEL (4G)	32
2.6 ARQUITETURA DO EVOLVED PACKET SYSTEM.....	35
3. EVOLVED PACKET CORE (EPC)	46
3.1 HOME SUBSCRIBER SERVER (HSS)	46
3.2 MOBILE MANAGEMENT ENTITY (MME)	48
3.3 SERVING GATEWAY (SGW).....	54
3.4 PACKET DATA NETWORK GATEWAY (PDN GW).....	57
4. ARQUITETURA E SERVIÇOS	64
5. VIRTUALIZAÇÃO DO EPC	74
6. CONCLUSÃO	88
6.1 DIFICULDADES ENCONTRADAS	89
6.2 TRABALHOS FUTUROS.....	89
REFERÊNCIAS	89

1. INTRODUÇÃO

A quarta geração de telefonia móvel, ou 4G, ou ainda, *Long Term Evolution* (LTE) é uma realidade na sociedade atual e uma tecnologia que revoluciona nosso dia-a-dia há alguns anos. A Agência Nacional de Telecomunicações salienta que 41% da população brasileira utilizam o LTE (ANATEL, 2017).

O 4G foi desenvolvido para ofertar internet banda larga de alto desempenho em comparação as tecnologias anteriores. Com o 4G, a banda larga móvel é finalmente uma realidade. Serviços como videoconferência, *e-mail*, mapas eletrônicos, acesso a bancos, *streaming* de vídeo, redes sociais e mensagens instantâneas são considerados por muitas pessoas como essenciais e necessários em qualquer lugar. Como consequência desta demanda, a procura pelo 4G se tornou cada vez maior (MAZZONI, 2014).

O *Third Generation Partnership Project* (3GPP), órgão padronizador do 4G, visualizou essa tendência de mercado e pautou seus esforços para uma Internet móvel, priorizando o aumento da velocidade para transferência de dados (chamado *throughput*), a eficiência espectral dos sistemas e a redução da latência da rede.

Levando em conta a diversidade de aplicações, é mais que necessária uma implantação rápida de uma rede 4G com qualidade e baixo custo, mas essa tarefa não é simples. Alguns aplicativos como *WhatsApp*, *YouTube*, *Netflix*, *Skype*, *Instagram*, exigem uma resposta diferente e mais eficiente da rede. Sendo assim, as operadoras de telefonia móvel necessitam de uma arquitetura de rede flexível, que facilite a implementação de novos serviços, de baixo custo e de fácil gerenciamento.

Tradicionalmente, as funções realizadas pela rede 4G foram implantadas por equipamentos com *hardwares* e *softwares* específicos, acoplados e dependentes. Esses equipamentos são produzidos por fornecedores que possuem patentes sobre os seus equipamentos.

Se for levado em conta um país como o Brasil, a problemática de implantar uma rede 4G fornecendo novos serviços é agravada devido à sua grande extensão territorial. Uma operadora possui um gasto demasiado em infraestrutura de rede. Também deve-se pensar em inúmeras dificuldades

físicas para a implantação de fibra ótica ou enlace de rádio frequência para atingir a cobertura máxima.

Considerando-se todos os pontos citados anteriormente e a necessidade imposta pelo usuário final. A virtualização dos equipamentos da rede 4G é a solução de diversos problemas que as operadoras enfrentam atualmente. Marco Di Costanzo, atual Diretor de Engenharia da TIM Brasil, afirma: “A virtualização da rede é uma tendência clara e irreversível e representa uma transformação que impulsionará a evolução dos serviços tradicionais para os digitais, promovendo o desenvolvimento da sociedade e economias digitais.” (OLHAR DIGITAL, 2016, *online*). A virtualização também representa uma peça chave para a próxima evolução da telefonia móvel, chamada de 5G.

O primeiro passo para ter-se uma rede virtualizada é o *Network Functions Virtualization* (NFV), que é a implementação da virtualização dos equipamentos e funções da rede. O segundo passo é a implementação do *Software Defined Networking* (SDN), que irá trazer toda a potencialidade da virtualização. Essa combinação é a base do núcleo de rede do 5G.

Pesquisas de mercado mostram que com o núcleo da rede 4G virtualizado. A operadora em cinco anos poderá obter uma eficiência de custo de operação de cerca de 25% sem alterar a sua arquitetura, resultando em uma economia de 20% ao longo de cinco anos. Alterando-se a sua arquitetura com a centralização da sinalização de controle da rede e a distribuição dos equipamentos que realmente lidam com o tráfego de dados do usuário, a eficiência dos custos de operação poderá chegar até 36% ao longo de cinco anos, resultando em uma economia de 30% para a operadora em cinco anos (IDC, 2016).

Uma rede virtualizada, além de representar economia de custos fixos e variáveis para a operadora, significa uma rede mais flexível. Isto ocorre pelo fato de que a virtualização é composta por um *hardware* único que implementa todas as funcionalidades do núcleo da rede LTE. Alguns dos principais benefícios da virtualização do 4G são: desacoplamento entre *hardware* e *software*, implementação de novos serviços de forma mais rápida e flexível e melhor gerenciamento das funções virtualizadas. Isto garante um aproveitamento dos recursos de infraestrutura, redundância de forma mais eficiente e rápida. Porém, há alguns desafios para a implementação da virtualização, principalmente a

comunicação dos sistemas virtualizados e os equipamentos tradicionais. Outro problema a ser levado em consideração é garantir que os sistemas virtualizados possuam um certo nível de resiliência contra falhas de software e hardware.

Este trabalho disserta sobre a virtualização do núcleo da rede LTE, fazendo um histórico da telefonia móvel até o 4G, apresentando a tecnologia e o seu funcionamento. O trabalho se propõe a discutir a virtualização do núcleo da tecnologia 4G, como será implementada essa transformação e abordará aspectos técnicos e econômicos.

1.1 JUSTIFICATIVA

A IDC Analyze the Future, empresa líder em inteligência de mercado, consultoria em telecomunicações e mercados de tecnologia, realizou um estudo em 2016, mostrando que as operadoras de telefonia móvel estão sofrendo baixo crescimento na receita de serviços móveis.

Um dos maiores problemas das operadoras é aumentar a eficiência nos investimentos para aquisição de novos equipamentos de rede. Estes equipamentos são caros, tanto em *hardware* quanto em *software*. Além disso, por serem de tecnologia proprietária dos fornecedores de redes de telecomunicações, a operadora depende dos fornecedores para realizar a implantação. Esta implantação na maioria das vezes é demorada, atrasando o tempo de mercado para implantar um novo serviço ou lançar ofertas. Vale citar que os projetistas de rede necessitam ter em mente diversos problemas derivados desta condição. Assim como o problema de gerenciamento da rede, as operadoras dependem de funcionários treinados pelos fornecedores para oferecer manutenção e gerenciar a rede, então há a necessidade de os fornecedores realizarem esse treinamento.

Deve-se levar em conta que é indispensável ter equipamentos em duplicidade, assim a operadora pode realizar manutenção ou transferir a carga caso algum equipamento apresente uma falha. Isto torna a rede ainda mais cara.

A virtualização da rede núcleo do 4G surge para extinguir uma série de problemáticas em um contexto de uma operadora. A virtualização dos elementos do núcleo do LTE, além de garantir à operadora, uma maior flexibilidade e baixos custos através do desacoplamento do *hardware* e *software*, é o primeiro passo para a evolução do núcleo do 4G para 5G. Por isso, a virtualização é um passo

inevitável para todas as operadoras de telefonia móvel para implementarem a tecnologia 5G.

1.2 OBJETIVO GERAL

O objetivo deste trabalho é primeiramente compreender a tecnologia LTE e o funcionamento do núcleo da rede, o *Evolved Packet Core* (EPC). Posteriormente, esclarecer de que maneira será implementada a virtualização dos equipamentos deste núcleo, desafios e benefícios técnicos e econômicos.

1.3 OBJETIVOS ESPECÍFICOS

- Explicar inteiramente a arquitetura de rede da quarta geração de telefonia móvel (4G);
- Descrever os componentes do núcleo da rede 4G;
- Abordar a virtualização do *Evolved Packet Core*.

1.4 METODOLOGIA

Este trabalho faz uma pesquisa descritivo-explicativa sobre a tecnologia 4G e a virtualização do seu núcleo. Foram realizadas pesquisas em livros, artigos acadêmicos, monografias, normas técnicas do órgão padronizador da tecnologia móvel e documentos elaborados pelos fornecedores de equipamentos de telefonia móvel.

Foi produzido um estudo baseado em bibliografias e documentos técnicos sobre a tecnologia de telefonia móvel de quarta geração (4G) e o funcionamento do núcleo desta tecnologia, conhecido como *Evolved Packet Core* (EPC). Assim foi analisada completamente a arquitetura da rede. Em seguida, será analisada a evolução do EPC através da virtualização dessas funções de rede.

1.5 ESTRUTURA DO TRABALHO

Além deste capítulo introdutório e das considerações finais, este trabalho está organizado do seguinte modo:

O capítulo 2 mostra uma visão geral sobre a evolução tecnológica da telefonia móvel e sua trajetória, desde o desenvolvimento da tecnologia antes da

primeira geração de telefonia móvel (1G) até à quarta geração de telefonia móvel (4G). Em seguida, descrevem-se a visão geral sobre a arquitetura da tecnologia 4G e diversos aspectos sobre esta tecnologia e o seu funcionamento.

O capítulo 3 descreve os componentes básicos do EPC que fornecem conectividade ao usuário na rede LTE e suas funcionalidades.

O capítulo 4 disserta sobre as interconexões entre os equipamentos do núcleo da rede LTE, a distribuição dos equipamentos na rede e quais serviços possuem vantagens de acordo com cada arquitetura de rede.

O capítulo 5 explica como será implementada a virtualização da rede e possíveis arquiteturas de rede. Apresenta desafios para *data centers*, e também aborda benefícios técnicos e econômicos desta tecnologia e tendências de mercado que virão com a virtualização das funções da rede LTE.

2. EVOLUÇÃO DA TELEFONIA MÓVEL

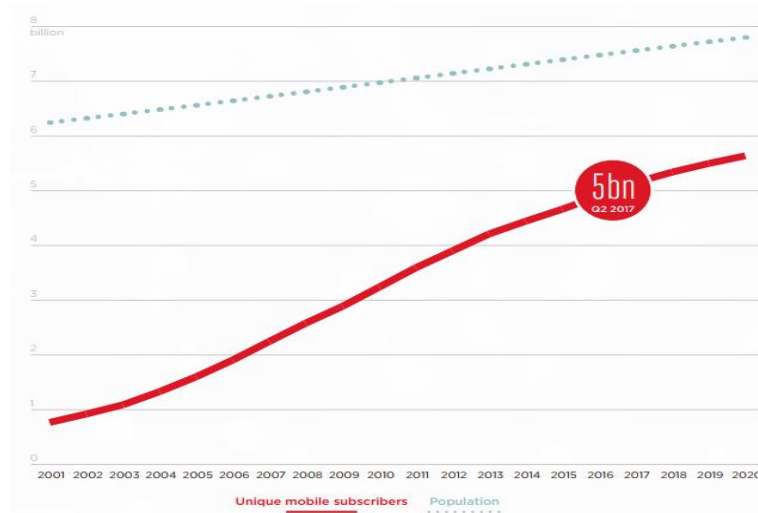
A telefonia móvel no mundo passou por grandes evoluções desde 1983, ano em que os Estados Unidos, através da empresa Ameritech, lançou o *Advanced Mobile Phone Service* (AMPS), o primeiro sistema comercial de telefonia móvel celular norte-americano. A implantação desse sistema ocorreu na cidade de Chicago, operando apenas com serviço de voz (GHOSH; ZHANG; MUHAMED, 2010).

Cumpramos ressaltar que, apesar de o AMPS não ter sido o primeiro sistema de telefonia móvel celular no mundo, foi o mais bem-sucedido, e suas variações continuaram a ser utilizadas mesmo após a implantação de outras tecnologias. Esse sucesso fomentou o desenvolvimento de tecnologias mais avançadas.

O serviço de voz impulsionou o crescimento da telefonia móvel no início de sua criação, todavia, atualmente, o serviço de internet móvel é o que mais se desenvolve e se tornou o foco para as operadoras. Nas últimas duas décadas, a Internet tornou-se indispensável para o mundo, mudando a forma de como se compartilham informações, conduzem-se negócios e busca-se entretenimento.

A *GSM Association* (GSMA), uma importante associação de operadoras móveis no mundo, fez uma publicação em setembro de 2017, apontando que o número de assinantes no mundo chegou a 5 bilhões, e que até o final desta década, esse número tende a aumentar para 5,7 bilhões (GSMA, 2017). Conforme a figura 1:

Figura 1 - A evolução dos assinantes móveis



Fonte: GSMA Intelligence (2017, p. 4)

Outra estatística que vale a pena ser comentada são os dados coletados do *International Telecommunication Union* (ITU), uma importante agência da Organização das Nações Unidas, onde são feitos padronizações e estudos sobre telecomunicações no mundo. Esses dados mostram que o Brasil é o quinto maior mercado de telefonia móvel do mundo em número de assinantes desde 2011.

2.1 AVANÇOS TECNOLÓGICOS ANTERIORES À PRIMEIRA GERAÇÃO DE TELEFONIA MÓVEL

O primeiro uso de sistemas de rádio móvel aconteceu em 1921, em que o departamento de polícia de Detroit instalou em viaturas um sistema analógico de frequência de rádio. A princípio, esse sistema só recebia ordens da central de polícia, mas posteriormente também passou se comunicar com o departamento de polícia. Esse sistema também foi utilizado pelo Departamento de Polícia Civil de Nova York, no qual aproximadamente 5.000 veículos compartilhavam apenas 11 canais de voz (GUEDES, 2014).

Em 1946, a empresa *American Telephone and Telegraph* (AT&T) introduziu o *Mobile Telephone System* (MTS), o primeiro serviço de telefonia móvel do mundo, um sistema manual com capacidade de suportar no máximo três chamadas simultâneas, em St. Louis, Missouri. Já em 1948, a AT&T expandiu o serviço para mais 100 cidades nos Estados Unidos, contando com mais de 5.000 clientes nesse ano (GHOSH; ZHANG; MUHAMED, 2010).

Nesse sistema, as Estações Rádio Base (ERB), infraestruturas de rádio que permitem o acesso do equipamento do usuário à rede da operadora, foram projetadas para ser amplificadores com altas taxas de potência, mas isso gerava um problema de gasto de energia em comparação com ERB's atuais. Esses equipamentos visavam cobrir a maior área possível, muitas vezes, uma cidade inteira, porém eram extremamente limitados em sua capacidade (GHOSH; ZHANG; MUHAMED, 2010).

Um exemplo disso foi a cidade de Nova York. O sistema poderia suportar 2.000 assinantes em 2,5 mil quilômetros quadrados. Mesmo assim, aqueles poucos clientes muitas vezes tinham que esperar até 30 minutos para fazer uma

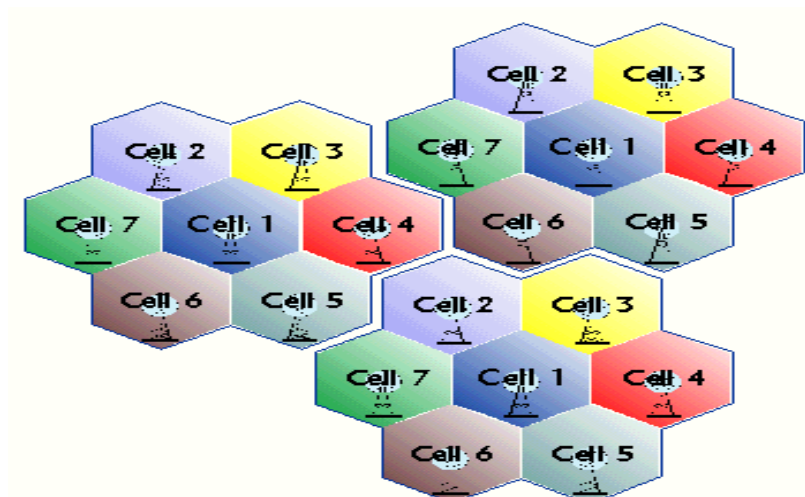
chamada. Isto acontecia pelo fato de que as linhas estavam sempre ocupadas. Havia uma crescente demanda por serviços móveis e uma pressão cada vez maior para encontrar uma nova maneira de suportar mais usuários (GHOSH; ZHANG; MUHAMED, 2010).

Em 1947 a empresa Bell Labs, pioneira no desenvolvimento de várias tecnologias que são usadas até hoje, idealizou o conceito de sistema de telefonia celular em telecomunicações e proporcionou a solução do problema de suportar mais usuários.

A ideia do sistema de telefonia celular consiste em: substituir uma única ERB, que antes cobria uma grande área, por várias ERB's com baixa potência e, assim, fornecendo uma área de cobertura menor e utilizando uma porção mais reduzida do espectro de rádio total. Com isso, as frequências poderiam então ser reutilizadas na área de cobertura do sistema, desde que as ERB's que utilizassem a mesma frequência estivessem suficientemente separadas umas das outras. Essas áreas menores atendidas por uma ERB foram chamadas de células (GHOSH; ZHANG; MUHAMED, 2010).

A figura 2 exemplifica o conceito criado pela Bell Labs.

Figura 2 - Representação do sistema celular



Fonte: Site GEOCITIES (2017, *online*)

Embora a Bell Labs tenha idealizado o conceito de sistema celular, a tecnologia necessária para sua implementação não foi disponibilizada até a década de 1970. A AT&T apresentou uma proposta ao *Federal Communications*

Commission (FCC), órgão regulador americano de telecomunicações, sobre o conceito de sistema de telefonia celular em 1971. Após vários anos de negociações e deliberações, em 1983, a FCC atribuiu 40MHz do espectro na banda de 800MHz. Isto desencadeou a implantação da primeira geração de telefonia móvel comercial nos Estados Unidos. Na tabela 1 mostra-se o avanço cronológico até a implementação do AMPS (GHOSH; ZHANG; MUHAMED, 2010).

Tabela 1 - Marcos históricos

Ano	Marcos Históricos Importantes
1934	Sistema de modulação analógica de rádio frequência implantado em 194 municípios dos EUA voltado à segurança pública.
1946	Primeiro serviço de telefonia móvel introduzido na AT&T em St. Louis, Missouri.
1960-1970	Bell Labs desenvolveu o conceito de telefonia móvel celular.
1971	AT&T apresenta proposta para a implantação do sistema móvel celular à FCC.
1979	NTT no Japão implementa o primeiro sistema celular comercial.
1983	FCC aloca 40 MHz no espectro de 800 MHz para o AMPS.
1983	<i>Advanced Mobile Phone Service</i> (AMPS) é lançado em Chicago.

Fonte: Autor (2017)

2.2. PRIMEIRA GERAÇÃO DE TELEFONIA MÓVEL (1G)

A primeira geração de telefonia móvel foi desenvolvida em três partes do mundo: Estados Unidos, Japão e Europa, e foi caracterizada pela sua modulação analógica que, em suma, consiste na ideia de modificar um sinal eletromagnético, usando uma onda cossenoidal para transportar a informação. O 1G foi desenvolvido principalmente para serviço de voz (GHOSH; ZHANG; MUHAMED, 2010).

A diferença entre os sistemas de telefonia móvel de primeira geração e as tecnologias anteriores foi a implantação do conceito de telefonia celular, no qual se providenciavam trocas entre células (*handover*) e transferências de chamadas em curso (GHOSH; ZHANG; MUHAMED, 2010).

Em 1979, a empresa japonesa *Nippon Telephone and Telegraph* (NTT) implementou o primeiro sistema celular comercial do mundo. Por sua vez, na Europa, o chamado *Nordic Mobile Phone* (NMT-400) foi implantado em 1981. O NMT-400 foi a primeira rede que suportou a transferência automática de chamadas e o *roaming* internacional (GHOSH; ZHANG; MUHAMED, 2010).

O AMPS, desenvolvido pela Bell Labs no final da década de 70, foi considerado o mais bem-sucedido sistema de primeira geração de telefonia móvel, em conjunto com as suas variações que vieram nos anos posteriores. Além dos EUA, o AMPS foi implantado em vários países da América do Sul, Ásia e América do Norte (GHOSH; ZHANG; MUHAMED, 2010).

Roaming é um conceito básico de telefonia móvel celular que é necessário comentar para o entendimento do trabalho. Quando o dispositivo do usuário está fora de sua Área de Mobilidade, o mesmo está em *roaming*. Sendo assim, ele é um usuário visitante no sistema celular daquela região. Área de mobilidade é definida como a área geográfica que é considerada de referência do dispositivo do usuário. (TELECO, 2002).

2.3. SEGUNDA GERAÇÃO DE TELEFONIA MÓVEL (2G)

O 1G possibilitou a realização de chamadas, porém a tecnologia era muito limitada. A segunda geração de telefonia móvel (2G) introduziu novos tipos de serviços como o *Short Message Service* (SMS) e o *Multimedia Message Service* (MMS). Essa tecnologia é utilizada até hoje no mundo, principalmente em lugares mais afastados dos grandes centros urbanos, como municípios do interior do Brasil.

O sistema 2G também visava principalmente o mercado de voz, mas ao contrário do 1G, que utilizava a modulação analógica, o 2G utiliza a modulação digital (GHOSH; ZHANG; MUHAMED, 2010). Dentre as várias melhorias no desempenho do sistema, podemos citar:

- O uso de codificadores de áudio digitais para possibilitar maior eficiência ao espectro de rádio.

- Multiplexação dos usuários no mesmo canal de frequência, através de técnicas de multiplexação por divisão de tempo ou divisão de código, possibilitando o aumento de usuários na rede.
- Criptografia simples para fornecer medidas de segurança contraespionagem e antifraude, que eram uma fonte de preocupação com os sistemas analógicos de primeira geração.

Entre as diversas tecnologias de multiplexação utilizadas no 2G, uma delas foi o *Time Division Multiple Access* (TDMA). O TDMA é uma tecnologia cujo funcionamento permite que o canal de comunicação seja dividido em *slots*, isto é, em intervalos de tempo alternados (ALECRIM, 2010).

O TDMA geralmente consegue suportar até três vezes mais conexões do que as tecnologias analógicas, utilizando a mesma quantidade de canais. Isso é possível porque o TDMA é uma tecnologia digital, os dados da comunicação são comprimidos, permitindo que ocupem apenas um terço da capacidade do canal. Com isso, os outros dois terços podem ser aproveitados para outras chamadas (ALECRIM, 2010).

Essa breve noção sobre o TDMA é importante devido ao seu uso no *Global System for Mobile Communications* (GSM), uma vez que foi a tecnologia que mais se destacou no 2G. O GSM foi criado na *Conference Europeenne des Postes et Telecommunications* (CEPT), cuja missão foi desenvolver um sistema único de comunicação por rádio frequência para toda a Europa, tendo em vista que o mercado celular europeu estava fragmentado devido à utilização de uma variedade de sistemas incompatíveis: países escandinavos utilizavam o NMT-400 e o NMT-900 (variação do NMT-400), a Alemanha tinha o *Funktelefonnetz-C-450*, o Reino Unido utilizava o *Total Access Communication System* (TACS) e a França utilizava a *Radiocom* (GHOSH; ZHANG; MUHAMED, 2010).

Em 1989, o *European Telecommunications Standards Institute* (ETSI) foi o responsável pelo desenvolvimento do padrão GSM e pela sua primeira versão, chamada GSM fase I. O padrão foi lançado no ano seguinte, em 1990. Pouco tempo depois, várias operadoras da Europa implantaram o GSM. Essa tecnologia também se espalhou pelo mundo alcançando 220 países, e com essa ampla adoção, o GSM se tornou a tecnologia que possibilitou o *roaming* internacional na Europa (GHOSH; ZHANG; MUHAMED, 2010).

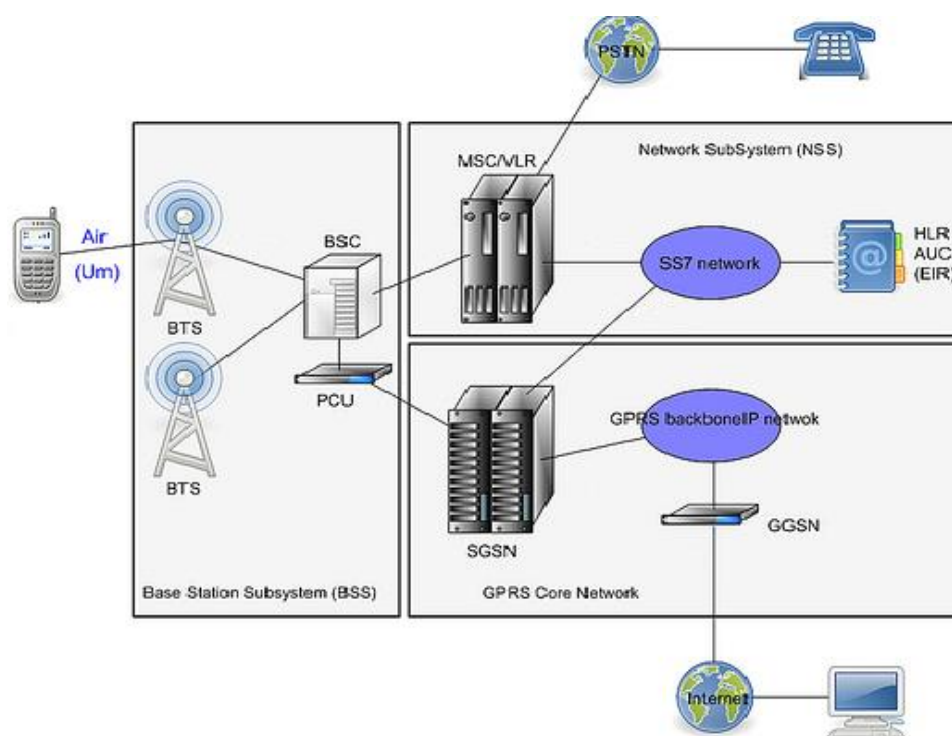
Uma característica bastante interessante da tecnologia GSM é o uso do dispositivo *Subscriber Identity Module* (SIM), conhecido no Brasil simplesmente como "chip" ou como "cartão SIM". Este minúsculo dispositivo armazena informações referentes ao assinante como, por exemplo, o perfil do usuário, número telefônico, serviços contratados, entre outros. Outro recurso de segurança que surgiu no GSM e que é utilizado até hoje é o *International Mobile Equipment Identify* (IMEI): trata-se de uma sequência numérica exclusiva para identificar cada dispositivo móvel, sendo definida pelo fabricante (ALECRIM, 2010).

2.3.1 Avanço da Segunda Geração de Telefonia Móvel

Para que a comunicação com serviços baseados em pacotes fosse realizada, o ETSI introduziu o *General Packet Radio Service* (GPRS). Essa tecnologia tinha a capacidade de se comunicar com os protocolos da Internet. O GPRS também utiliza múltiplos *slots*, mas como trabalha com comutação de pacotes, esses *slots* permanecem ativos apenas quando há envio ou recebimento de dados (GHOSH; ZHANG; MUHAMED, 2010). Uma curiosidade: o GPRS teoricamente poderia alcançar 171,2 Kb/s, mas dificilmente ultrapassava 80 Kb/s. As velocidades teóricas são testadas em ambientes favoráveis e sem ruídos. Alguns exemplos de problemas que ocorrem e inviabilizando taxas teóricas são superlotação de ancoragem de dispositivos em uma mesma ERB, causas naturais como chuva intensa, edifícios, distância do dispositivo e a ERB (ALECRIM, 2010).

A última evolução da tecnologia 2G antes da terceira geração foi o *Enhanced Data Rates for GSM Evolution* (EDGE). Essa tecnologia também é utilizada até hoje e possui características que se assemelham bastante ao GPRS. A grande diferença é a sua velocidade máxima teórica de 473,6 Kb/s, mas que normalmente não ultrapassa 380 Kb/s (ALECRIM, 2010).

Figura 3 - Arquitetura GSM e GPRS



Fonte: Tutorials Point (2012, *online*)

A figura 3 mostra como o sistema GSM, que utiliza apenas comutação por circuito, foi atualizado para o GPRS. Desta forma, a arquitetura passou a realizar comutação por pacotes para o transporte de aplicações e serviços de dados. O GPRS introduziu basicamente dois elementos à rede: o *Serving GPRS Support Node* (SGSN), que é o elemento de controle da rede, e o *Gateway GPRS Support Node* (GGSN). Este por sua vez fornece a funcionalidade de roteador de pacotes IP e conecta à rede GPRS a Internet e outras redes IP. A evolução dessa nova arquitetura de rede é a base para as futuras redes móveis de altas taxas de transmissão.

2.4 TERCEIRA GERAÇÃO DE TELEFONIA MÓVEL (3G)

A segunda geração proporcionou um aumento significativo na capacidade de voz, aprimorando a qualidade desse serviço, e iniciou o suporte para aplicativos de dados. Mas o paradigma de comutação de circuitos com base no qual esta tecnologia foi implementada criou sistemas 2G muito ineficientes para dados (GHOSH; ZHANG; MUHAMED, 2010).

A terceira geração de telefonia móvel (3G) veio com a proposta de que os usuários poderiam ter acesso à Internet móvel com qualidade similar às conexões fixas de banda larga da época. Desta forma os usuários poderiam aproveitar recursos como *streaming* de vídeo, aplicações de áudio, mensagens multimídia, entre outros.

O trabalho sobre o 3G começou no início da década de 1990, quando o ITU iniciou propostas para a criação do sistema 3G (conhecida como IMT-2000). O objetivo do ITU era criar uma especificação globalmente harmonizada para comunicação móvel que facilitasse o *roaming* internacional e a redução de custos. O ITU apresentou os seguintes requisitos de *throughput* como critério para o IMT-2000 (GHOSH; ZHANG; MUHAMED, 2010):

- 2Mbps em ambientes fixos ou em edifícios;
- 384kbps para pedestres ou perímetros urbanos;
- 144kbps para veículos em movimento;

Além do aumento na taxa de transferência de dados, os sistemas 3G fornecem um controle via qualidade de serviço (QoS), fornecendo uma melhor gestão da rede e priorizando determinados tráfegos de dados. O QoS são parâmetros definidos em termos de largura de banda, latência e *jitter*, de forma que a aplicação possa obter prioridade na rede, proporcionando uma melhor experiência do usuário com a tecnologia (CISCO, 2006).

O marco que deu início à tecnologia 3G oficialmente foi o desenvolvimento da tecnologia *Universal Mobile Telephone Service* (UMTS). O UMTS foi desenvolvido originalmente pelo ETSI como um sistema 3G na IMT-2000, baseado na evolução do GSM. O projeto foi continuado pelo *Third Generation Partnership Project* (3GPP) em 1998. O 3GPP é um grupo formado em colaboração entre sete órgãos padronizadores de tecnologia móvel em todo o mundo (GHOSH; ZHANG; MUHAMED, 2010).

O 3GPP completou e publicou o padrão 3G UMTS em 1999, e esse padrão é frequentemente chamado de UMTS 99. O *UMTS Release 99*, foi amplamente implantado em todo o mundo e obteve grande sucesso.

O UMTS manteve a arquitetura do GSM/GPRS. Todavia, a interface aérea do 3G, chamada *Wide-Band Code Division Multiple Access* (W-CDMA) foi considerada uma mudança radical em relação à interface aérea do 2G.

O CDMA realiza uma propagação do espectro de rádio diretamente. Para realizar isto, os dados do usuário são multiplicados por códigos pseudoaleatórios que fornecem uma canalização, sincronização e codificação. O sistema opera em uma largura de banda maior de 5MHz, capaz de suportar mais de 100 chamadas de voz simultâneas e fornecer melhores taxas de transferência de dados. Porém, mesmo assim, o WCDMA alcançou taxas de transferência de somente 384 Kbps (GHOSH; ZHANG; MUHAMED, 2010).

Visando-se melhorar o 3G, com mais velocidade e capacidade para Internet, foram feitos aprimoramentos pelo 3GPP. O *High Speed Packet Data Access* (HSPA) foi uma atualização da tecnologia WCDMA com foco de aumentar o desempenho na transferência de pacotes de dados. A introdução foi feita em etapas: a primeira foi o *High Speed Downlink Packet Access* (HSDPA), introduzido no *Release 5* (3GPP, 2017).

Utilizando a tecnologia HSDPA, o UMTS suporta taxas de até 14,4 Mbps, de acordo com o 3GPP. Naturalmente, a velocidade real varia de acordo com a qualidade do sinal e o número de usuários conectados à mesma NodeB, mas mesmo com condições adversas, as taxas sempre são mais altas que o WCDMA (3GPP, 2017).

Após a introdução do HSDPA, o *High Speed Uplink Packet Access* (HSUPA) foi introduzido no *Release 6*. O HSUPA é um padrão atualizado, que complementa as melhores taxas do HSDPA. As principais melhorias são as relacionadas às taxas de *upload* que chegam até 5.8 Mbps, de acordo com o 3GPP. Por ser apenas uma extensão do UMTS e não um novo padrão 3G, os investimentos necessários para migrar as redes são relativamente pequenos, tendo em vista que é preciso apenas substituir alguns equipamentos nas torres e redimensionar a estrutura de roteamento, sem exigir o licenciamento de novas faixas de frequência ou substituição de antenas (3GPP, 2017).

É importante enfatizar que o HSDPA e o HSUPA são dois padrões complementares e não concorrentes. O HSDPA melhora as taxas de *download* em relação ao WCDMA, enquanto o HSUPA melhora as taxas de *upload*. De acordo com os equipamentos usados, as operadoras podem suportar ambas as atualizações, oferecendo tanto *download* quanto *upload* mais rápidos, ou suportar apenas o HSDPA (3GPP, 2017).

Segundo o 3GPP, o HSDPA e o HSUPA foram combinados e foram adicionadas algumas melhorias, resultando na evolução do HSPA, também conhecido com HSPA+ ou HSPA evoluído. Essa atualização foi feita no *Release 7*. Com o HSPA+, as taxas de *download* chegaram a até 21Mbps e as de *upload* chegaram a até 11Mbps (3GPP, 2017).

2.5 QUARTA GERAÇÃO DE TELEFONIA MÓVEL (4G)

O crescimento da Internet na última década (2000-2010) é o principal motivo para a criação da quarta geração de telefonia móvel (4G). A Internet atualmente oferece grandes variedades de aplicações e se tornou o meio pelo qual se busca informação, comunicação e entretenimento.

O surgimento da quarta geração ocorreu em 2005, quando o 3GPP documentou a TR25.913. Um *Technical Report* (TR) do tipo TR.XX.9XX são documentos contendo o início de uma série de padronizações sobre determinada tecnologia. Dentro da TR25.913 havia metas de capacidades, complexidades, implementação e arquitetura da tecnologia 4G (TR25.913). Os principais objetivos que estão contidos nessa especificação técnica são:

- Taxa de transferência de dados significativamente aumentada, sendo de 100 Mbps para *Downlink* e 50 Mbps para *Uplink*;
- Um espectro de rádio frequência com melhor eficiência;
- Latência para acesso à rede em torno de 10ms;
- Largura de banda escalável;
- Flexibilidade no espectro de frequência de rádio chegando a até 20MHz;
- Latência para o controle do usuário em torno de 100ms;

O LTE amplia a experiência do usuário com a telefonia móvel, tendo-se em vista que as novas aplicações carecem de uma largura de banda maior. O LTE foi a primeira tecnologia móvel a fornecer suporte ao tráfego de pacotes IP com QoS de ponta a ponta, ou seja, do usuário até à aplicação (MAZZONI, 2014).

Uma preocupação do 3GPP para a criação do LTE foi a redução do custo de cada *megabyte*, sendo que houve crescimento significativo no uso de dados da rede, principalmente dos dados relacionados a aplicações de vídeo. O 3GPP reconheceu esse problema e transformou-o em um critério chave para o

desenvolvimento do LTE (MAZZONI, 2014). Abaixo estão alguns critérios que estão ligados diretamente à eficiência de custos:

- Capacidade de implantação no espectro existente e reutilização de *sites* e equipamento de transmissão;
- Interoperabilidade com sistemas padronizados pelo 3GPP, para permitir uma migração suave e econômica;
- Interoperabilidade com sistemas que não fazem parte do 3GPP;
- Uma arquitetura *flat* (explicação sobre arquitetura *flat* na sessão 2.6 deste trabalho) com menos componentes e protocolos de rede;
- Arquitetura totalmente IP;
- Um único núcleo de pacotes IP para voz e dados;
- Suporte para redes de transmissão baseadas em protocolos IP de baixo custo;
- eNodeB com menores requisitos de energia e espaço. Estes eNodeBs poderiam ser colocados dentro dos gabinetes existentes ou montados ao lado deles;
- Suporte para autoconfiguração e otimização a fim de reduzir custos de instalação e gerenciamento.

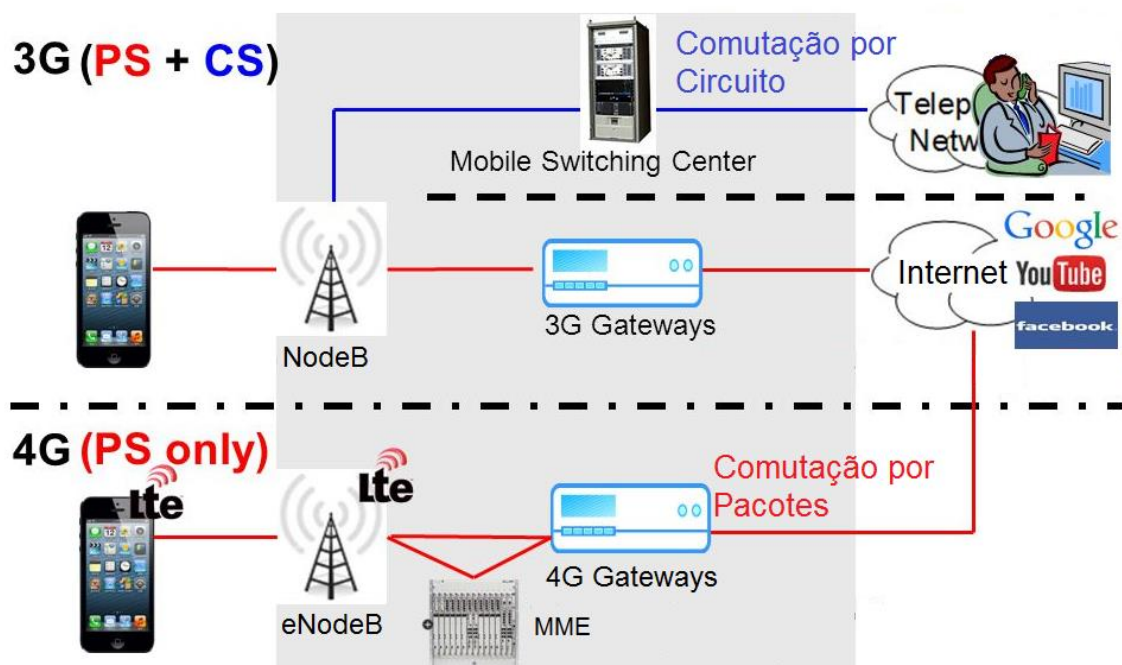
A tecnologia LTE foi especificada no *Release 8* do 3GPP e implantada comercialmente no final de 2009 no Japão. Desde então, os números de redes comerciais cresceram fortemente em todo o mundo. O LTE se tornou a tecnologia de desenvolvimento mais rápida no sistema móvel, mas somente em 2012 a Anatel ofereceu o leilão da faixa de 2,5MHz, possibilitando a implementação do LTE no Brasil (MAZZONI, 2014). Apesar da introdução tardia do LTE, a Anatel aponta que em 2017, 41% da população brasileira já utiliza 4G.

Na teoria, o LTE alcançou 300 Mbps de *downlink* e 75 Mbps de *Uplink* em testes realizados em laboratórios. A taxa padronizada pelo 3GPP foi de 100 Mbps de *downlink* e 50 Mbps de *uplink*, mas levando-se em consideração o número de usuários ancorados no mesmo eNodeB, ruídos na transmissão e as limitações da infraestrutura da rede de cada operadora, esses valores podem ser menores (MAZZONI, 2014).

Figura 4 ilustra a principal diferença entre o LTE e as tecnologias relacionadas com o 3G e 2G. O LTE opera somente com comutação por pacotes,

eliminando a comutação por circuito do sistema. Esta arquitetura foi projetada para oferecer o máximo de *throughput* para o usuário final e economizar custos.

Figura 4 - Comparação com arquitetura das redes 3G e 4G



Fonte: alterada de Hua Tu et al. (2015, online)

Outra característica da tecnologia LTE é o QoS de ponta a ponta. Na tecnologia GPRS, todos os pacotes dos usuários competem igualmente pelos recursos da rede e nenhuma garantia é fornecida sobre quando ou se os pacotes vão chegar ao destino. A tecnologia 3G já possuía QoS para gerenciar os *Backbones* e a rede interna da operadora de acordo com cada serviço. Mas somente no LTE essa funcionalidade das redes IP foi estendida para o usuário final, possibilitando assim a introdução de novos serviços que necessitam de requisitos específicos para a sua operação. Com o QoS de ponta a ponta do LTE, os recursos da rede tanto de rádio quanto do núcleo da rede são negociados garantindo uma experiência máxima do usuário na rede (CISCO,2006).

Finalizando os aspectos gerais sobre a tecnologia LTE, será exibida uma tabela com as taxas de transferência das tecnologias padronizadas pelo 3GPP. A tabela 2 relaciona cada uma das tecnologias padronizadas pelo 3GPP e os valores de velocidade de *download*, *upload* e latência.

Tabela 2 - Tecnologias padronizadas pelo 3GPP e suas taxas de transferência

Padronizações	3GPP Release	Velocidade de <i>Download</i>	Velocidade de <i>Upload</i>	Latência
GPRS	Release 97/99	40–80kbps	40–80kbps	600–700ms
EDGE	Release 4	237–474kbps	237kbps	350–450ms
UMTS (WCDMA)	Release 4	384kbps	384kbps	<200ms
HSDPA/UMTS	Release 5	14,4Mbps	2Mbps	<120ms
HSUPA	Release 6	14,4Mbps	5.4kbps	<100ms
HSPA+	Release 7 e 8	21Mbps	11Mbps	<80ms
LTE	Release 8	100Mbps	50Mbps	<10ms

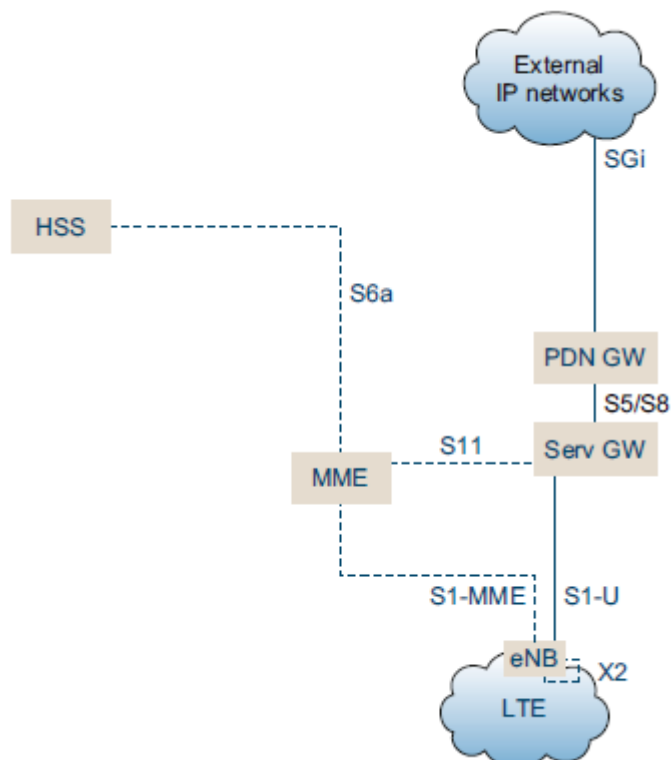
Fonte: Autor (2017)

2.6 ARQUITETURA DO EVOLVED PACKET SYSTEM

O *Evolved Packet System* (EPS) engloba o acesso de rádio frequência do LTE, o núcleo da rede e os terminais que compõem o sistema móvel geral, basicamente todo o sistema do LTE. O EPS é formado pelo *Evolved UTRAN* (ou E-UTRAN) e o *Evolved Packet Core* (EPC) e foi projetado para interoperar com redes padronizadas pelo 3GPP como, por exemplo a tecnologia 3G, e redes não-padronizadas pelo 3GPP como WLAN (OLSSON et al., 2009).

A figura 5 mostra a arquitetura mais básica do LTE.

Figura 5 - Arquitetura básica para o LTE



Fonte: Olsson et al. (2009, p. 39)

Antes de seguir com a descrição de como ocorre uma conexão fim-a-fim do EPS, vale a pena comentar sobre a definição de entidade de rede segundo o 3GPP. Uma entidade de rede é responsável por uma função única na arquitetura, mas isso não implica que os equipamentos físicos reais produzidos pelos fornecedores e implantados em redes comerciais devam corresponder a somente uma entidade. Neste trabalho, uma entidade é representada como um bloco, isto ilustrado na figura 5. O presente trabalho também pode utilizar a palavra nó, a mesma pode ser considerada um sinônimo da palavra entidade (OLSSON et al., 2009).

Fornecedores de equipamentos de telecomunicações podem produzir um equipamento com uma única função, ou podem escolher combinar diferentes entidades no mesmo produto/*hardware*. Um exemplo são as entidades como o *Serving Gateway* (SGW) e *Packet Data Network Gateway* (PDN GW) padronizadas pelo 3GPP. Essas entidades podem ser implementadas pelos fornecedores em somente um *hardware* ou em equipamentos separados. Essa

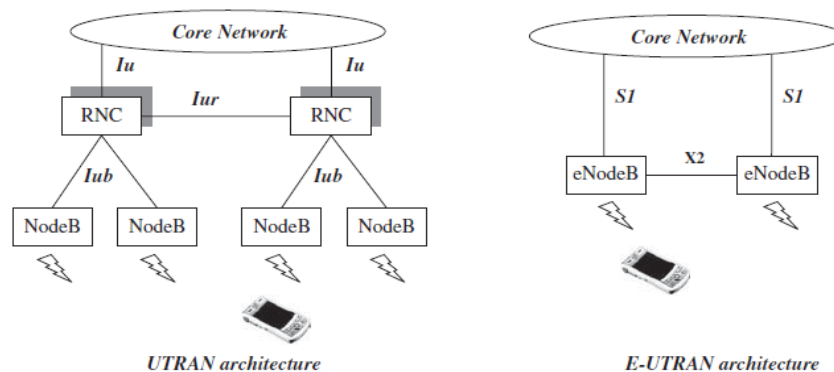
decisão depende da arquitetura de rede planejada pela operadora e dos serviços ofertados para o usuário final (OLSSON et al., 2009).

Ainda sobre conceitos utilizados pelo 3GPP, o termo ponto de referência é utilizado para denotar uma associação entre duas entidades. Existem algumas diferenças na definição formal de ponto de referência e interface, mas, para o propósito deste trabalho, as discrepâncias não têm implicações práticas. Portanto, interface será o termo utilizado neste trabalho (OLSSON et al., 2009).

As interfaces no 3GPP geralmente possuem uma letra como prefixo e outras letras como sufixos adicionais. No GPRS por exemplo, a maioria das interfaces começa com a letra 'G', enquanto no EPC a maioria das interfaces começa com a letra 'S'. (OLSSON et al., 2009).

Para retomar o foco na arquitetura. Será comentado os dois principais motivos que nortearam a criação do *design* da arquitetura do EPS. O primeiro motivo era o desejo de otimizar o tratamento do tráfego de dados do usuário, através de uma arquitetura "flat". Uma arquitetura *flat* neste contexto significa o menor número possível de nós envolvidos no processamento do tráfego de dados do usuário. A arquitetura *flat* permite uma melhor eficiência na infraestrutura em que os dados são trafegados. Isto se torna vantajoso tendo em vista que o volume de dados móveis está crescendo rapidamente e deve crescer ainda mais rápido no futuro com a introdução de novos serviços. A figura 6 mostra a diferença entre as tecnologias de acesso do 3G e do 4G e a extinção do *Rádio Network Controller* (RNC), reduzindo-se o número de elementos da rede.

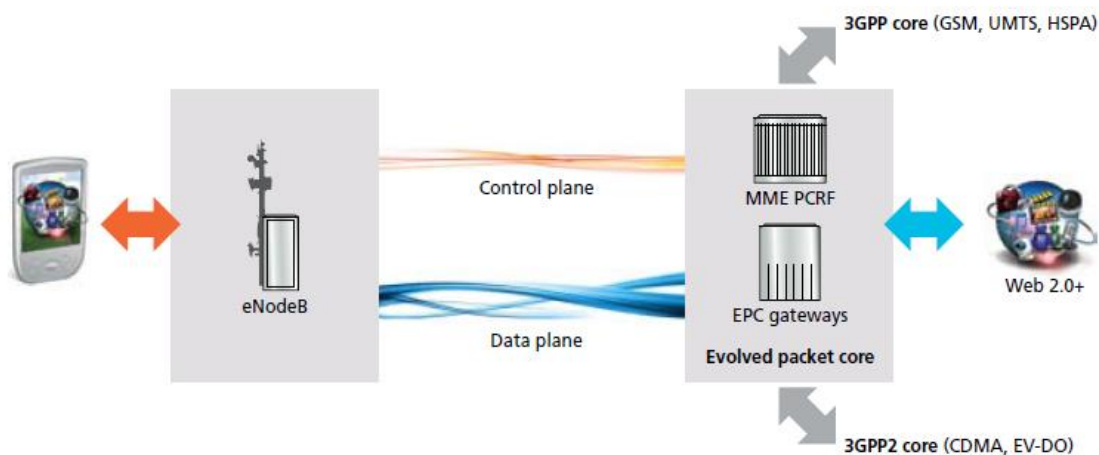
Figura 6 - Mudança na arquitetura UTRAN e Evolved UTRAN



Fonte: Lescuyer e Lucidarme (2008, p. 38)

O segundo motivo era separar a sinalização de controle, conhecidas como plano de controle e o tráfego de dados do usuário, conhecido como plano do usuário. Isso foi motivado por vários fatores. Um deles foi a necessidade de escalonar de forma independente o plano do usuário e o plano de controle. Isto ocorre devido ao plano de controle tender a ser dimensionado de acordo com o número de usuários. Enquanto o plano do usuário é mais dependente das taxas de dados e volume de dados oriundos de novos serviços ou aplicativos. Na figura 5, o plano do usuário é representado por uma linha contínua e a interface usada é o S1-U, enquanto o plano de controle é representado por uma linha tracejada e a interface usada é a S1-MME (OLSSON et al., 2009).

Figura 7 - Divisão entre plano de dados e plano de controle



Fonte: Acatel Lucent (2009, p. 7)

A figura 7 demonstra a divisão do plano de usuário e o plano de controle, também mostra como os *gateways* do EPC operam com os dados do usuário, enquanto o *Mobility Management Entity* (MME) e o *Policy Charging Rules Function* (PCRF) operam somente com sinalização na rede.

Examinando-se a figura 5, o primeiro elemento que se destaca é o ponto de acesso do LTE, conhecido como eNodeB. O eNodeB fornece a interface de rádio e executa o gerenciamento de recursos de rádio para LTE. O eNodeB também suporta o uso de cabeçalhos IP e a criptografia dos dados do usuário. Os eNodeBs estão interligados entre si através de uma interface chamada X2. Esta interface tem vários usos como, por exemplo, troca de sinalização entre os eNodeB no momento da execução do *handover* (OLSSON et al., 2009).

Os eNodeBs estão conectados ao EPC através da interface S1. Esta interface é dividida em plano do usuário e plano de controle. A interface do plano de controle é referida como S1-MME e termina no MME. A interface S1-U, entretanto, termina no SGW e lida com o tráfego do usuário. Em uma operadora de tamanho razoável pode haver milhares de eNodeBs (OLSSON et al., 2009).

Os eNodeBs estão conectados a pelo menos um MME. O MME trabalha com toda a sinalização do plano de controle da rede LTE. Já os dados dos usuários são tratados pelo *Serving Gateway* (SGW) e pelo *Packet Data Network Gateway* (PDN GW) (OLSSON et al., 2009).

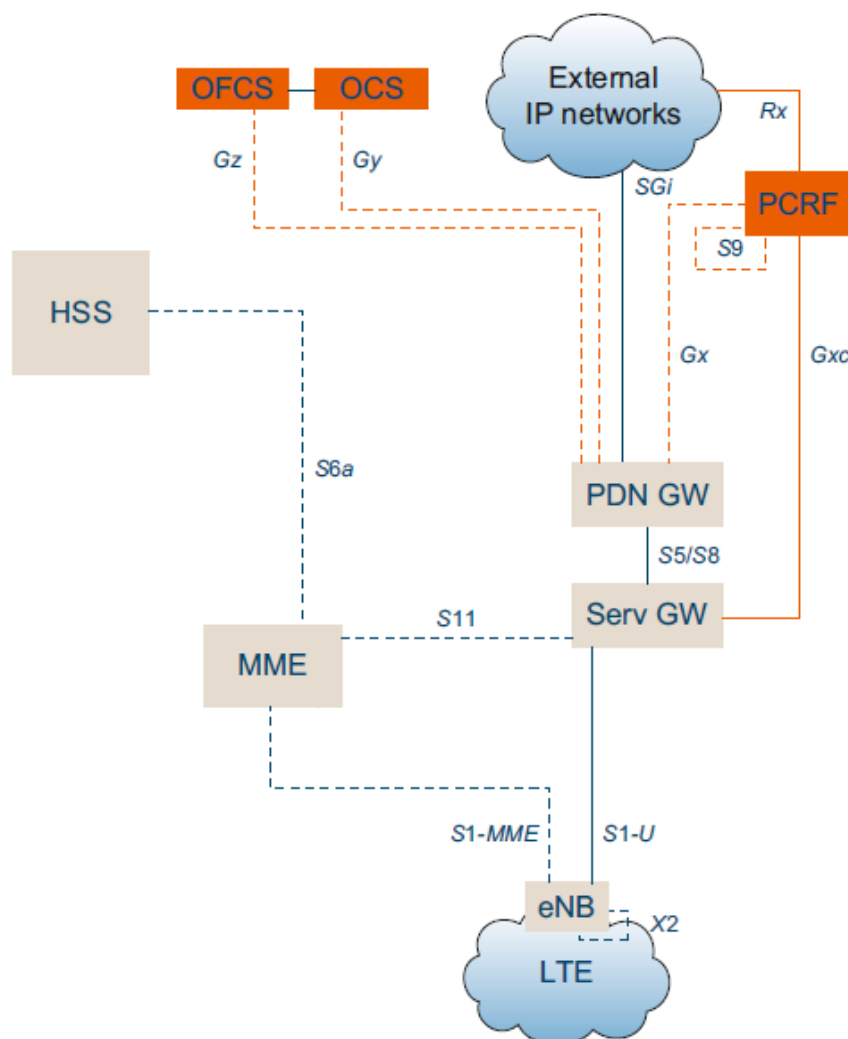
O SGW recebe os dados dos eNodeBs através da interface S1-U. O SGW também trabalha como um ponto de âncora para conexões dentro da rede LTE, auxiliando os eNodeBs quando ocorre um *handover*. Também age como um ponto de âncora, opcionalmente, para outras redes do 3GPP (OLSSON et al., 2009).

O PDN GW é o ponto de interconexão com redes IP externas da operadora através da interface SGi. O PDN GW inclui a funcionalidade para alocação de endereços IP, cobrança e filtragem de pacotes. O PDN GW também tem um papel fundamental no suporte de QoS para o usuário final (OLSSON et al., 2009).

O SGW e o PDN GW ficam conectados através de uma interface chamada S5, caso o usuário esteja conectado à sua rede de origem, mas caso o usuário esteja em *roaming*, a interface utilizada será a S8 (OLSSON et al., 2009).

expandindo a arquitetura básica do LTE, serão introduzidos três novos elementos e suas interfaces associadas a cada um, conforme a figura 8: o *Policy and Charging Rules Function* (PCRF), o *Online Charging System* (OCS) e o *Offline Charging System* (OFCS). Para entender como estes elementos operam, vai ser conceituado o controle de fluxo IP.

Figura 8 - Adicionando os nós PCRF/OCS/OFCS à arquitetura básica do EPC



Fonte: OLSSON et al. (2009, p. 43)

O fluxo IP (conhecido como *IP Flow*) são pacotes IP fluindo através de uma rede. Esses dados estão associados a uma aplicação específica em uso como, por exemplo, uma navegação na web ou *streaming* de vídeo. O PCRF é uma entidade chave na arquitetura do EPC. O PCRF foi projetado para permitir o controle da rede baseado em fluxo IP. O PCRF é o orquestrador da rede LTE, assim, usando o conceito de *IP flow*, ele pode controlar o tráfego de dados da

rede LTE aplicando diferentes políticas de tratamento para usuários e aplicações. No contexto da arquitetura do 3GPP, políticas são regras de tratamento baseadas em um fluxo IP específico como, por exemplo, como os dados devem ser cobrados ou qual QoS deve ser atribuído a este serviço. Para que os dados sejam cobrados e que haja um controle de políticas, os fluxos IP são classificados no PDN GW. Esses dados são operados em tempo real (OLSSON et al., 2009).

O OFCS e o OCS são entidades lógicas que interagem com o PDN GW (através das interfaces Gz e Gy, respectivamente) e suportam vários recursos relacionados ao carregamento de crédito e serviços aos usuários finais com base em diferentes parâmetros, como tempo, volume, evento, etc.

A rede LTE pode oferecer banda larga móvel com maior capacidade de dados e menor latência. No entanto, não existe um domínio de voz por comutação de circuitos no LTE. A indústria de telefonia móvel adotou uma solução de chamadas de voz e vídeo baseada em IP que opera globalmente para a LTE, conhecida como *Voice over LTE* (VoLTE) (ERICSSON, 2014).

Existem três opções para prestação de serviço de voz na quarta geração de telefonia móvel: *Circuit Switch Fallback* (CSFB), *Voice over LTE via Generic Access* (VoLGA) e o VoLTE. No entanto, este trabalho irá focar apenas no VoLTE. (OLSSON et al., 2009)

VoLTE representa uma nova experiência na utilização de serviços de voz para o usuário, incluindo serviços como voz HD, vídeo-chamada, conferência HD, mensagens instantâneas e gerenciamento de contato (ERICSSON, 2014).

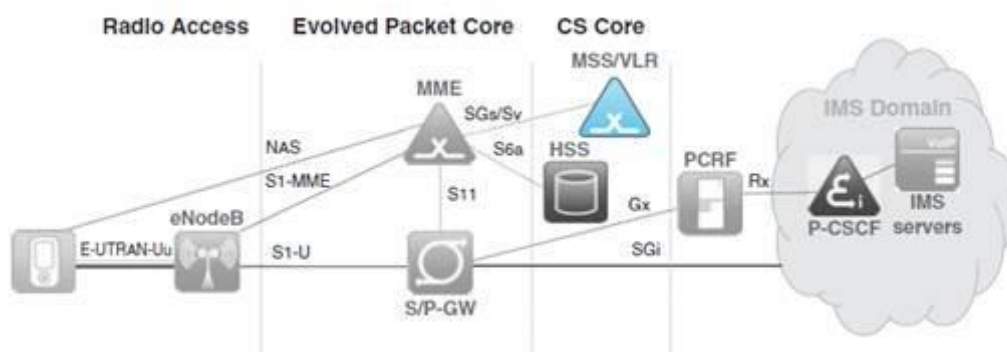
Nas tecnologias anteriores padronizadas pelo 3GPP, o núcleo da rede era responsável pela mobilidade, conectividade do usuário e as funcionalidades do controle de chamada. Quando se trata de VoLTE, o EPC é responsável pela mobilidade e conectividade, enquanto o *IP Multimedia Subsystem* (IMS) fornece o controle de chamadas (TELECO, 2014).

Os elementos existentes no EPC terão as seguintes funções na implementação do VoLTE (TAKAKI; BAZZO, 2012):

- PCRF – responsável pelo controle direto dos recursos, ou seja, disponibiliza a qualidade do serviço de acordo com o perfil do usuário.

- MME – responsável pelo controle das funções relacionadas ao gerenciamento de recursos, de capacidade e de conexão.
- SGW – responsável pela mobilidade e retenção de informações quando o dispositivo do usuário está inativo.
- PDN GW – responsável pela alocação do endereço IP do dispositivo, garantia de QoS de acordo com o serviço utilizado e tarifação de acordo com regras do PCRF. Além disso, filtra os pacotes do usuário de acordo com as regras de QoS para os recursos existentes.

Figura 9 - Arquitetura EPS para VoLTE



Fonte: Poikselka et al. (2012, *online*)

A figura 9 apresenta uma arquitetura para a implantação do VoLTE. O objetivo deste trabalho é fornecer um breve entendimento do serviço, sem detalhar todos os procedimentos e toda arquitetura utilizada no VoLTE.

O EPC e suas interfaces são primordiais na funcionalidade do VoLTE, e juntamente com o sistema IMS, que é uma arquitetura que possibilita serviços multimídias em IP, podem prover qualquer serviço que é oferecido pela telefonia móvel atualmente, incluindo chamadas de voz e SMS (TELECO, 2014).

Nesse contexto de serviços, é importante ressaltar que fornecer conectividade ao usuário não se trata apenas de obter um endereço IP. Trata-se também de como esses dados irão ser transportados na rede, de forma que o usuário tenha uma boa experiência acessando os variados tipos de serviços oferecidos pela rede. Conceituar o que é um *bearer* é importante para entender

como os diferentes tipos de dados com diferentes prioridades trafegam na rede LTE (TELECO, 2014).

O *bearer* fornece um canal de transporte lógico dentro da rede EPS entre o dispositivo e o PDN GW. Cada *bearer* está vinculado a um conjunto de parâmetros de QoS. A função do QoS é descrever as propriedades do canal como, por exemplo, taxas de bits, atraso e taxa de erro de bits (OLSSON et al., 2009). Quando um dispositivo do usuário é ligado e se conecta à rede, a rede fornece um *default bearer*, ou seja, um canal lógico padrão onde os dados irão trafegar.

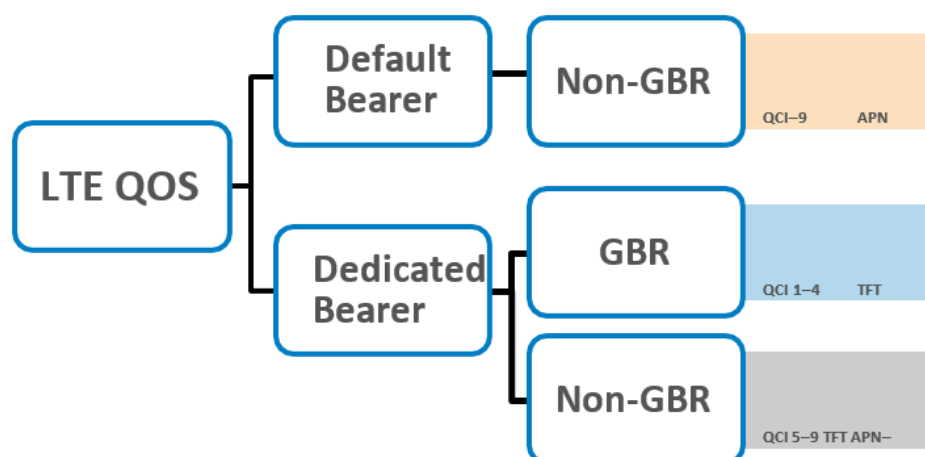
Dedicated bearers são *bearers* para serviços que necessitem de alto desempenho da rede como serviços de voz e vídeo. O *dedicated bearer* pode ser ativado sob demanda e são configurados apenas quando são necessários (OLSSON et al., 2009).

Em termos simples, o *dedicated bearer* fornece um túnel dedicado a um ou mais tráfegos específicos. Esses *bearers* são os portadores secundários que são criados em cima do *default bearer* existente, compartilhando o endereço IP previamente estabelecido pelo *default bearer* (BEC TECHNOLOGIES, 2010).

Os *dedicated bearers* são utilizados no VoLTE, no entanto para compreender como funcionam os *dedicated* e *default bearers* nesse serviço, precisa-se entender os parâmetros de QoS que são utilizados no mesmo. Estes parâmetros são configurados no *dedicated bearer*, e a entidade que ordena a configuração destes parâmetros é o PCRF.

Na tecnologia LTE, os parâmetros de QoS são divididos em dois: com garantia de taxa de transferência de dados (*guaranteed bit rate*) e de taxa de transferência de dados não garantida (*non-guaranteed bit rate*). O *default bearer*, por ser um canal lógico de propósito geral, não possui a garantia de transferência de bit (BEC TECHNOLOGIES, 2010) como mostrado na figura 10.

Figura 10 - Parâmetros de QoS na rede LTE



Fonte: BEC Technologies (2010, p. 1)

O QoS utilizado na rede LTE possui uma divisão de acordo com uma classe de serviços, esta classe de serviço é chamada de QCI (*QoS class of identifier*). Essas classes são divididas de 1 a 9. A figura 11 mostra o QCI correspondente a cada serviço, a sua prioridade na rede, a latência e se possui taxa garantida de bits.

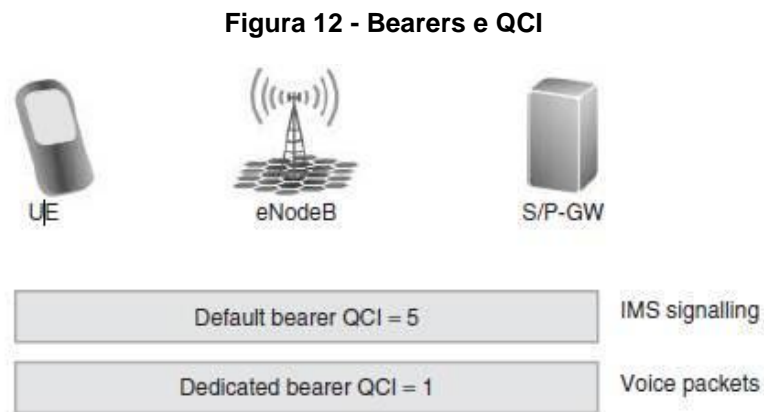
Figura 11 - Classes de QoS

QCI	RESOURCE TYPE	PRIORITY	PACKET DELAY BUDGET (MS)	PACKET ERROR LOSS RATE	EXAMPLE SERVICES
1	GBR	2	100	10^{-2}	Conversational voice
2	GBR	4	150	10^{-3}	Conversational video (live streaming)
3	GBR	5	300	10^{-6}	Non-conversational video (buffered streaming)
4	GBR	3	50	10^{-3}	Real-time gaming
5	Non-GBR	1	100	10^{-6}	IMS signaling
6	Non-GBR	7	100	10^{-3}	Voice, video (live streaming), interactive gaming
7	Non-GBR	6	300	10^{-6}	Video (buffered streaming)
8	Non-GBR	8	300	10^{-6}	TCP-based (for example, WWW, e-mail), chat, FTP, p2p file sharing, progressive video and others
9	Non-GBR	9	300	10^{-6}	

Fonte: Sridhar (2009, p. 18)

Cada classe possui um parâmetro de latência e um parâmetro de perda de pacotes. No caso do VoLTE, o número da classe em que este serviço se encontra é o QCI 1 como mostrado na figura 11, sendo assim considerado de grande prioridade na rede LTE (BEC TECHNOLOGIES, 2010).

A figura 12 ilustra os *dedicated* e *default bearers* em uma chamada VoLTE com seus respectivos QCI's utilizados. Quando ocorre uma chamada de VoLTE, a sinalização na rede é enviada pelo *default bearer*, essa sinalização possui QCI 5. É interessante observar que a sinalização IMS tem prioridade máxima na rede. Após essa sinalização ser enviada, a rede cria um *dedicated bearer* que possui QCI 1 com garantia de bits para o transporte dos dados da chamada. E, assim, a chamada é estabelecida.



Fonte: Poikselka et al. (2012, *online*)

3. EVOLVED PACKET CORE (EPC)

No capítulo anterior, foi feita uma análise sobre as tecnologias de telefonia móvel com ênfase no LTE. Neste capítulo serão descritas mais detalhadamente as diferentes entidades de rede que fazem parte do núcleo do LTE, o *Evolved Packet Core* (EPC).

Essa parte da rede foi padronizada pelo 3GPP na TS 23.002, especificação técnica que mostra uma visão geral sobre a arquitetura e descreve todos os componentes do EPC, que são:

- *Home Subscriber Server*;
- *Mobility Management Entity*;
- *Serving Gateway*;
- *Packet Data Network Gateway*.

Algumas das funções do núcleo da rede LTE são:

- Gestão de mobilidade – o núcleo possui a capacidade de rastrear dispositivo conforme eles se movem entre os *eNodeB*;
- Gerenciamento de sessões – configuração do *default bearer*;
- Gerenciar o fluxo de dados de determinado serviço ou aplicativo;
- Segurança e privacidade - autenticação, criptografia e privacidade do usuário.

3.1 HOME SUBSCRIBER SERVER (HSS)

O HSS é uma base de dados que armazena as informações de cada usuário na rede. Essas informações serão utilizadas pelas entidades que processam as sessões e chamadas como, por exemplo, o MME (3GPP TS 23.002, 2014).

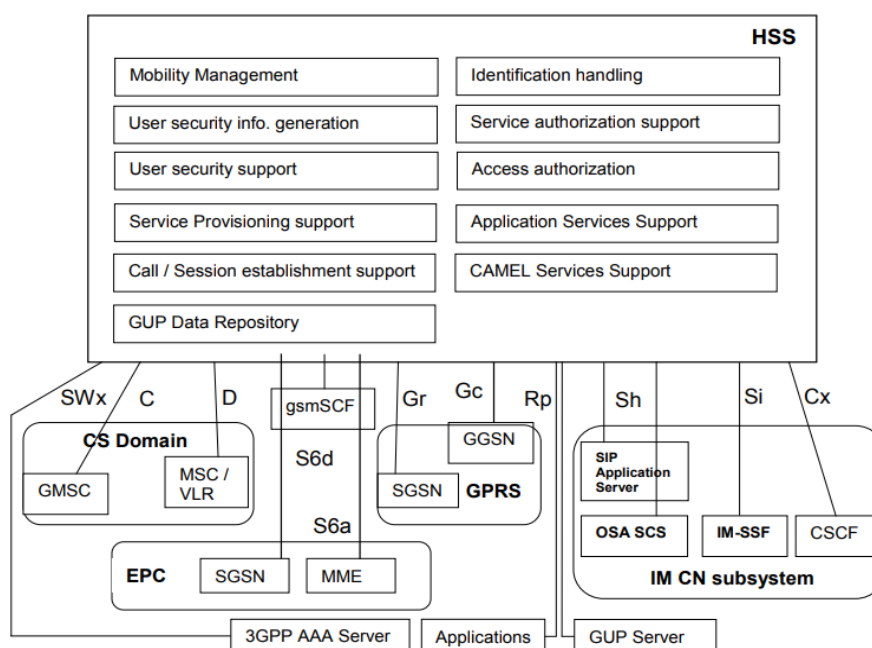
O HSS é responsável por manter as seguintes informações relacionadas ao usuário:

- Identificação do usuário: numeração;
- Informações de segurança do usuário: informações que são usadas para o acesso à rede, autenticação e autorização;
- Informações de localização;
- Informações do perfil do usuário, como por exemplo, parâmetros de QoS de acordo com o serviço utilizado.

O HSS gera informações de segurança do usuário para autenticação, verificação da integridade da comunicação e criptografia. Com base nessas informações, o HSS é responsável por suportar as entidades de controle de chamadas e gerenciamento de sessões dos diferentes domínios e subsistemas. Uma operadora pode conter um ou vários HSS's, este número depende da quantidade de assinantes, capacidade do equipamento e organização da rede.

A figura 13 ilustra as funções, interfaces, subsistemas e tecnologias padronizadas pelo 3GPP que fazem parte do HSS.

Figura 13 - Funções do HSS e suas interfaces



Fonte: 3GPP TS 123 002 (2014, p. 22)

Algumas das funcionalidades do HSS são:

- Gestão de Mobilidade: o HSS suporta a mobilidade do usuário através do domínio da comutação por pacotes e para o subsistema IMS;
- Suporte para o estabelecimento de chamada ou conexão IP: o HSS suporta os procedimentos de estabelecimento de chamada/sessão no domínio de comutação por pacotes e no subsistema IMS;
- Geração de informação de segurança do usuário;

- Suporte de segurança do usuário: o HSS suporta os procedimentos de autenticação;
- Identificação do usuário: o HSS fornece todas as informações que determinam de forma exclusiva o assinante no sistema;
- Suporte de autorização de serviço.

Quando o usuário se conecta à rede LTE, o MME se comunica com o HSS de origem do usuário através da interface S6a, fornecendo o *International Mobile Subscriber Identity* (IMSI) para o HSS. O IMSI é a identidade do assinante na rede. O HSS de origem do usuário por sua vez mantém alocado o IMSI do usuário permanentemente. Logo após receber o IMSI, o HSS realiza o procedimento de autenticação do usuário. Quando isso ocorre, o HSS fornece para toda a rede chaves criptografadas, garantido a segurança, confidencialidade, integridade do usuário. O HSS também fornece para a rede informações sobre serviços que o usuário pode realizar (OLSSON et al., 2009).

Quando o usuário se desloca entre as redes GSM / WCDMA e LTE, não deve haver informações inconsistentes. Por exemplo, o dispositivo se desloca da rede LTE para uma área que só possui cobertura GSM, os dados que existem no HSS e no *Home Location Register* (HLR), base de dados do 2G/3G, não podem ser diferentes na rede. Isso significa que o HLR e o HSS precisam compartilhar de um único conjunto de dados ou garantir que os dados sejam iguais através de outros meios, como uma conexão física entre os dois equipamentos. O 3GPP não especifica nenhuma solução detalhada para este problema. Os fornecedores em geral para solucionar este problema combinam as duas entidades em um único equipamento, deixando a rede mais otimizada (OLSSON et al., 2009).

3.2 MOBILE MANAGEMENT ENTITY (MME)

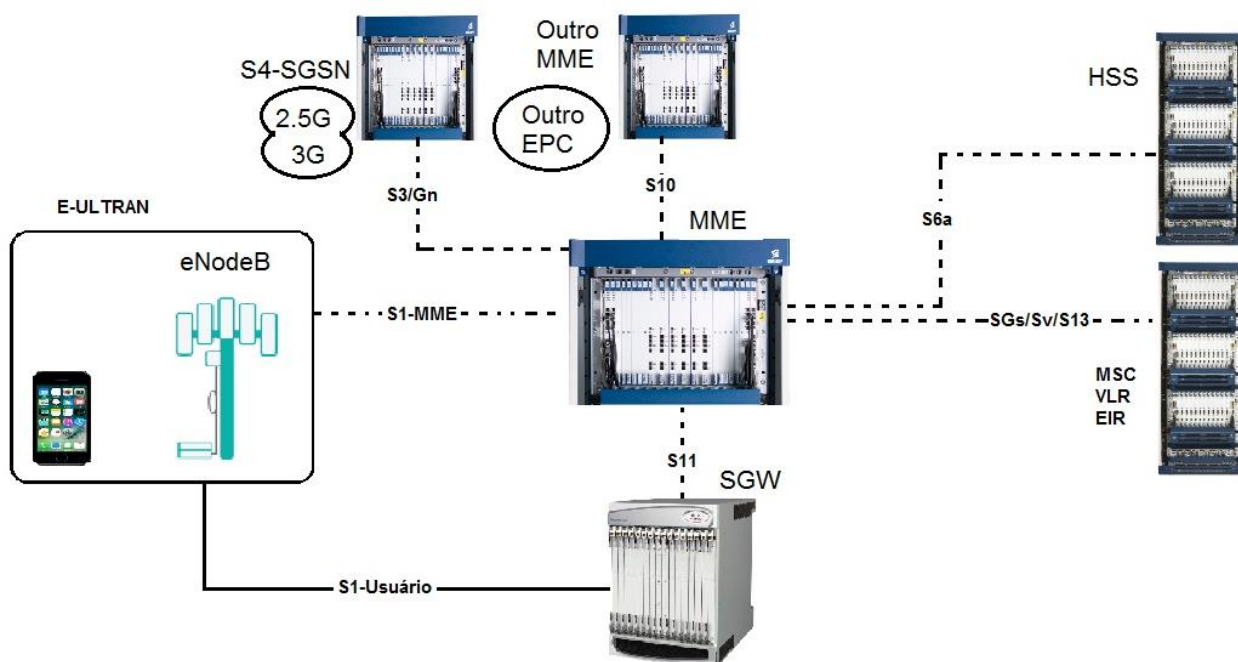
A próxima entidade do núcleo da rede LTE que será apresentada é o MME. O MME é a entidade que realiza o controle da rede e sinalização, ou seja, atribui recursos para o usuário e realiza a gestão da mobilidade para apoiar o rastreamento do usuário (*paging*, *roaming* e *handovers*). O 3GPP especifica

essas funções principalmente na TS 23.401 e na TS 23.002. De acordo com as padronizações do 3GPP, as funções e procedimentos do MME são:

- Sinalização com o protocolo *Non Access Stratum* (NAS);
- Sinalização de mobilidade com as tecnologias anteriores padronizadas pelo 3GPP (através da interface S3);
- Seleção do PDN GW e do SGW;
- Seleção do MME para casos de *handover*;
- Seleção do SGSN para casos de *handover* para tecnologias 2G/3G;
- *Roaming* através da interface S6a com o HSS;
- Autenticação do equipamento do usuário;
- Gerenciamento dos *dedicated/default bearers*;
- Gerenciamento da lista de *Tracking Areas* e atualização das mesmas;
- Interface de comunicação com *Mobile Switching Centre* (MSC) – para *paging* (o conceito de *paging* será abordado nesta seção).

A figura 15 ilustra o MME e todas as suas interfaces. Nem todas as entidades e interfaces serão descritas em detalhe, tendo em vista que este trabalho está direcionado para serviços de dados realizados via comutação por pacotes.

Figura 14 - MME e a comunicação com as entidades do EPS



Fonte: Autor (2017)

O MME possui um protocolo próprio para realizar o controle entre o equipamento do usuário e o MME. O protocolo *Non Access Stratum* (NAS) é o mais importante para o gerenciamento do EPS e foi padronizado pelo 3GPP na TS 24.301.

Segundo o 3GPP, o protocolo NAS promove o suporte à mobilidade do usuário na rede e os procedimentos para estabelecer conexão entre o dispositivo do usuário e o PDN GW. Outra função adicional do protocolo NAS é a segurança na rede. Através do protocolo NAS há uma garantia de troca de mensagens seguras. A interface utilizada para essa comunicação é a S1-MME, ilustrada na figura 14 (3GPP TS 24.301).

O Protocolo NAS é dividido em dois seguimentos: gestão de mobilidade do EPS e o gerenciamento de seções do EPS.

A gestão de mobilidade é responsável pela comunicação entre o dispositivo do usuário e a rede, informando o dispositivo sobre serviços específicos da rede. O dispositivo por sua vez pode informar a rede sobre sua capacidade (OLSSON et al., 2009). Os procedimentos de gestão de mobilidade do EPS são:

- *Attach* – É o procedimento de conexão do dispositivo do usuário à rede da operadora de telefonia móvel. O procedimento de *attach* é realizado para permitir a recepção de serviços da rede. O procedimento de *Attach* inclui um estabelecimento de um *default bearer*, garantindo a conectividade IP do usuário (OLSSON et al., 2009);
- *Detach* – É o procedimento de remoção dos *dedicated/default bearers*. O *detach* pode ocorrer em vários casos, por exemplo, quando o dispositivo do usuário está desligado. O *detach* também pode ser usado pela rede para remover os *dedicated/default bearers* de um dispositivo que parou de executar a atualização *paging*, porque está fora da cobertura do eNodeB (OLSSON et al., 2009);
- Autenticação – O MME realiza o procedimento de autenticação do equipamento do usuário. O mecanismo de autenticação é chamado de *Authentication and Key Agreement (AKA)*. AKA é o procedimento que leva o dispositivo e a rede para se autenticar mutuamente, fornecendo proteção, integridade e confidencialidade (CISCO, 2016). Essa autenticação mútua é baseada no fato de que a rede e o dispositivo do usuário têm acesso à mesma chave de segurança (IMSI).

Outra função do MME realizada através do procedimento de gestão de mobilidade do EPS é o gerenciamento de mobilidade do usuário. Em uma rede LTE é comum existir milhares de eNodeBs. O MME possui a função de realizar atualizações periódicas sobre a localização do dispositivo do usuário. Porém, acompanhar o equipamento do usuário sempre que o mesmo se movesse entre células diferentes não seria prático e causaria muita sinalização na rede. Assim como procurar o dispositivo em toda a rede para cada evento como, por exemplo, uma chamada recebida também não seria favorável. Assim, os eNodeBs são agrupados em Áreas de Registro (AS). As áreas de registros no EPS são chamadas de *Tracking Area (TA)*. O MME mantém uma lista com todas as TA's que o usuário se locomoveu, atualizando essa lista sempre que o usuário se locomove para uma TA diferente. (OLSSON et al., 2009).

Uma função de fundamental importância que também é realizada pelo MME é o *paging*. O *paging* é usado para localizar dispositivos ociosos e estabelecer uma conexão de sinalização. Este é o procedimento de atualização da localização do equipamento do usuário.

O gerenciamento de seções do EPS é responsável por gerenciar os *bearers*. Estes procedimentos são usados para gerenciar as conexões com o PDN GW de um equipamento do usuário (OLSSON et al., 2009).

Os procedimentos realizados pelo gerenciamento de seções do EPS são iniciados pela rede, mas o dispositivo também pode solicitar a modificação de algum recurso dos *dedicated/default bearers* ou requisitar a rede para executar procedimentos de ativação ou desativação dos *dedicated/default bearers* (OLSSON et al., 2009). Os procedimentos do gerenciamento de seções do EPS são:

- Procedimento de ativação do *default bearer*;
- Procedimento de ativação de um *dedicated bearer*;
- Procedimento de modificação do estado de um *dedicated* ou *default bearer*;
- Quando o dispositivo do usuário solicita procedimento de conexão com um PDN GW;
- Quando o dispositivo do usuário solicita procedimento de desconexão com um PDN GW;
- Quando o dispositivo do usuário solicita a modificação de um recurso do *dedicated* ou *default bearers*;
- Procedimento de desativação do *default bearer*;
- Procedimento de desativação do *dedicated bearer*.

O MME também foi projetado para possuir comunicação com as demais redes padronizadas pelo 3GPP e redes que não são padronizadas pelo 3GPP. Em uma rede comercial, essa comunicação é de fundamental importância, pois muitas vezes a rede não possui uma cobertura totalmente 4G. (CISCO,2016).

Outra função do MME é a seleção do PDN GW. Essa função usa informações do assinante fornecidas pelo HSS. Para cada um dos PDN GW's contido na rede (CISCO,2016), o HSS fornece:

- Um endereço IP de um PDN GW e uma APN¹;
- Em caso de *roaming* com outras operadoras de telefonia móvel, o HSS indica uma APN e aloca um PDN GW para o equipamento do usuário visitante.

O HSS também indica a APN padrão para o dispositivo, para estabelecer conexão com um outro PDN GW quando o dispositivo já está conectado a um ou mais PDN GW's (CISCO,2016).

O MME é também responsável pela seleção do SGW para a rede LTE. Essa seleção é baseada na topologia da rede, ou seja, o SGW selecionado está geograficamente localizado mais próximo do dispositivo do usuário. No caso de sobreposição de áreas de serviço. O MME pode preferir o SGW com uma área de serviço que possui uma probabilidade de mudança reduzida. Isso é definido conforme o dispositivo do usuário se locomove pela rede (OLSSON et al., 2009).

Em uma visão geral da rede então, o dispositivo do usuário inicia o procedimento de *Attach* enviando uma solicitação para o eNodeB. Dentro desta mensagem estão contidos diversos identificadores da rede como, por exemplo, última TA visitada (se houver), capacidade de rede do dispositivo, entre outros.

O eNodeB seleciona o MME. Após a seleção do MME, o eNodeB através da interface S1-MME encaminha a mensagem de *Attach* para o MME escolhido. Caso o MME desconheça o dispositivo, o MME envia uma solicitação de Identidade para o equipamento solicitando o IMSI. O equipamento do usuário responde então a solicitação com o IMSI.

Neste momento acontece o procedimento de autenticação AKA. O MME envia o IMSI para o HSS de origem do usuário. O HSS por sua vez realiza o AKA, concedendo todas as chaves cifradas para a rede. A partir deste momento a rede sabe de fato a identidade do usuário. Como mencionado anteriormente, a interface utilizada entre o MME e o HSS é a S6a.

¹*Access Point Name* (APN) é um identificador usado pela rede. A APN é composta por dois identificadores, o *APN Network Identifier* (APN-NI), cuja função é definir o PDN GW no qual o dispositivo solicitará a conectividade e, opcionalmente, um serviço solicitado pelo equipamento do usuário. Esta parte da APN é obrigatória. O segundo identificador é o *APN Operator Identifier* (APN-OI), cuja função é definir em qual rede da operadora o PDN GW está localizado. Esta parte da APN é opcional (OLSSON et al., 2009).

O MME envia uma solicitação de atualização da localização do usuário (identidade do MME, IMSI,) para o HSS. O HSS responde com dados do assinante (IMSI, Dados de Subscrição). Esses dados contém a lista de todas as APN's que o dispositivo está autorizado a acessar. Cada APN, por sua vez, possui um conjunto de parâmetros de QoS.

Após o recebimento da mensagem do HSS, o MME seleciona o SGW e o PDN GW, após essa seleção o usuário irá obter um *default bearer* e ganhará conectividade com a internet.

3.3 SERVING GATEWAY (SGW)

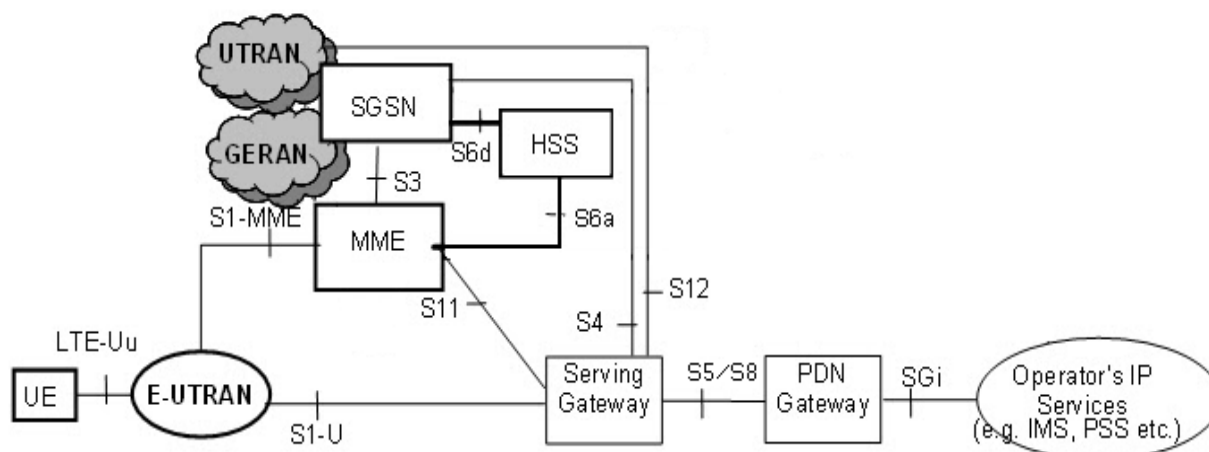
O SGW está padronizado pelo 3GPP na TS 23.002 e TS 23.401. Segundo o 3GPP, o SGW é o *gateway* que converge os dados do usuário entre o E-UTRAN e o EPC (3GPP TS 23.401). A cada dispositivo está associado um SGW. As funções do SGW incluem:

- Âncora para a mobilidade do usuário, entre o LTE e outras tecnologias do 3GPP (retransmitindo o tráfego de dados para os sistemas 2G e 3G);
- Roteamento de pacotes de dados do usuário para PDN GW;
- Ponto de âncora para a mobilidade do usuário entre os eNodeBs da rede LTE (*handover* sem a perda de conexão e dados);
- Envio de marcadores finais para o eNodeB, ou envio de marcadores finais através da interface S4/S12 para tecnologias anteriormente padronizadas pelo 3GPP (2G e 3G respectivamente);
- Interceptação Legal;
- Marcação no cabeçalho do pacote de *uplink* e *downlink* para a rede LTE, como, por exemplo, aplicando o *DiffServ Code Point*, baseado no QCI do *dedicated* ou *default bearer* ao qual o serviço está associado;
- Contabilização de pacotes para a tarifação para um usuário de uma operadora visitante.

Ao longo desta seção as funções e conceitos aplicados do SGW serão descritos, bem como seu funcionamento. A figura 15 ilustra as entidades e

interfaces que se comunicam com o SGW. O *GSM EDGE Radio Access Network* (GERAN) é o acesso de rádio da tecnologia 2G, enquanto o *UMTS Terrestrial Radio Access Network* (UTRAN) é o acesso de rádio da tecnologia 3G.

Figura 15 -SGW e a comunicação com as entidades do EPS



Fonte: Alterada de 3GPP TS 23.401 (2010, p. 16.)

A primeira funcionalidade que será explicada é a comunicação para a mobilidade do usuário. O SGW se conecta com o SGSN através das interfaces S4 e S12. Desta forma é criado um ponto de ancoragem comum entre a tecnologia LTE e as tecnologias 2G (GERAN) e 3G (UTRAN), isto está ilustrado na figura 15. Isto significa que todo o tráfego de dados do usuário relacionado a um dispositivo que está em *roaming* passará pelo SGW, independentemente da rede de rádio que está sendo usada. Isto difere como o *roaming* é tratado nas tecnologias 2G e 3G. Nas tecnologias anteriores, o próprio SGSN realizava o *roaming* entre as tecnologias 2G e 3G (OLSSON et al., 2009).

Com todo o tráfego de dados de *roaming* passando por um único ponto na rede, a operadora pode controlar e monitorar o tráfego de forma consistente, potencialmente aplicando políticas nos fluxos de dados (OLSSON et al., 2009).

O SGW é responsável também pelo roteamento de pacotes para o PDN GW. Através desse roteamento o dispositivo ganha conectividade com a Internet. O SGW e o PDN GW se comunicam através da interface S5/S8, usando o protocolo GTP. O *GPRS Tunneling Protocol* (GTP) é um protocolo padronizado pelo 3GPP que é usado nas tecnologias 2G, 3G e no LTE. O GTP é dividido em

GTP-U, para transporte de dados do usuário, e o GTP-C, utilizado para realizar o controle da rede (OLSSON et al., 2009).

O SGW possui ainda a função de marcação de pacotes finais chamados de *end-makers*. Este marcador, que vai no cabeçalho do pacote, informa à rede que este pacote encerra o recebimento de dados do usuário. Desta maneira, o equipamento do usuário entrará em modo ocioso, também chamado de *Idle Mode*. Desta maneira a rede economiza recursos do EPC. A conexão só será restabelecida quando chegar um pacote destinado a este usuário. O SGW por sua vez irá informar ao MME que ocorreu a chegada de um pacote endereçado para o usuário. O MME realizará o *paging* para atualizar a rede a respeito da localização do dispositivo e ativar o *dedicated ou default bearer* (OLSSON et al., 2009).

O SGW também realiza a interceptação legal, que é um dos requisitos regulatórios que as operadoras devem satisfazer como obrigação legal para as autoridades governamentais na maioria dos países em que estão operando (OLSSON et al., 2009).

A interceptação legal permite que as autoridades competentes realizem a interceptação do tráfego de usuários específicos. Esta ativação ou desativação requer documento legal como, por exemplo, um mandado, interrogação e outros tipos de procedimentos legais.

O SGW é responsável ainda por marcar os pacotes de *downlink* e *uplink* para a camada de transporte entre as entidades do EPC e do E-UTRAN. Os equipamentos de transporte que conectam as entidades entre o EPC e o E-UTRAN não possuem conhecimento dos *dedicated/default bearers* do EPS. No entanto a camada de transporte deve identificar um serviço diferenciado em andamento para poder dar prioridade na rede. Para isso o SGW marca o cabeçalho do pacote.

Essa marcação está especificada na TS 36.414. A marcação de *Differentiated Services Code Point* (DSCP) é feita de acordo com características da classe QCI e outros parâmetros de tráfego. As categorias de tráfego são específicas e podem ser determinadas a partir dos parâmetros da aplicação.

Assim, considerando a visão geral do sistema, após a autenticação do dispositivo, o MME seleciona o SGW. O MME solicita a criação do *default bearer* para o SGW, tendo em vista que o MME não se comunica diretamente com o

PDN GW. O SGW envia uma mensagem chamada de *create default bearer request* para o PDN GW. Dentro da mensagem estão contidos parâmetros como a APN do dispositivo, IMSI, endereço para alocação do PDN entre outros parâmetros (CISCO, 2016).

Após a criação do *default bearer* no PDN GW, o SGW recebe uma mensagem confirmando a criação do *default bearer*, e envia uma mensagem para o MME confirmando a requisição. Com o *default bearer* criado, o SGW realiza suas funções como, por exemplo, âncora de mobilidade para *handovers* e roteamento para o PDN GW (CISCO, 2016).

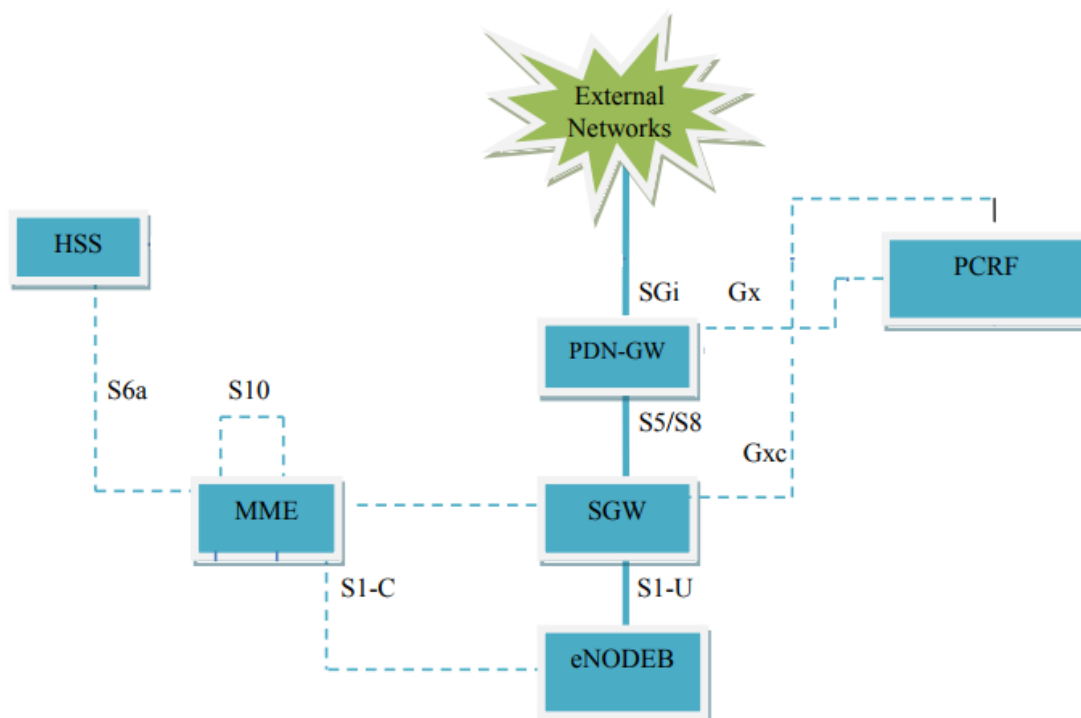
3.4 PACKET DATA NETWORK GATEWAY (PDN GW)

O PDN GW é o *gateway* que termina a rede da operadora e proporciona a conexão do dispositivo do usuário com a rede mundial de computadores através da interface SGi. Este *gateway* atua como ponto de âncora para outras redes não confiáveis e possui uma importante função para a tarifação e controle de políticas no EPS. As funções do PDN GW no EPS são definidas na TS 23.401 e na TS 23.402. Dentro dessas especificações técnicas, as funções do PDN GW na arquitetura do EPS são:

- Inspeção de pacotes como, por exemplo, *Deep Packet Inspection* (DPI). O DPI será abordado nesta seção;
- Interceptação Legal;
- Alocação do endereço IP para o equipamento do usuário;
- Marcação do pacote IP tanto de *Uplink* quanto de *Downlink*;
- Suporte para tarifação;
- Execução da taxa de transferência de Downlink baseado no *Aggregate Maximum Bit Rate* (AMBR). O AMBR limita o *throughput* máximo que o usuário pode utilizar;
- Âncora de mobilidade para sistemas padronizados pelo 3GPP e sistemas não padronizados pelo 3GPP.

A figura 16 ilustra o PDN GW se comunicando com os elementos da rede EPS. Com base nesta figura serão comentadas as funções e comunicações realizadas pelo PDN GW.

Figura 16 - PDN GW e a comunicação com as entidades do EPS



Fonte: Alterada de Singh (2009, p. 23)

A inspeção de pacotes é uma função realizada pelo PDN GW, sendo vista como essencial para uma operadora. Essa inspeção auxilia as operadoras a manter o controle e fazer uso eficiente dos recursos de rede disponíveis. Também permite a criação de serviços diferenciados e reduz as preocupações com segurança. Além disso, as preferências dos assinantes e a eficácia das campanhas de marketing podem ser analisadas e avaliadas para identificar novas oportunidades de negócios. Com a inspeção de tráfego de dados, a operadora pode cobrar de forma diferente cada serviço. (ERICSSON, 2010).

No contexto das operadoras de telefonia móvel, a inspeção de pacotes é dividida em três: rasa, profunda e heurística. Essa inspeção é feita de acordo com as camadas do modelo *Open Systems Interconnection* (OSI) definido pela *International Organization for Standardization* (ISO) de número 7498-1. Esta padronização define sete camadas de abstração para redes de computadores. As implementações das inspeções são feitas de acordo com o planejamento do fornecedor.

A inspeção rasa ou também chamada de *Shallow Packet Inspection* (SPI) examina até a camada 4 do modelo OSI. Com o SPI, os administradores podem impedir que os dispositivos de clientes recebam arquivos flash do YouTube ou arquivos de imagens de sites de redes sociais, por exemplo. O SPI pode priorizar alguns pacotes sobre outros, para isto é examinado até a quarta camada. No entanto, o SPI nem sempre determina o conteúdo do pacote. Para isso é necessária uma inspeção profunda (PARSONS, 2012).

A inspeção profunda ou também chamada de *Deep Packet Inspection* (DPI) permite em tempo real que as operadoras identifiquem com precisão a origem e o conteúdo de cada pacote que está trafegando pelo PDN GW. Esta a inspeção mais interessante no contexto de uma operadora. A inspeção de pacotes até a camada 7, permite o fornecimento de informações cruciais para a operadora e sistemas de suporte empresarial, sem comprometer outros serviços (PARSONS, 2012).

Através da utilização do DPI, a operadora recebe relatórios sobre cada usuário, identificando seu perfil e gerando novas oportunidades de negócio. Os fornecedores oferecem duas soluções para o DPI: uma em que o DPI fica alocado dentro do PDN GW e uma solução de DPI à parte do PDN GW. Esta última diminui o processamento no *gateway*, porém aumenta o número de nós na rede, podendo ocasionar um *delay* (PARSONS, 2012).

Os métodos de SPI e DPI não são suficientes no caso de tráfego *peer-to-peer*². Os protocolos de muitas aplicações *peer-to-peer* são projetados para evitar a detecção, o mesmo requer uma abordagem diferente. Ao invés de inspecionar pacotes individuais, são analisados os padrões de bits e a frequência de pacotes em diferentes sessões. Essas estatísticas são então combinadas com informações sobre o diâmetro da rede (distância média entre pares de nós que estão se comunicando) e *hosts* que atuam tanto como servidores quanto como clientes. Este tipo de inspeção de tráfego geralmente é referido como

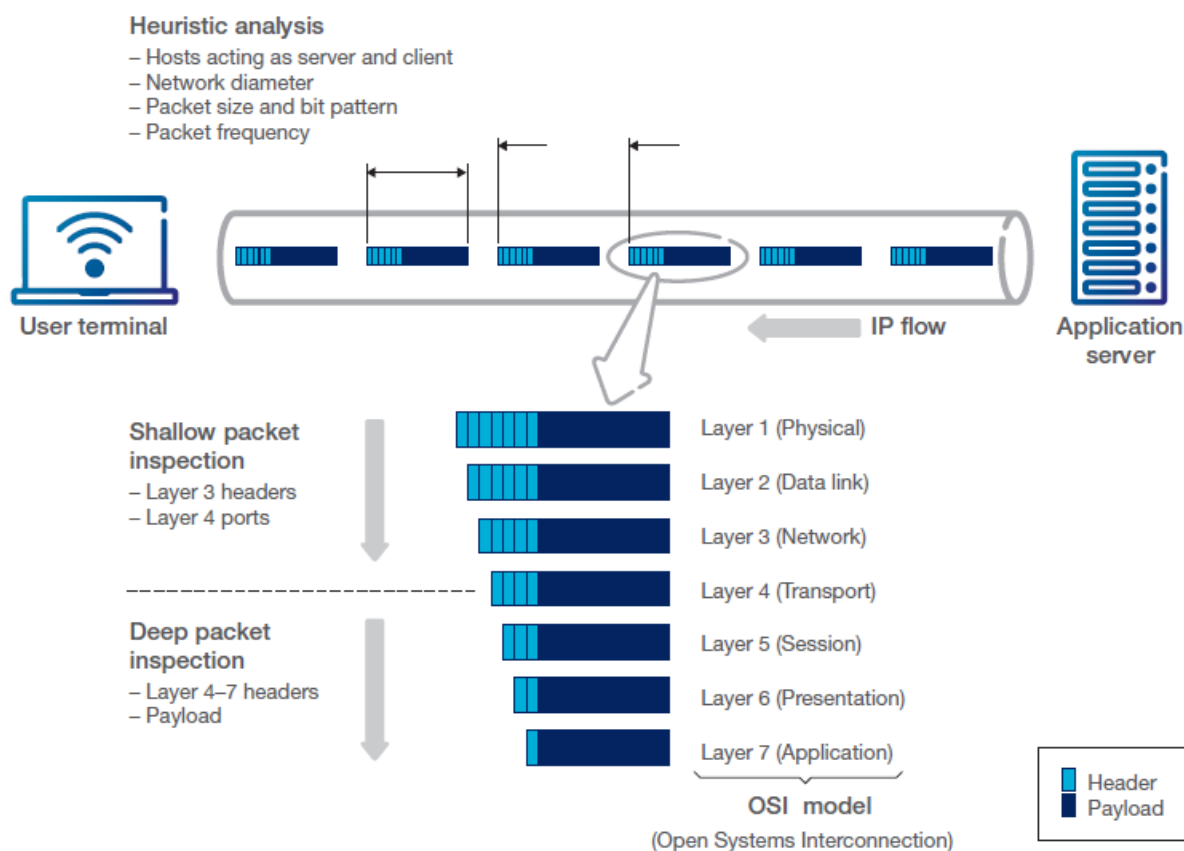
²Em termos gerais, é uma arquitetura de software que permite que pontos (indivíduos) se comuniquem diretamente com outros pontos através de computadores em uma rede intermediária. Mais especificamente, uma rede *peer-to-peer* não é baseada no conceito cliente-servidor. A principal característica do *peer-to-peer* é que pares trocam informações e serviços mutuamente usando conexões diretas sem intervenção do servidor. A principais aplicações do *peer-to-peer* são *torrents* como uTorrent e aplicações de chamadas de vídeo/áudio como Skype.

análise heurística e pode ser usado para identificar a maioria das aplicações *peer-to-peer* (ERICSSON, 2010).

Deve-se notar que a análise heurística nem sempre é completamente precisa, uma vez que se baseia em reconhecimento de padrões. Novos protocolos e aplicativos *peer-to-peer* aparecerão, e a inteligência da rede para análise heurística terá que ser atualizada continuamente (ERICSSON, 2010).

A figura 17 ilustra as inspeções SPI, DPI e heurística.

Figura 17 - Inspeção de tráfego: SPI, DPI e análise heurística



Fonte: ERICSSON (2010, p.7)

O PDN GW também possui função de suporte para tarifação. Em contraste com as arquiteturas padronizadas anteriormente que realizam um controle estático que causava problemas para criação e gerenciamento de novos serviços, o LTE promove um controle centralizado, fornecendo uma camada de controle com consciência global dos elementos da rede. Com o controle centralizado de políticas, há uma simplificação do processo de aplicação de regras para todos os *gateways* da rede. Esse controle permite o uso de políticas

dinâmicas que aloca de recursos dinâmicos na rede, gerenciando de forma mais eficiente a capacidade e os recursos da rede. Além disso, com as políticas dinâmicas são possíveis implementar novas regras *online* na rede, facilitando a implementação de novos serviços e alterando os existentes (CISCO, 2016).

Esses conceitos de inspeção de pacote e diferentes políticas estão padronizados no *Policy Control and Charging* (PCC). O PCC fornece à operadora uma série de ferramentas para diferentes parâmetros de QoS, controle de serviços e tarifações. Tendo-se em vista que existem diferentes serviços e cada um possui diferentes requisitos de parâmetros de QoS. É importante garantir que os serviços possam coexistir e que cada serviço possua um caminho de transporte apropriado (OLSSON et al., 2009).

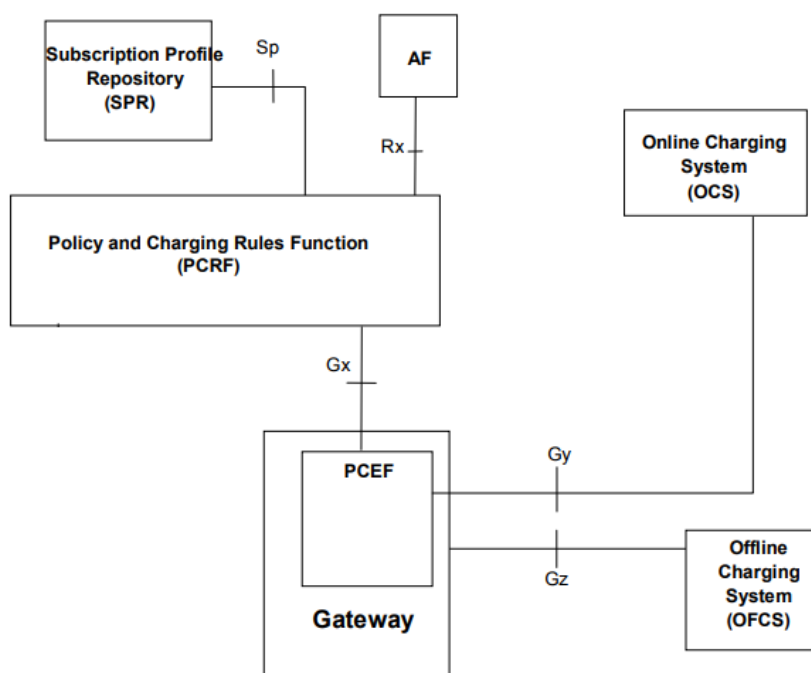
A arquitetura do PCC para o EPS foi definida no *Release 7*. A arquitetura do PCC também foi projetada para operar quando o usuário está em *roaming*. Desta forma, o usuário tem acesso aos mesmos serviços se estiver ancorado em um *gateway* que não faz parte da sua rede doméstica.

A arquitetura do PCC está ilustrada na figura 18. O *Subscriber Profile Repository* (SPR) é uma base de dados que contém informações como políticas e dados específicos de cada usuário na rede. O *Online Charging System* (OCS) é um sistema de gerenciamento de crédito para cobrança pré-paga. Essa comunicação ocorre através da interface Gy e permite a operadora oferecer a modalidade pré-paga para os seus usuários, isto ocorre em tempo quase real. O *Offline Charging System* (OFCS) se comunica com o PCEF e gera registro de tarifações para cobrança pós-paga, comunicação feita pela interface Gz. O *Application Function* (AF) interage com aplicativos ou serviços que exigem regras dinâmicas. O AF extrai a informação da sessão através da sinalização do aplicativo e fornece-a ao PCRF. Um exemplo disso é uma chamada via VoLTE. A aplicação sinaliza para o AF, que, por sua vez, informa o PCRF, através da interface Rx (SINGH, 2009).

O PCRF é o ponto de decisão do controle de políticas do PCC. O PCRF possui a função de decisão de políticas e controle das mesmas, funcionalidades relacionadas ao controle da tarifação do usuário, procedimentos de segurança, decisões de como as aplicações devem ser tratados, autoriza recursos de QoS. Com essas funções atribuídas ao PCRF, a operadora possui uma visão geral da

rede, recebendo relatórios de receitas, consumo dos usuários, possíveis novos serviços, entre outros (OLSSON et al., 2009).

Figura 18 - Visão geral da arquitetura do PCC



Fonte: alterada de 3GPP TS 23.203 (2009, p. 19)

O bloco lógico chamado de *Policy Charging Enforcement Function* (PCEF) faz a execução das diferentes políticas de tratamento do usuário ou de serviços, como descrita na TS 29.212. Esta entidade pode estar localizada dentro do PDN GW ou pode ser um elemento de rede a parte. Com arquitetura do PCC, a rede pode criar uma prioridade de serviço de acordo com os parâmetros de QoS, isto é considerado uma importante função do PCC. Essas regras por sua vez são transmitidas do PCRF para o PCEF (OLSSON et al., 2009).

Outra importante função do PDN GW é a alocação do endereço IP para o dispositivo. É importante ressaltar que existem diversas aplicações e serviços no mercado, as diferentes aplicações podem utilizar IPv4 ou IPv6. Cabe a rede poder fornecer conexão IP para as múltiplas aplicações utilizando IPv4 e/ou IPv6. Uma conexão IP do PDN GW deve, portanto, fornecer conectividade usando a versão apropriada do protocolo IP. Atualmente, muitas redes IP e aplicações

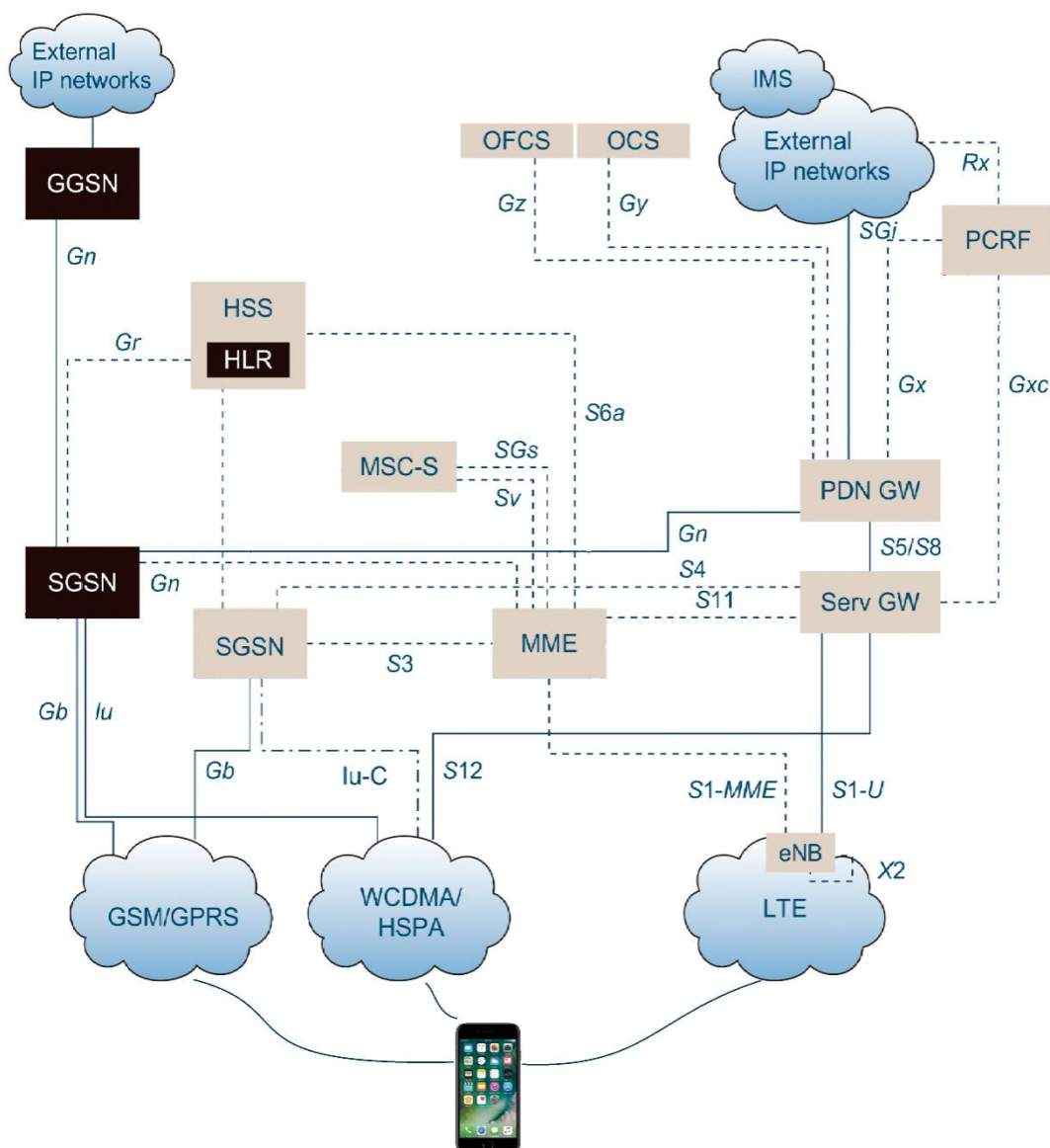
ainda são baseadas em IPv4. Ou seja, o usuário receberá um endereço IPv4 e acessará os serviços baseados em IPv4. Mas existe a tendência de as conexões IPv6 se tornarem mais comuns (OLSSON et al., 2009). Segundo o Google, a adoção do IPv6 no Brasil é de 21,67% até o dia 31 de outubro de 2017 (essa estatística é atualizada semanalmente).

Em uma visão geral do funcionamento do PDN GW. O SGW envia uma mensagem para a criação do *default bearer* para o PDN GW. Essa mensagem contém o IMSI, endereço IP do dispositivo, AMBR e diversas identidades que são usadas para identificar o usuário e informações de localização. O PDN GW interage com o PCRF para obter as regras padrão do PCC para serem aplicadas ao equipamento. O PDN GW fornece ao PCRF as informações de AMBR, IP do dispositivo, IMSI, entre outras, esta mensagem tem como base a mensagem recebida pelo SGW. O PDN GW envia uma mensagem respondendo a requisição do SGW. Esta mensagem contém o endereço IP conforme a aplicação o solicita, podendo ser IPv4 ou IPv6. A partir deste momento, o fluxo de dados de *downlink* pode ser enviado para o SGW, porém o SGW armazena esses dados até o procedimento de *attach* estar completamente concluído. O SGW envia uma mensagem para o MME. O MME, por sua vez, aceita a solicitação de *attach*, envia a mensagem destinada para o eNodeB e o mesmo estabelece o *bearer* de rádio para o dispositivo. Após o estabelecimento do *default bearer*, o PDN GW pode realizar as demais funções como a inspeção de pacotes e a execução das regras definidas pelo PCC (CISCO, 2016).

4. ARQUITETURA E SERVIÇOS

No capítulo 3 foi detalhado o EPC, suas entidades e funcionalidades. Este capítulo irá mostrar a arquitetura do LTE de forma mais ampla, com as entidades se comunicando. A figura 19 ilustra a arquitetura de rede com as tecnologias de acesso de rádio padronizadas pelo 3GPP. Nessa figura é possível visualizar de maneira clara todo o sistema, com as interfaces conectando as entidades.

Figura 19 - Arquitetura EPS com as tecnologias padronizadas pelo 3GPP



Fonte: Alterado de Olsson et. al. (2009, p. 33)

A comunicação com os sistemas padronizados pelo 3GPP é fundamental para uma operadora comercial, tendo em vista que nenhuma tecnologia móvel possui cobertura em toda a área territorial de um país ou região. Se for levado

em conta um país como o Brasil por exemplo, é improvável que toda a infraestrutura de acesso de rádio seja totalmente LTE, devido à sua extensão territorial. Existem algumas soluções de arquitetura para que um usuário possa continuar conectado à rede, mesmo que se locomova de uma tecnologia de acesso para outra. É válido mencionar que algumas aplicações são sensíveis a essa troca de tecnologias de acesso de rádio (OLSSON et al., 2009).

O procedimento em questão é válido somente para dispositivos compatíveis com a tecnologia 4G que possuem uma conexão IP. Estes dispositivos sempre têm o seu ponto de referência no EPC, ou seja, mesmo se estiverem ancorados em um ponto de acesso da tecnologia 3G, os dados do usuário serão transferidos para um PDN GW. Normalmente é usada a APN do dispositivo para determinar em qual PDN GW o equipamento será ancorado (OLSSON et al., 2009).

O motivo pelo qual a rede foi projetada desta maneira vem do fato de que se o usuário simplesmente realizasse a troca de tecnologia, seria necessário um novo procedimento de *attach*. Além do mais, a conexão iria ser totalmente perdida e o dispositivo do usuário iria ganhar outro endereço de IP (OLSSON et al., 2009).

O procedimento de *handover* da tecnologia LTE para as tecnologias de acesso 2G e 3G é semelhante. Portanto será abordado somente o procedimento da tecnologia LTE para o 3G. Quando o dispositivo do usuário se movimenta de eNodeB para NodeB, o eNodeB emite uma mensagem para o MME, informando que o usuário está saindo de sua área de cobertura e se dirigindo para uma área de cobertura que não possui eNodeB's. O MME por sua vez inicia o processo de transferência. Para isso, o MME seleciona o SGSN, através da interface S3. Essa seleção será baseada em qual RNC o dispositivo irá ficar ancorado. Após isso, o MME mapeará o contexto do usuário. Nesse contexto estão parâmetros como o *default bearer* e os parâmetros de QoS. Essas informações serão enviadas ao SGSN (OLSSON et al., 2009).

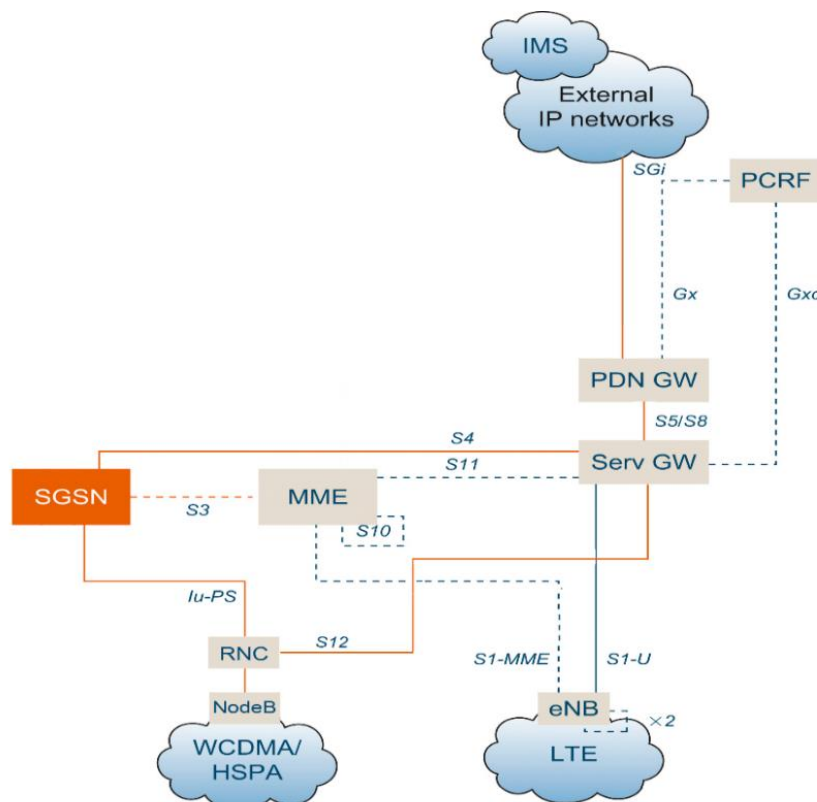
O SGSN selecionará o SGW em que o dispositivo será ancorado através da interface S4. O SGSN também fornece informações necessárias para estabelecer os recursos de rádio para o RNC. Após a alocação desses recursos, o RNC envia uma mensagem ao SGSN informando que concluiu a alocação de

recursos e está preparado para receber dados do usuário que chegarão do SGSN ou do SGW, através da interface S12 (OLSSON et al., 2009).

O eNodeB que realizará o procedimento de *handover* é informado de que os recursos do RNC estão reservados e que o processo de preparação está concluído. O eNodeB sincroniza com a célula do UTRAN que passa a receber dados do usuário. Por fim, o SGSN é informado sobre o sucesso do *handover* e envia uma mensagem ao MME sobre o mesmo (OLSSON et al., 2009).

A figura 20, através da linha contínua alaranjada, mostra o caminho que o pacote percorre na arquitetura quando um dispositivo 4G está ancorado em uma rede 3G. Como mencionado anteriormente, o 3GPP padronizou uma interface lógica S12 que comunica diretamente o RNC e o SGW.

Figura 20 - Caminho percorrido pelo pacote

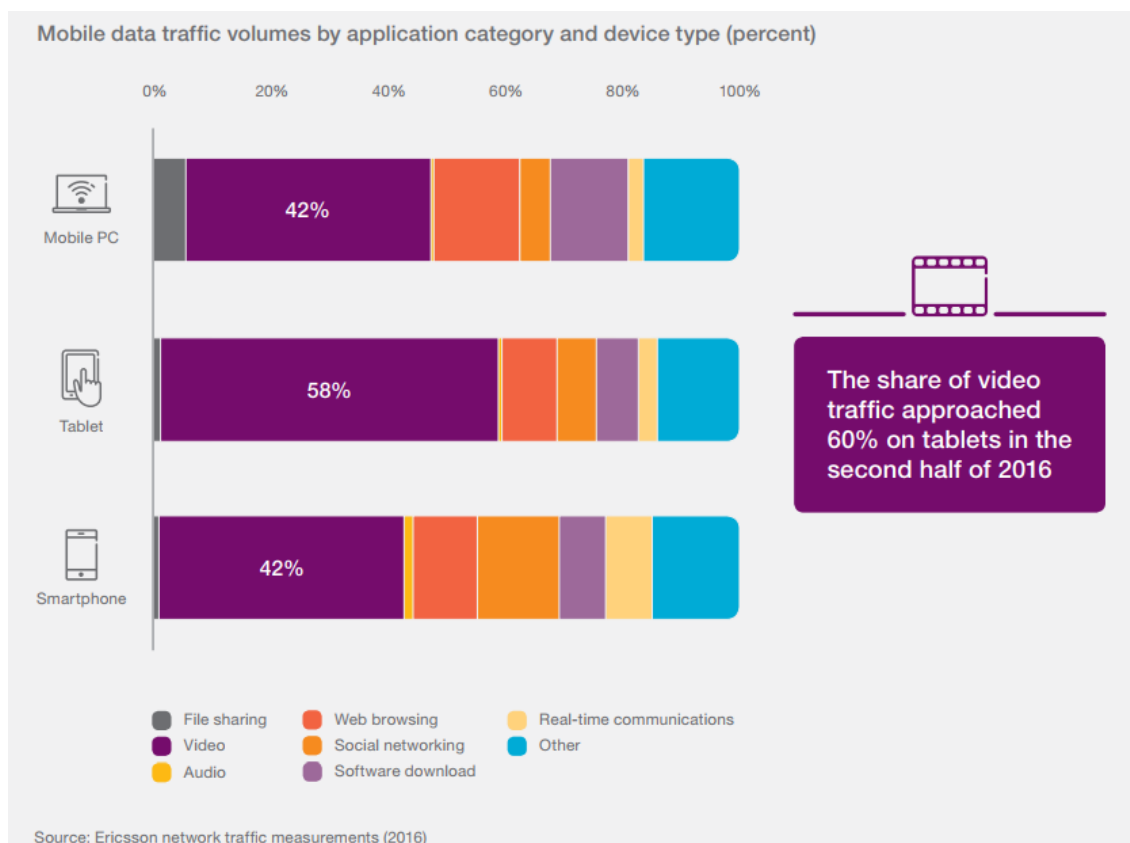


Fonte: Alterada de Olsson et. al. (2009, p.44)

Há uma diversidade de aplicações e serviços que podem ser utilizados pelo usuário móvel na rede LTE, afinal, este foi o propósito da sua criação, operar fim-a-fim com protocolo IP. Cada uma dessas aplicações e serviços exige uma resposta da rede de forma mais eficiente. Para entender o contexto geral do uso de dados por aplicação é apresentada a figura 21, ilustrando a quantidade de

dados que são utilizados pelos usuários de acordo com a aplicação e com o dispositivo.

Figura 21 - Volume de dados utilizados relacionando tipos de dispositivos e aplicações.



Fonte: Ericsson (2016, p.10)

A aplicação de vídeo será a primeira a ser comentada neste trabalho. A figura 21 mostra que a aplicação de vídeo é a mais consumida pelos usuários e o maior desafio para as operadoras. Segundo projeções feitas pela Ericsson, em 2022 o tráfego de vídeos na rede será de 75%. Sendo assim, essa aplicação é crítica em um planejamento de rede.

Uma operadora pode criar serviços dedicados a *streaming* como, por exemplo, um serviço dedicado de vídeos para o *Netflix*. Na rede 4G, a forma de implementação desses serviços é feita com base na tabela de QCI. Na tabela de QCI ilustrada na figura 11, existe uma classe para vídeos não conversacionais com taxa garantida de bits, classe 3. Existem também classes sem a taxa de garantia de bits, mas que possuem um baixo *delay* na rede como, por exemplo, os QCI's 6 e 7.

Outro serviço que deve ser levado em conta no momento de planejamento de uma rede 4G é o VoLTE, serviço de voz na rede LTE. O VoLTE é a tendência para o mundo e já foi lançado em mais de 100 redes em 55 países. A Ericsson faz uma projeção de que o número de assinantes que utilizará o VoLTE irá chegar a 4,6 bilhões até o final de 2022. O VoLTE é realidade em muitas redes dos EUA, Japão, Coréia do Sul e Canadá. Em algumas redes, o percentual de chamadas realizadas pelo VoLTE chega a 80%. A tecnologia VoLTE também será utilizada como base para permitir serviços de comunicação nas futuras redes 5G (ERICSSON, 2016).

O VoLTE é baseado na interação do IMS e o EPC. Apesar de o volume de dados usados pelo VoLTE ser muito inferior ao volume de dados de uma aplicação de vídeo, o VoLTE exige da rede uma latência baixa. Por ser um serviço conversacional, os usuários que estão se comunicando não podem perceber atrasos durante a comunicação. O VoLTE utiliza a classe 1 da tabela do QCI, como mostrado na figura 12. Isto garante a máxima experiência do usuário ao utilizar o serviço. O EPC deve alocar recursos especiais para a realização da chamada, tendo em vista que o VoLTE exige a criação de um *dedicated bearer*.

Outro fator importante para o planejamento da rede é como os equipamentos do EPC serão implementados. Essa implementação poderá ser distribuída ou centralizada e dependerá de vários fatores, alguns deles são listados a seguir:

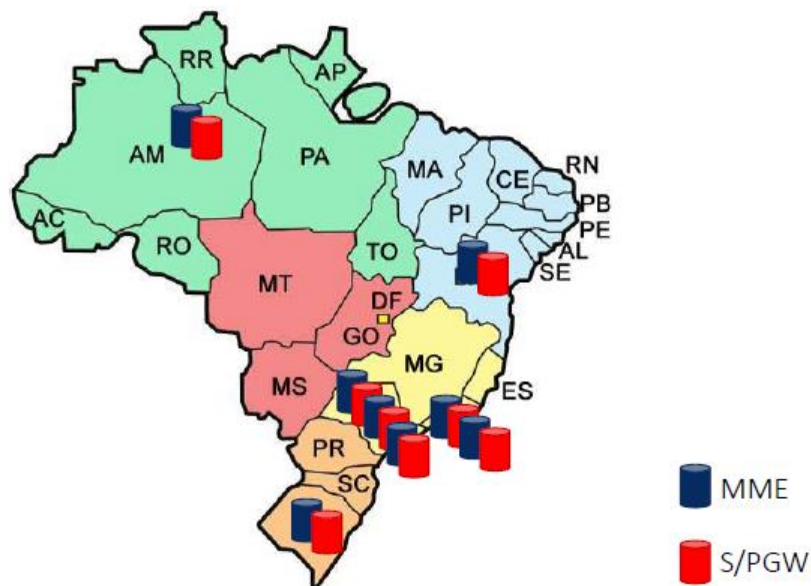
- Arquitetura da rede de transmissão – o importante de ser analisado neste fator é como a rede de transmissão (*backbones* e roteadores secundários) está distribuída na rede.
- Custo da rede de transmissão – com introdução de novos equipamentos na rede é importante fazer o planejamento dos custos com a conexão física dos equipamentos e a sua transmissão.
- Custo para implantação e operação do EPC – os equipamentos do EPC são caros e sua operação depende de técnicos especializados.
- Saídas de Internet, ou seja, cabos de fibra ótica submarinos - os exemplos que serão utilizados neste trabalho se basearão no mapa

do Brasil. É importante ressaltar que o Brasil possui quatro saídas de internet, localizadas em São Paulo, Rio de Janeiro, Bahia e Ceará.

- Serviços implementados – VoLTE, serviços específicos para vídeos, etc.
- Aplicações e serviços acessados.
- Localização dos provedores de serviços – quais servidores são acessados pelos usuários e onde se localizam como, por exemplo, os servidores do Google, que normalmente estão localizados nos EUA.
- Cache na rede e sua localização – em geral as operadoras armazenam cópias dos principais servidores que são usados pelos usuários, diminuindo a latência da rede.

Serão apresentadas algumas imagens com as diferentes distribuições dos equipamentos do EPC na rede e, será comentado qual dessas distribuições favorece determinado serviço ou aplicação. A figura 22 é uma arquitetura com equipamentos centralizados. Essa configuração favorece aplicações como *Web Browsing* ou apenas navegação em redes sociais, ou seja, aplicações que não são críticas e não dependem de uma resposta com baixa latência ou uma taxa garantida de bits. Pelo fato de possuir custos de implantação menores e pela centralização dos equipamentos, diminuindo o custo de operação. Esse modelo possui um bom custo-benefício para as operadoras.

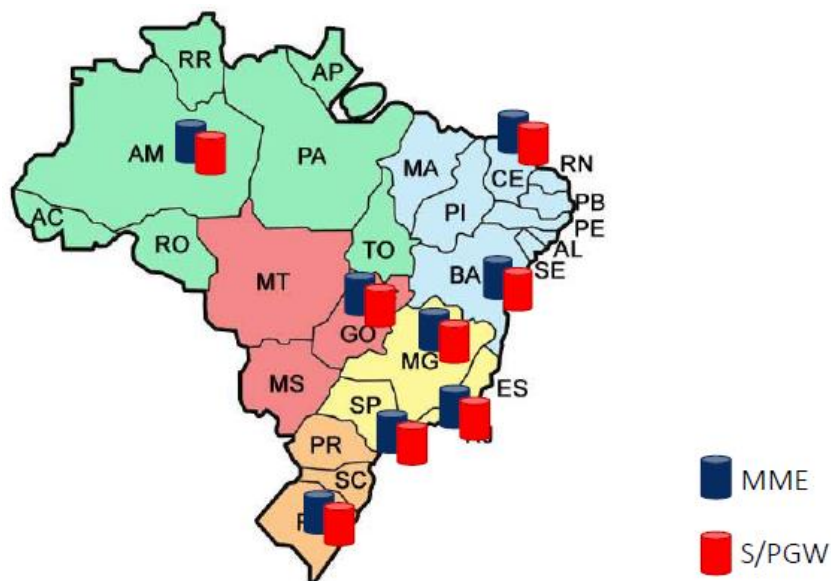
Figura 22 - Cenário de arquitetura centralizada



Fonte: Bitar (2017, p. 17)

A figura 23 ilustra os equipamentos totalmente distribuídos, espalhando os equipamentos que lidam com o tráfego de dados do usuário e equipamentos que realizam o controle da rede. Esse tipo de configuração é favorável para qualquer tipo de serviço implementado, principalmente para o VoLTE, porém essa implantação possui um custo elevado. Levando-se em conta que sempre deve haver duplicidade na rede, para caso um equipamento apresente uma falha ou realize manutenção do mesmo, essa configuração se torna ainda mais cara. Outro fator importante a ser levado em conta é a operação dos equipamentos que se torna mais difícil.

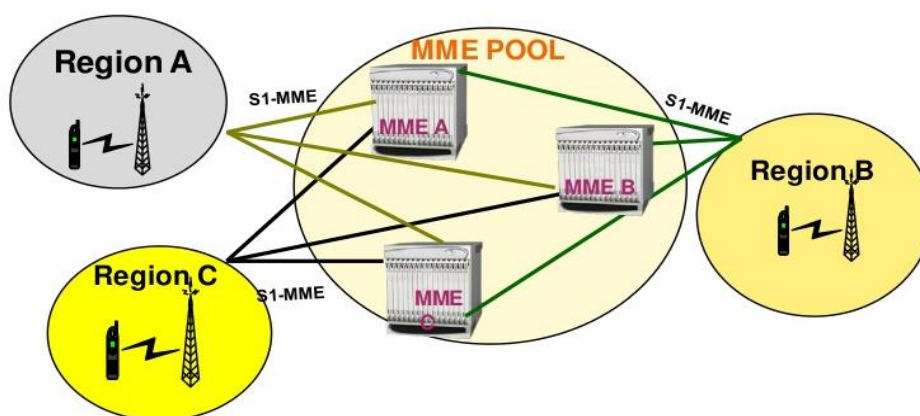
Figura 23 - Cenário de arquitetura totalmente distribuída



Fonte: Bitar (2017, p. 14)

Antes de continuar com as diferentes formas de distribuir os equipamentos do EPC. É necessário explicar o que é *pool* de MME. Essa configuração é definida como uma área em que o dispositivo pode se locomover pela rede sem necessidade de alterar o MME de atendimento. Uma área de *pool* de MME possui um ou mais MME's em paralelo. *MME Pool Areas* são um conjunto de áreas de MME (CISCO,2016). A figura 24 ilustra três MME's que estão ligados a todas as regiões de atendimento.

Figura 24 - Representação do pool de MME



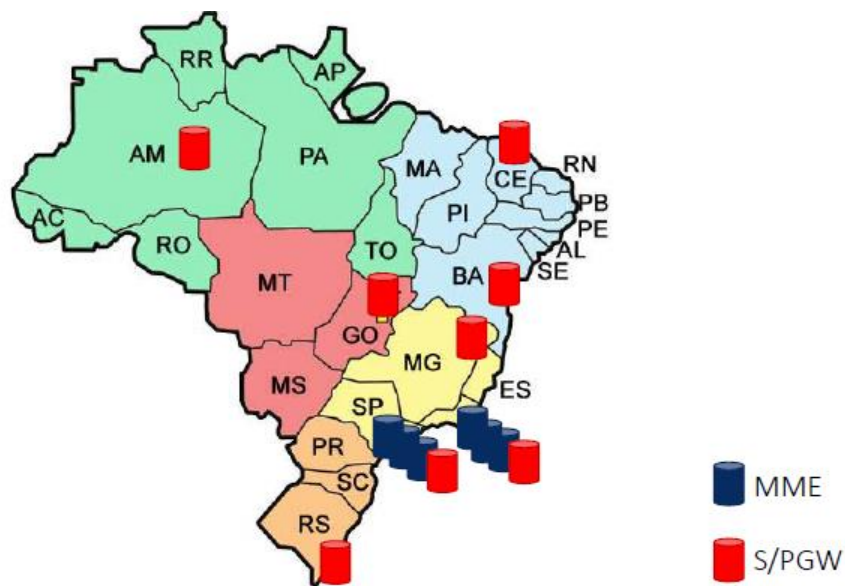
Fonte: Cisco e afiliados (2011, online)

Os benefícios do *pool* de MME são:

- Como a figura 4 mostra, o *pool* de MME permite a redundância geográfica, pois um *pool* pode cobrir todos os eNodeBs.
- Aumenta a capacidade total, pois permite o compartilhamento de carga entre os MME's.
- Converte o rastreamento do equipamento do usuário que antes pertencia a um MME de outra área de registro, para o rastreamento de um grupo de MME's que administra múltiplas áreas de registro. Isso reduz a carga de sinalização e os atrasos de transferência de dados.
- Facilita a introdução de novos MME's e a substituição de MME's antigos, pois a carga de um MME pode ser transferida para outro MME, facilitando a substituição do equipamento antigo.
- Elimina o ponto de falha entre um eNodeB e um MME.
- Permite o agendamento de manutenção num momento em que a carga não sobrecarrega a rede.

A configuração de *pool* do MME permite a centralização dos elementos de controle da rede. Isto é reforçado se for considerado que o MME lida somente com sinalizações e normalmente é planejado de acordo com o número de assinantes. Isso não afeta a experiência do usuário com a tecnologia. Esse tipo de configuração com o PDN GW e o SGW distribuídos favorece todos os tipos de serviços e aplicações, diminuindo os custos relacionados à operação e à manutenção dos MME's. Esse cenário pode ser observado na figura 25.

Figura 25 - Cenário de arquitetura com somente os GW's distribuídos



Fonte: Bitar (2017, p. 15)

Com os diferentes serviços e aplicações mostrados neste capítulo. O planejamento de rede e a distribuição dos equipamentos do núcleo da rede são dependentes das necessidades de cada serviço. Uma rede, por exemplo, que não possui grande demanda de dados e os serviços ofertados não causam impacto no EPC, tem como melhor opção de implantação a centralização de seus equipamentos. Já uma rede com grande demanda de dados, utilização de VoLTE e serviços de vídeos, é interessante ter os PDN GWs e SGWs distribuídos pela extensão territorial de acordo com a demanda.

5. VIRTUALIZAÇÃO DO EPC

Em outubro de 2012, o ETSI publicou um *whitepaper* introduzindo o conceito de *Network Functions Virtualization* (NFV) e anunciou o *Network Functions Virtualization Industry Specification Group* (NFV ISG), um grupo formado por mais de 150 empresas de diversas áreas, incluindo 28 provedores de serviços de Tecnologia da Informação (TI) com o objetivo de padronizar o NFV.

Essa publicação fez um levantamento de diversos problemas enfrentados pelas operadoras de telefonia móvel, introduzindo uma proposta para a solução destes problemas e apresentando desafios para a implementação do NFV. A virtualização das funções da rede promove uma mudança de paradigmas na indústria de telecomunicações. Durante toda a evolução da telefonia móvel no mundo, as funções de rede foram implementadas por *hardwares* e *softwares* que precisam ser obrigatoriamente do mesmo fabricante.

A solução tradicional para iniciar um novo serviço na rede ocorre desta maneira: as operadoras se comunicam com os fornecedores para adquirir uma variedade de novos equipamentos e encontrar espaço para acomodá-los, tornando cada vez mais difícil a sua implementação. Outro ponto importante é a lentidão deste processo (ETSI, 2012).

Problemas decorrentes da atual arquitetura são o aumento dos custos de energia, desafios para conseguir capital de investimento, projetar a atualização da rede, integrar os equipamentos e operar equipamentos cada vez mais complexos. Além disso, esses equipamentos atingem o fim do seu período útil rapidamente, exigindo a repetição do ciclo de implantação e integração, com pouco ou nenhum benefício de receita (ETSI, 2012).

Para solucionar esses problemas, os integrantes do NFV ISG criaram o conceito de NFV (2012, p. 5):

Network Functions Virtualization aims to transform the way that network operators architect networks by evolving standard IT virtualization technology to consolidate many network equipment types onto industry standard high volume servers, switches and storage, which could be located in Datacenters, Network Nodes and in the end user premises. It involves the implementation of network functions in software that can run on a range of industry standard server hardware, and that can be

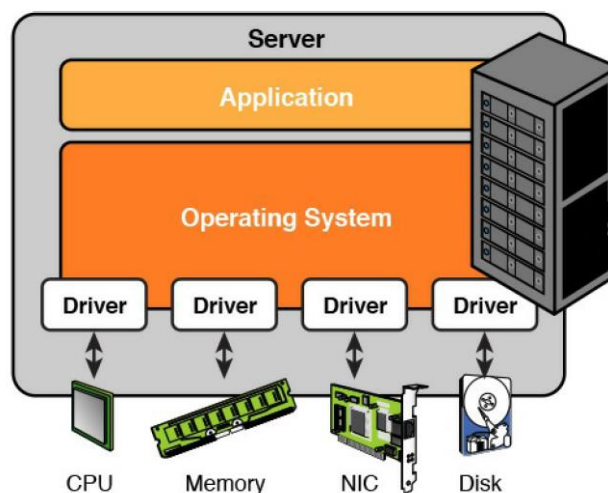
moved to, or instantiated in, various locations in the network as required, without the need for installation of new equipment.

A virtualização é uma tecnologia utilizada pelo mercado de TI há muitos anos, oferecendo uma camada de abstração dos verdadeiros recursos de *hardware* do sistema computacional e provendo um *hardware* virtual para cada sistema virtualizado. O objetivo disso é abstrair as características físicas e a forma que os sistemas operacionais e aplicações interagem com os recursos computacionais (LIMA; COSTA; SILVA, 2011).

A virtualização permite que recursos sejam divididos ou compartilhados em vários ambientes virtuais simultaneamente. Estes ambientes podem interagir entre si ou ser totalmente isolados uns dos outros e são conhecidos como Máquinas Virtuais ou, em inglês, *Virtual Machines* (VM). As VM's são ambientes que executam sistemas operacionais. Os sistemas operacionais que são executados nas VM's são conhecidos como sistemas convidados ou o seu termo em inglês *guests* (BUYTAERT et al., 2007).

Para melhor entendimento de um ambiente virtualizado, primeiramente é necessário entender como funciona um típico computador não virtualizado. Este possui um conjunto de dispositivos de *hardware* em que é instalado um sistema operacional (Windows, Linux, etc.), e uma ou mais aplicações são instaladas neste sistema operacional. A figura 26 ilustra este ambiente (MARSHALL; REYNOLDS; MCCRORY,2006).

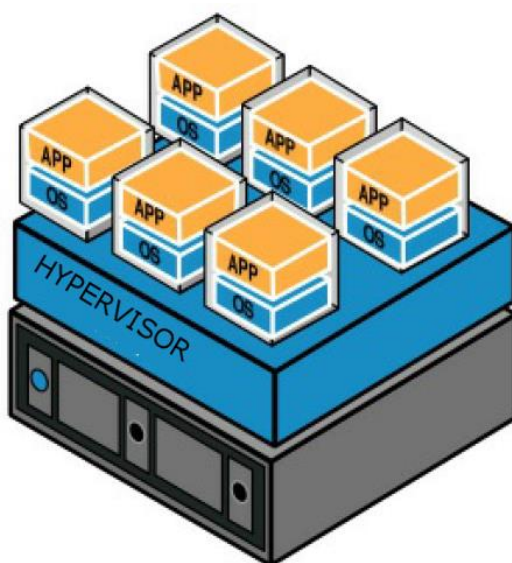
Figura 26 - Arquitetura de um sistema tradicional



Fonte: modificada de Doherty (2016, p.25)

Em um ambiente virtualizado, o arranjo é ligeiramente diferente, pois o computador possui uma plataforma de virtualização na qual um ou mais VM's são criados. Cada VM possui recursos de *hardware* definidos pelo usuário, possibilitando a instalação de um sistema operacional e aplicações conforme mostra a figura 27. Esta arquitetura é conhecida como arquitetura *hosted* (MARSHALL; REYNOLDS; MCCRORY, 2006).

Figura 27 - Arquitetura de um sistema virtualizado

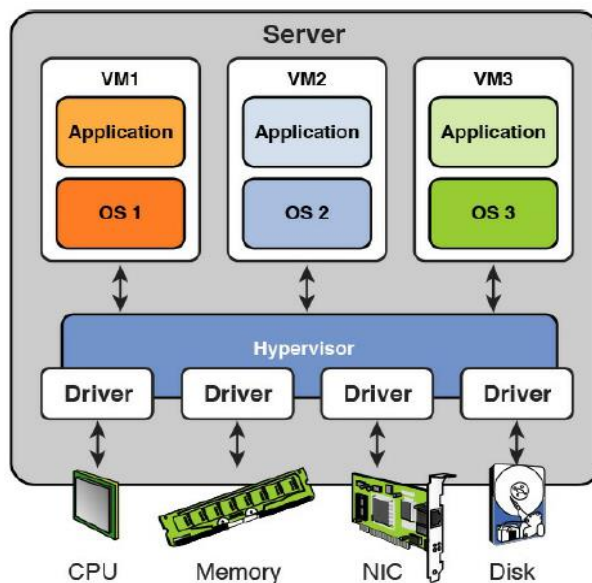


Fonte: modificada de Doherty (2016, p. 32)

A figura 27 ilustra o componente chamado *hypervisor*, conhecido também como *Virtual Machine Monitor* (VMM). Os *hypervisors* supervisionam e multiplexam uma variedade de sistemas operacionais usando algoritmos altamente eficientes e sofisticados. O *hypervisor* é uma camada de *software* bem isolada e necessária em um ambiente virtualizado. Esta camada deve trabalhar com um mínimo de recursos do sistema e possuir privilégios para a supervisão de todas as VM's. Os *hypervisors* definem e aplicam fortes políticas aos sistemas operacionais para as VM's operarem de forma cooperativa (REVELLE, 2011).

Em uma perspectiva de fácil compreensão, os *hypervisors* são uma camada de *software* que permite que os recursos de *hardware*, ou seja, memórias, discos de armazenamento, etc. compartilhem seus recursos entre VM's. Podemos considerar os *hypervisors* como um sistema operacional para sistemas operacionais. A figura 28 ilustra o *hypervisor*.

Figura 28 - Hypervisor agindo como interface entre os recursos de hardware e os VM's



Fonte: Doherty (2016, p. 27)

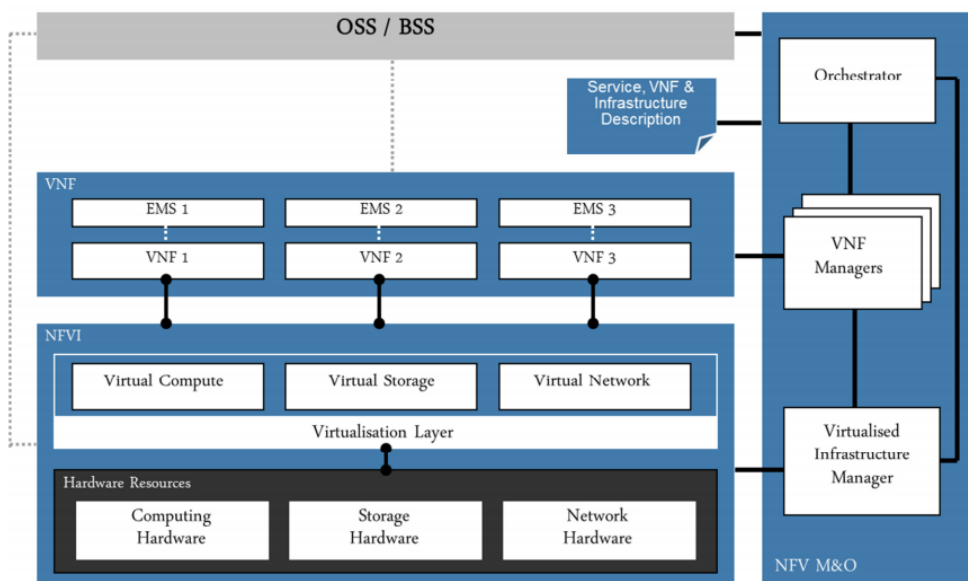
Voltando-se a focar no NFV, com a introdução da tecnologia de virtualização na indústria das telecomunicações, o provisionamento de serviços da maneira que ocorre atualmente sofre uma grande alteração. As principais mudanças com a introdução do NFV são:

- Desacoplar o *software* do *hardware*: como o elemento de rede não é mais uma coleção de entidades integradas de *hardware* e *software*, a evolução de ambos é independente um do outro. Isso permite que o *software* avance separadamente do *hardware* e vice-versa.
- Implementação de uma rede flexível: a separação entre *software* e *hardware* ajuda a redistribuir e compartilhar recursos de infraestrutura. Desta forma, o *hardware* e o *software* podem executar diferentes funções em vários momentos.
- Operação dinâmica: o desacoplamento da funcionalidade da função de rede em componentes de *software* instanciáveis proporciona maior flexibilidade para dimensionar o desempenho real de uma função de rede específica, de forma mais dinâmica e de acordo com a demanda de tráfego. Por exemplo, a operadora

precisa provisionar mais capacidade em determinado período do dia.

Para a aplicação das premissas levantadas anteriormente, o NFV ISG padronizou uma arquitetura que suporta a aplicação do NFV em diversos contextos da indústria de telecomunicações. Essa arquitetura é ilustrada na figura 29.

Figura 29 - Arquitetura do NFV



Fonte: ETSI (2013, p. 11)

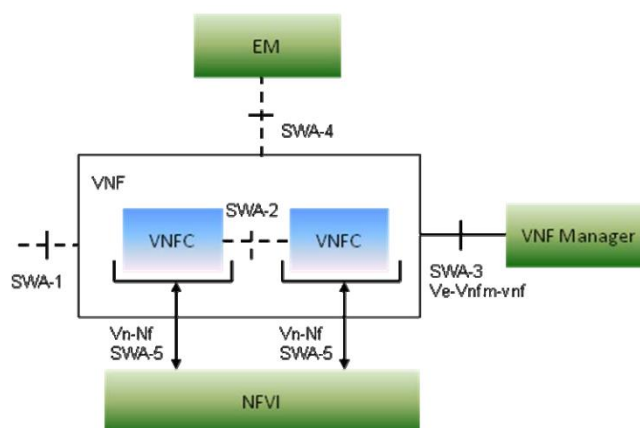
Os principais componentes da arquitetura do NFV são:

- *Virtual Network Functions (VNF):* é a virtualização de uma ou mais funções da rede. Um exemplo de funções da rede são as entidades padronizadas pelo 3GPP como, por exemplo, um MME. O comportamento funcional desta entidade é independente do uso da virtualização. Isto significa que o MME continuará realizando as suas funções de rede padronizadas pelo 3GPP como descritas no capítulo 3. A VNF pode ser acompanhada por um *Element Management System (EMS)*. O EMS executa a funcionalidade de gerenciamento de uma ou várias VNF's e também coordena o gerenciamento dessas funções entre os VM's subjacentes (ETSI GS NFV 002).

- O *NFV Infrastructure* (NFVI) fornece os recursos necessários para suportar a execução das funções de rede virtualizadas. Os *hardwares* de processamento, armazenamento e rede que fornecem conectividade e recursos computacionais à VNF são abstraídos através de uma camada de virtualização (por exemplo, um *hypervisor*). Com isso, há um desacoplamento entre os recursos de *hardware* e o *software* da VNF. Uma VNF pode ser implantada em um ou vários VM's (ETSI GS NFV 002).
- *NFV Management and Orchestration* (NFV M&O): abrange o gerenciamento e a orquestração dos recursos físicos e/ou dos *softwares* que suportam o NFVI e o gerenciamento das VNF's. O NFV M&O se concentra nas tarefas de gerenciamento específicas da virtualização (ETSI, 2013).
- O NFV M&O também interage com o *Operations Support Systems* (OSS) e o *Business Support Systems* (BSS). OSS são sistemas que suportam a operação da rede de forma geral, e BSS são sistemas que fornecem suporte ao negócio. Desta forma, o NFV pode ser integrado aos sistemas de gerenciamento da rede existentes. Como o NFV tem uma proposta de tornar a rede mais flexível, deixando mais fácil a introdução de novos serviços, essa interação com o OSS/BSS permite à operadora visualizar uma nova oportunidade de negócio e rapidamente fazer a introdução da mesma (ETSI, 2013).

É importante ressaltar que a VNF foi definida como a virtualização de uma função da rede com as mesmas interfaces e comportamento que foram padronizados anteriormente. A figura 30 ilustra a arquitetura interna do VNF, composta por uma ou múltiplas *VNF Components* (VNFC). O VNFC é uma entidade de *software* produzida por fornecedores de VNF.

Figura 30 - Arquitetura do VNF



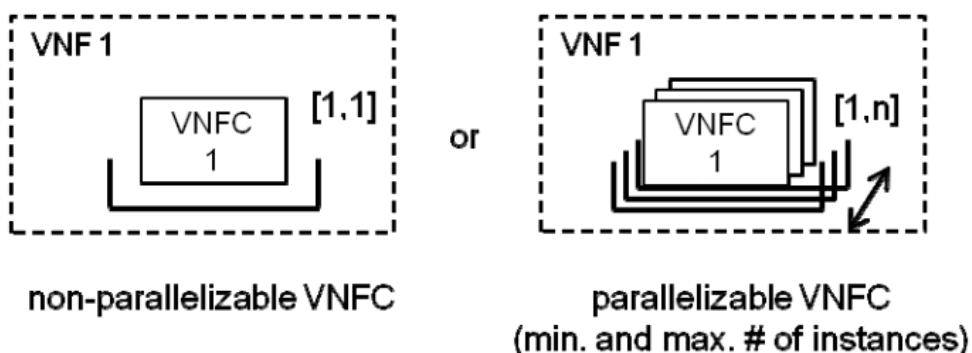
Fonte: ETSI GS NFV SWA 001 (2014, p. 14)

Uma maneira fácil de compreender o funcionamento do VNFC é considerar o mesmo como uma solução de *software* produzida pelos fornecedores. Um fornecedor pode criar uma VNF com apenas uma entidade de rede como, por exemplo, um PCRF, mas também pode criar uma VNF com todas as entidades do EPC juntas.

Uma VNF com todas as entidades do EPC pode ser decomposta em vários VNFC's. Cada VNFC pode ser uma entidade isolada como um MME ou um grupo de entidades como um SGW e um PDN GW. Um VNFC pode também conter todas as entidades do EPC. A decisão de como isto ficará implementado dependerá do fornecedor e será baseada em qual aplicação ou serviço a VNF suportará. A comunicação entre os VNFC's é feita através da interface SWA-2 (ETSI GS NFV SWA 001, 2014).

O VNFC pode ser instanciado paralelamente com as mesmas características múltiplas vezes. A limitação do número de vezes que o VNFC poderá ser instanciado irá depender da capacidade do *hardware* virtual no qual o VNFC está sendo executado ou do número de instâncias que o VM suporta. Desta forma, o NFV M&O pode criar várias instâncias de VNFC caso isto seja necessário. Um exemplo disso é o aumento abrupto de chamadas de dados: rapidamente o NFV M&O pode instanciar vários VNFC's para responder a nova demanda (ETSI NFV SWA 001, 2014). Isso é mostrado na figura 31.

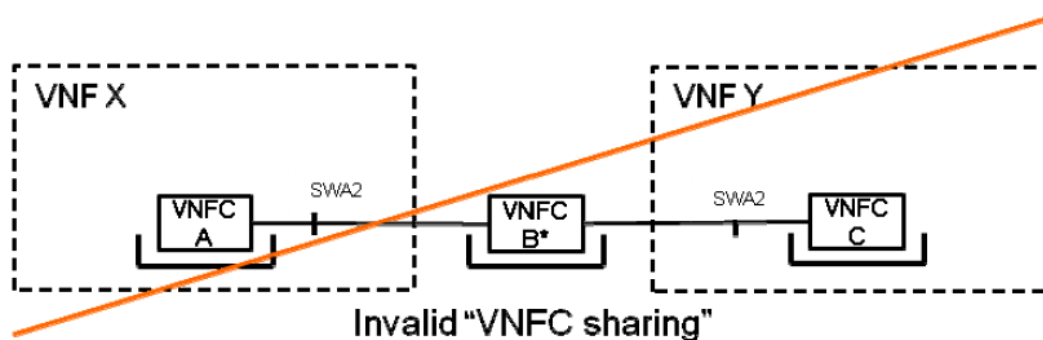
Figura 31 - Instância de VNF



Fonte: ETSI GS NFV-SWA 001(2014, p. 21)

Um sistema de uma operadora pode possuir diversas VNF's com múltiplos VNFC's. Porém, um VNFC não pode ser compartilhado entre VNF's. A figura 32 ilustra a ideia, deixando claro que o VNFC é uma classe que pertence somente a um VFN (ETSI NFV SWA 001, 2014).

Figura 32 - Invalidando o compartilhamento de VNFC

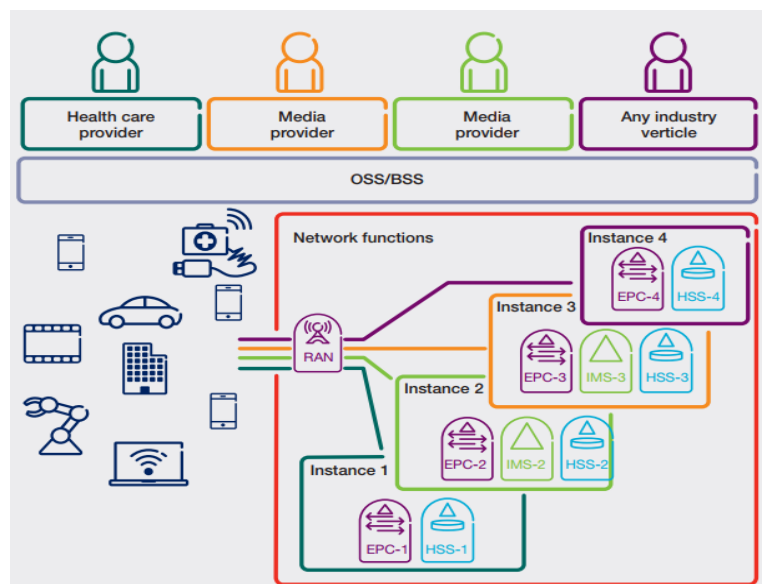


Fonte: ETSI GS NFV-SWA 001 (2014, p. 26)

Outro aspecto importante a ser comentado é o fatiamento da rede (*network slice*) na virtualização. Foi explicado como uma VNF pode possuir múltiplos VNFC's e, com isto, é possível criar um conjunto de funções diferentes e separadas logicamente. Quando um conjunto de VNF's é configurado para determinado uso, isto é chamado de *network slice* ou instância lógica. Assim, a operadora pode criar diversos tipos de VNF's para determinados tipos de usuários ou determinados tipos de serviços. A figura 33 ilustra a ideia de instâncias separadas logicamente com diversos tipos de componentes. Por exemplo, uma operadora que oferece VoLTE pode configurar uma instância com

EPC e IMS para este serviço e outra instância apenas com EPC para serviços de navegação na web (ERICSSON, 2015).

Figura 33 - Network Slice e suas aplicações



Fonte: Ericsson (2015, p. 6)

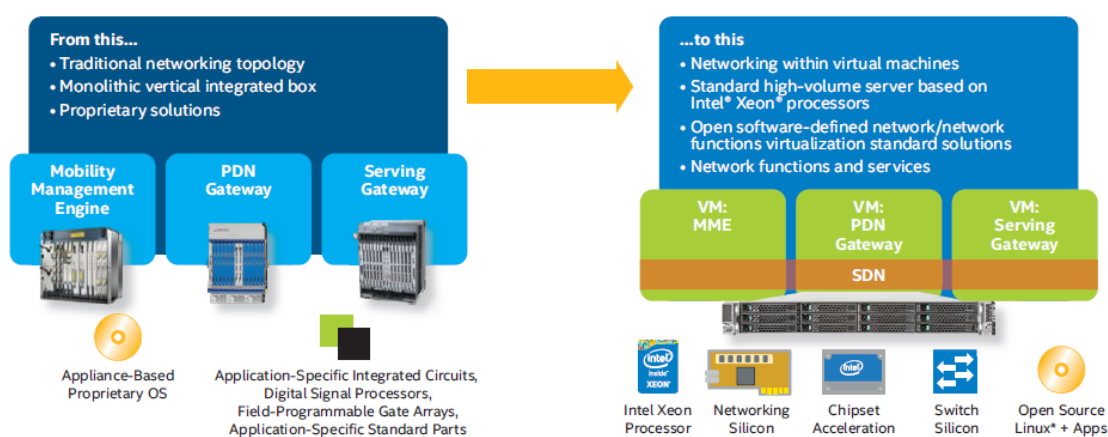
A virtualização do EPC visa reduzir a complexidade da rede e problemas relacionados à operação dos equipamentos. Com o EPC virtualizado, a operadora ganha mais flexibilidade no dia-a-dia, desacoplando *hardware* e *software*, ganhando uma eficiência nunca vista anteriormente. Em casos de demandas inesperadas como uma grande catástrofe natural, a rede tende a ter um aumento abrupto na demanda de serviços de voz. Com a virtualização, rapidamente a rede poderia criar recursos para tais demandas (ETSI GS NFV 001,2014). As vantagens da virtualização do EPC são:

- Melhoria da eficiência na utilização da rede devido à alocação flexível das funções da rede: isto acontece porque os recursos de hardwares estão em *pool* (ETSI, 2013).
- Elasticidade: uma função da rede pode ser dinamicamente modificada de acordo com a carga real na rede, e com isso, aumentar a escalabilidade da rede (ETSI, 2013).
- Reconfiguração de topologia: a topologia da rede pode ser reconfigurada dinamicamente para otimizar o desempenho (ETSI, 2013).

- Criação de uma rede que ofereça inovação e expansão de serviços, de forma mais rápida e acessível (NOKIA, 2015).
- Redução do custo da aquisição de equipamentos: os equipamentos de TI são mais acessíveis economicamente (NOKIA, 2015).
- Aumento na velocidade do *Time to Market*: as operadoras podem inovar em serviços de maneira mais rápida e prática. Atualmente um estudo para a implantação de algum serviço ou expansão da rede é demorado e custoso e depende de mão de obra especializada (ETSI, 2013).
- Redução do consumo de energia através do gerenciamento de recursos de energia de servidores: por exemplo, baseando-se em técnicas de virtualização, seria possível concentrar a carga de trabalho em um número menor de servidores durante as horas em que a demanda é menor como, por exemplo, durante a noite. Assim, os outros servidores podem ser desligados ou mudar para o modo mais econômico de energia (ETSI, 2013).

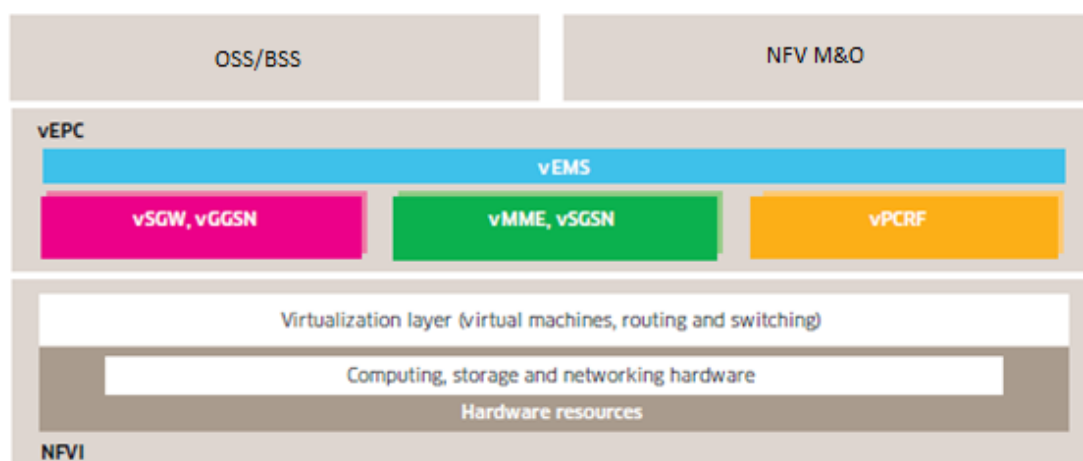
As figuras 34 e 35 mostram de forma sintetizada o que é virtualizar o EPC, eliminando-se *hardwares* de alto custo e promovendo-se todas as vantagens técnicas desta tecnologia.

Figura 34 - Virtualização do EPC



Fonte: Intel (2015, p. 3)

Figura 35 - Arquitetura do vEPC baseada na arquitetura do NFV



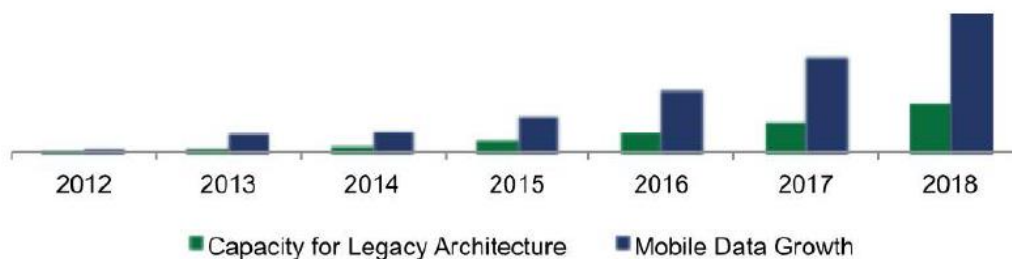
Fonte: modificada de Acatel Lucent (2014, p. 5)

Foram abordados benefícios técnicos da virtualização, mas também devem-se levar em conta os benefícios econômicos. Como mencionado anteriormente, a Ericsson faz projeções de que até 2020 o tráfego de vídeo na rede serão 75% do tráfego total. O crescimento do VoLTE já uma realidade em muitos países, e o crescimento do tráfego de redes sociais é uma realidade em praticamente todo o mundo (ERICSSON, 2015).

Ainda em previsões feitas pela Ericsson, em 2022 são esperados 29 bilhões de dispositivos conectados à rede, e destes 29 bilhões, aproximadamente 18 bilhões serão dispositivos IoT. Entre eles incluem-se carros, máquinas, medidores, sensores e terminais de ponto de venda (exemplo: máquinas de venda por cartão de crédito e débito) (ERICSSON, 2015).

A figura 36 ilustra o crescimento dos dados trafegados na rede e a capacidade da rede sem a virtualização. Fica claro que a rede sem a virtualização não conseguirá suportar o tráfego esperado. A estratégia tradicional de implantação se mostra insustentável.

Figura 36 - Crescimento dos dados na rede e arquitetura clássica de implementação



Fonte: Affirmed Networks and Cisco VNI (2015, p. 4)

A virtualização é de grande importância para as operadoras. A AT&T anunciou que até 2020, 75% da sua rede estarão virtualizados (ENTERPRISE NETWORK PLANET, 2016). As grandes operadoras do mundo já estão realizando testes de virtualização em suas redes (ACG RESEARCH, 2015).

A ACG Research analisou operadoras com a abordagem tradicional de implantação e operadoras que virtualizarão suas redes. O estudo analisa um período de cinco anos (2015-2019) com os seguintes resultados:

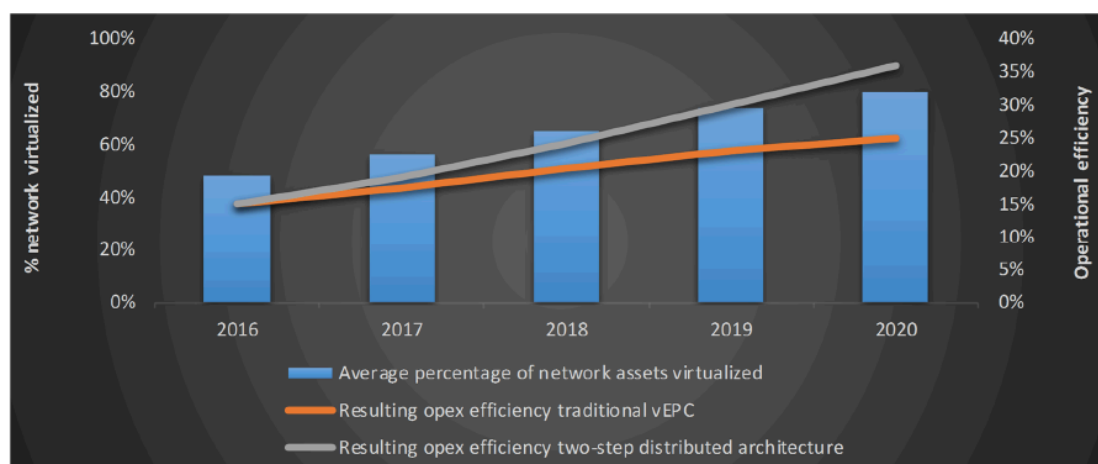
- Mudar para vEPC reduziu as despesas operacionais das operadoras em 67% em comparação com as operadoras que continuaram com a abordagem tradicional de implantação (ACG RESEARCH, 2015);
- As operadoras que mudaram para uma plataforma baseada em NFV começaram a economizar recursos no primeiro ano de implantação. Três anos após a implantação, os custos foram pagos (ACG RESEARCH, 2015).
- As operadoras que adotaram a virtualização da rede foram capazes de implementar novos serviços com mais agilidade (menos de 6 meses), enquanto as operadoras que usaram uma abordagem tradicional demoraram em média 15 meses (ACG RESEARCH, 2015).
- Aceleração do *Time-to-Market* para novos serviços em 10 meses em comparação com redes não virtualizadas (ACG RESEARCH, 2015).

- Fornecimento de uma melhor qualidade de experiência para os clientes com uma grande capacidade da rede (ACG RESEARCH, 2015).

A IDC Analyze The Future realizou um estudo que aponta os benefícios econômicos do vEPC. Para isso, foram utilizadas duas abordagens. O primeiro passo é a virtualização do EPC, mas que continua implementado de forma centralizada. E o segundo passo é a virtualização do EPC e a distribuição de *datacenters* na extensão territorial. Essa distribuição ocorre a partir do segundo ano de implantação do vEPC.

As projeções obtidas pela IDC foram: a implantação sem a distribuição da rede aumentou a eficiência dos custos de operação em 25% em cinco anos, e após cinco anos de implantação houve uma economia total de 20%. Com a implantação de forma distribuída, houve um aumento na eficiência dos custos de operação em 36%, e após cinco anos de implantação houve uma economia total de 30%. A figura 37 ilustra os resultados obtidos.

Figura 37 - Eficiência dos custos de 2016 até 2020



Fonte: IDC (2016, p. 6)

O vEPC é uma tecnologia que, como o que foi mostrado, pode oferecer vários benefícios às operadoras, mas há importantes desafios técnicos e operacionais que devem ser superados. Os principais desafios das operadoras para a implantação de um vEPC são:

- **Portabilidade/interoperabilidade:** a capacidade de executar funções da rede virtuais em diferentes *datacenters*, produzidos por diferentes fornecedores para diferentes operadoras: o desafio é definir um *hypervisor* que desacopla claramente os *softwares* da VNF e o *hardware* do NFVI. O vEPC deve coexistir com outros equipamentos tradicionais já implantados na rede da operadora (ETSI, 2013).
- **Performance:** com o desacoplamento do *hardware* e do *software*, uma provável diminuição no desempenho deve ser levada em consideração. Os *hypervisors* devem escalonar os recursos de *hardwares* para manter o desempenho que os equipamentos atuais possuem (ETSI, 2013).
- **Simplicidade:** Assegurar que as plataformas de rede virtualizadas serão mais fáceis de operar do que as que existem hoje (ETSI, 2013).
- **Garantir o nível adequado de resiliência às falhas de *hardware* e *software*** (ETSI, 2013).

Outra tendência em conjunto com o NFV, é o *Software Defined Network* (SDN), que representa uma nova abordagem em tecnologia de rede, projetada para melhorar a natureza complexa e estática das arquiteturas de redes distribuídas tradicionais. O conceito de SDN é baseado na arquitetura dividida, em que o plano de controle é separado da infraestrutura física subjacente e fornece uma abstração à rede. Uma vez que todas as funções de controle são movidas para um controlador programável, os administradores podem facilmente alterar o comportamento da rede em tempo real e adaptá-la às condições necessárias, executando aplicativos de controle apropriados. Com o SDN e o NFV, a rede ganha uma grande flexibilidade e simplicidade na sua operação, e passamos para uma nova era no mercado de telecomunicações (TOMOVIC; PEJANOVIC-DJURISIC; RADUSINOVIC, 2014).

6. CONCLUSÃO

Este trabalho teve como objetivo discorrer sobre a virtualização do núcleo da tecnologia LTE. Para isso, foi inicialmente estudado o LTE e, posteriormente, detalhado o núcleo da rede, chamado de *Evolved Packet Core* (EPC). O entendimento do LTE criou a base de fundamentos necessária para analisar o EPC virtualizado (vEPC), explorando a mudança da arquitetura, formas de como ocorrerá a implantação da virtualização, benefícios técnicos e econômicos e os desafios a serem batidos.

Os resultados obtidos com esta análise foram de maneira geral bastante vantajosos no contexto de uma operadora de telefonia móvel. Foi mostrado que com o desacoplamento do *hardware* e *software* e a implantação das funções da rede em *datacenters*, espera-se uma economia de custos relacionados à operação da rede. A virtualização promete uma melhor elasticidade da rede, adequando a demanda de serviços de acordo com períodos de tempo e até catástrofes naturais.

Com o vEPC, torna-se mais fácil administrar e implantar novos serviços. Isso vem do fato de que com a virtualização é possível criar fatias logicamente isoladas para cada serviço, tornando a administração da rede menos complexa. Essa vantagem de ofertar novos serviços rapidamente se estende ao mercado, permitindo que as operadoras possam inovar rapidamente.

Porém, há importantes desafios para adequar a tecnologia de TI dos *datacenters* às necessidades de uma operadora de telefonia móvel. Os equipamentos precisam possuir uma robustez elevada, e há a necessidade do desenvolvimento de uma camada *hypervisor* que seja capaz de abstrair os recursos de hardware sem perdas de desempenho para adquirir a melhor eficiência do sistema. Além disso, um dos maiores desafios será integrar os *datacenters* aos equipamentos existentes nas redes das operadoras.

Com isso, conclui-se que a virtualização do EPC é a principal maneira de economizar gastos, criar novos serviços e facilitar o gerenciamento da rede. Mas, não deixará de ser um grande desafio para as operadoras essa mudança de tecnologia, pois pode-se dizer que será uma mudança de paradigmas no mercado de telecomunicações. A virtualização do EPC apresenta-se como o futuro imediato para a indústria móvel, trazendo grandes vantagens técnicas e econômicas e preparando o núcleo de rede para o tão esperado 5G.

6.1 DIFICULDADES ENCONTRADAS

A principal dificuldade encontrada foi a escassez de literatura relacionada ao EPC e à virtualização, problema minimizado com documentos compartilhados pela orientadora. Outro problema é a língua, pois todos os documentos estão em inglês, e em virtude da linguagem técnica, torna-se por vezes difícil a tradução.

6.2 TRABALHOS FUTUROS

A virtualização do EPC abre margem para estudos em diversas áreas de TI e telecomunicações. Cada área mencionada tem desafios a serem superados. Há ainda muito a ser estudado, levando-se em conta as diversas fases da implementação do vEPC, e sendo importante observar que cada fase apresenta uma série de peculiaridades a serem detalhadas. Propostas de estudos futuros são o desempenho da infraestrutura de *datacenters* para redes de telecomunicações, o gerenciamento da rede virtualizada e a evolução do vEPC para o 5G.

REFERÊNCIAS

3GPP. High Speed Packet Data Access (HSPA). Disponível

em: <<http://www.3gpp.org/technologies/keywords-acronyms/99-hspa>> . Acesso em: 10 de Outubro de 2017.

3RD GENERATION PARTNERSHIP PROJECT. 3GPP TS 23 002. Disponível em:

<http://www.etsi.org/deliver/etsi_ts/123000_123099/123002/12.05.00_60/ts_123002v120500p.pdf>. Acesso em: 12 de Outubro de 2017;

_____. **3GPP TS 23.203 V10.6.0 (2012-03).** Disponível em:

<<http://www.qtc.jp/3GPP/Specs/23203-a60.pdf>> Acesso em: 01 de outubro de 2017.

_____. **3GPP TS 23.401 V8.0.0 (2007-12).** Disponível em:

<<http://www.qtc.jp/3GPP/Specs/23401-800.pdf>> Acesso em: 30 de setembro de 2017.

_____. **3GPP TS 24 301.** Disponível em:

<http://www.etsi.org/deliver/etsi_ts/124300_124399/124301/10.03.00_60/ts_124301v100300p.pdf>. Acesso em: 20 de Setembro de 2017.

_____. **3GPP TR 125 913,v.8 , n.1, 2009.** Disponível em:

<http://www.etsi.org/deliver/etsi_tr/125900_125999/125913/08.00.00_60/tr_125913v080000p.pdf> . Acesso em: 19 de Setembro de 2017.

_____. **3GPP TS 29 212.** Disponível em:

<http://www.etsi.org/deliver/etsi_ts/129200_129299/129212/12.06.00_60/ts_129212v120600p.pdf> . Acesso em: 13 de Setembro de 2017.

_____. **3GPP TS 33.106 2011.** Disponível em:

<http://www.etsi.org/deliver/etsi_ts/133100_133199/133106/10.00.00_60/ts_133106v100000p.pdf>. Acesso em: 20 de Novembro de 2017.

_____. **3GPP TS 36 414**. Disponível em:
<http://www.etsi.org/deliver/etsi_ts/136400_136499/136414/12.01.00_60/ts_136414v120100p.pdf>. Acesso em: 20 de Outubro de 2017.

_____. **3GPP TS23.401**. Disponível em: <<http://www.qtc.jp/3GPP/Specs/23401-800.pdf>>. Acesso em: 20 de Outubro de 2017

ACG RESEARCH. **Total Cost of Ownership Study Virtualizing the Mobile Core**. Sponsored by Affirmed Networks and VMWare. 2015.

ALCATEL LUCENT. **alcatel-lucent evolved packet core solution: delivering technical innovation for the new lte mobilecore**. Disponível em:
<http://www.pexx.net/pdfs/whitepapers/alcatel_lucent/mpr9500/EPC_Solution_wp_0309.pdf>. Acesso em: 10 de Outubro de 2017.

ALECRIM, Emerson. **Tecnologias 2g e 2,5g: TDMA, CDMA, GSM, GPRS E EDGE**. Disponível em: <<https://www.infowester.com/2g.php>> . Acesso em: 17 de Setembro de 2017.

ANATEL. **Cobertura de redes 4g no brasil**. Disponível em:
<<http://www.anatel.gov.br/dados/destaque-1/218-destaque-6>>. Acesso em: 10 de Setembro de 2017.

BEC TECHNOLOGIES. **Quality of Service (QoS) in LTE**. Disponível em:
<http://www.bectechnologies.net/main/newsletter_images/QoS.pdf>. Acesso em: 22 de Outubro de 2017.

BUYTAERT, Kris; DITTNER, Rogier; GARCIA Juan R; et al.; **The Best Damn Server Virtualization**. Book Period. Syngress Publishing Inc, 2007.

CISCO. **Implementing Cisco quality of service**. p. 23. Disponível em:
<<http://www.itsolutions.pro/images/stories/docs/implementingciscoqualityofserviceqosv2.2volumes12.pdf>>. Acesso em: 13 de Setembro de 2017.

_____. **Mobility Management Entity Overview**. Disponível em:
<https://www.cisco.com/c/en/us/td/docs/wireless/asr_5000/20/MME/b_20_MME_Admin/b_20_MME_Admin_chapter_01.pdf>. Acesso em: 20 de Setembro de 2017.

_____. **PDN Gateway Overview**. Disponível em:
<https://www.cisco.com/c/en/us/td/docs/wireless/asr_5000/20/P-GW/b_20_PGW_Admin/b_19_PGW_Admin_chapter_01.pdf> . Acesso em 13 de Setembro de 2017.

_____. **Serving Gateway Overview**. Disponível em:
<https://www.cisco.com/c/en/us/td/docs/wireless/asr_5000/20/SGW/b_20-SGW-Bookmap/b_20-SGW-Bookmap_chapter_01.pdf>. Acesso em: 29 de Outubro de 2017.

DENGO, Kan. **GSM-GPRS Architecture**. Disponível em: <<https://www.flickr.com/photos/35547713@N02/4239683146/in/photostream/>>
Acesso: 06 de setembro de 2017.

DOHERTY, Jim; **SDN and NFV simplified** Pearson Education; Inc ISBN-13: 978-0-13-430640-7. 2016.

ERICSSON. **Policies Protect Profits**. p. 7. Disponível em:
<<https://archive.ericsson.net/service/internet/picov/get?DocNo=8/28701>>.
Acesso em: 25 de Outubro de 2017.

_____. **Voice and Video Calling Over LTE**. Disponível em:
<<https://www.ericsson.com/assets/local/publications/white-papers/wp-voice-and-video-calling-over-lte.pdf>>. Acesso em 22 de Outubro de 2017.

_____, **Ericsson Mobility Report**. Disponível em:
<<https://www.ericsson.com/assets/local/mobility-report/documents/2017/ericsson-mobility-report-june-2017.pdf>> Acesso em: 15 de novembro de 2017

_____. **Gearing up support systems for software defined and virtualized networks**. Ericsson review: The communications technology journal since 1924. 2015

_____. **Network Functions Virtualization and software management**. Ericsson White paper. 2014

EUROPEAN COMPUTER MANUFACTURERS ASSOCIATION. ISO 7498-1.

Disponível em: <[https://www.ecma-](https://www.ecma-international.org/activities/Communications/TG11/s020269e.pdf)

[international.org/activities/Communications/TG11/s020269e.pdf](https://www.ecma-international.org/activities/Communications/TG11/s020269e.pdf)> . Acesso em:

10 de Outubro de 2017.

EUROPEAN TELECOMMUNICATIONS STANDARDS INSTITUTE – ETSI.

Network Functions Virtualisation (NFV). SDN and OpenFlow World

Congress. 2013.

GEOCITIES. **Telefonia Móvel**. Disponível em:

<<http://www.geocities.ws/redescefetpi/feitos/telefonica/telefmovel.html>>. Acesso em 10 de Setembro de 2017.

GHOSH, Aruanabha; ZHANG, Jun; ANDREWS, Jeffrey G.; MUHAMED, Rias.

FUNDAMENTALS OF LTE. ed. Pearson Education, p.464, 2010

GOOGLELLC. **Adoção do Ipv6 no Brasil**. Disponível em:

<<https://www.google.com/intl/pt-BR/ipv6/statistics.html#tab=ipv6-adoption&tab=ipv6-adoption>>. Acesso em: 31 de Outubro de 2017.

GSMA INTELLIGENCE. **GLOBAL MOBILE TRENDS – 2017**. Disponível em:

<<https://www.gsmaintelligence.com/research/?file=3df1b7d57b1e63a0cbc3d585feb82dc2&download>>. Acesso em: 18 de Setembro de 2017.

GUEDES, Luís C. dos S.; VASCONCELOS, Renan R.; DUARTE, Otto. **UMTS,**

HSPA E LTE. Dissertação (Trabalho de Conclusão de Curso) - Faculdade de

Redes de Computadores. Universidade Federal do Rio de Janeiro, Rio de

Janeiro, 2009

INTEL. **Virtualized Evolved Packet Core**. NFV Strategies: Global Service Provider. 2015.

ITU. **Mobile-cellular Subscriptions**. Disponível em: <<http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx>>. Acesso em: 20 de Setembro.

LAMMEL, Iuri. **dez anos de smartphone: a revolução em nossas vidas**. Disponível em: <<http://diariodesantamaria.clicrbs.com.br/rs/economia-politica/noticia/2017/07/dez-anos-de-smartphone-a-revolucao-em-nossas-vidas-9845275.html>>. Acesso em: 25 de Outubro.

LIMA, Cristiano J.; COSTA, Danilo M.; SILVA, Cláudio L. H.; **Virtualização de servidores: um estudo de caso na empresa sênior engenharia**. Revista Pensar Tecnologia, v. V1, p. 116-136, 2012.

MARSHALL; REYNOLDS, Wade A.; MCCRORY, Dave.; **Advanced server virtualization – VMware and Microsoft Platforms in the Virtual Data Center**. Auerbach Publications, 2006.

MAZZONI, Victor de Souza. **Análise Histórica e Funcional das Redes 4G LTE**, 2014, Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas, Departamento de Ciência da Computação, Bacharelado em Sistemas de Informação.

NETMANIAS. **LTE QoS: SDF and EPS Bearer QoS**. Disponível em: <<https://www.netmanias.com/en/?m=view&id=techdocs&no=10434&xtag=eps-lte-qos-sdf&xref=lte-qos-sdf-and-eps-bearer-qos>> Acesso em: 13 de novembro de 2017

NETWORK FUNCTIONS VIRTUALIZATION INDUSTRY SPECIFICATION GROUP. **ETSI GS NFV 001- Network Functions Virtualization (NFV): Use cases**. 2013.

_____. **ETSI GS NFV-SWA 001 - Network Functions Virtualization (NFV): Virtual Network Functions Architecture.** 2014.

_____. **ETSI GS NFV 002 - Network Functions Virtualization (NFV): Architectural Framework.** 2013.

NOKIA. **Nokia virtualized EPC Delivering on the promise of NFV and SDN.** whitepaper. 2015.

OLSSON, Magnus; SULTANA, Shabnam; ROMMER, Stefan; FRID, Lars; MULLIGAN, Catherine. **SAE and the evolved packet core: driving the mobile broadband revolution.** 2009. Book Academic Press ©2009 ISBN:0123748267 9780123748263.

PARSONS, Christopher. **Deep Packet Inspection in Perspective.** tracing its lineage and surveillance potentials outubro 2008. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.513.3177&rep=rep1&type=pdf>>. Acesso em: 04 de Outubro de 2017.

POIKSELKÄ, Miikka; HOLMA, Harri; HONGISTO, Jukka; KALLIO, Juha; TOSKALA. **voice over lte: volte.** ISBN: 978-1-119-95168-1 266.

REVELE, Don.; **Hypervisors and Virtual Machines Implementation Insights on the x86 Architecture.** ; login: The Usenix Magazine V. 36, e.5 2011.

SANTOS, Gustavo Mendes. **LTE: Soluções de Voz II.** 17 de Fevereiro de 2017. Disponível em: <<http://www.teleco.com.br/tutoriais/tutorialltevoz2/default.asp>>. Acesso em: 12 de Outubro de 2017.

SINGH, Amandeep. **Overview of the Evolved Packet Core Network.** p. 5-22. Disponível em: <<http://ece.engineering.ualberta.ca/Graduate/~~/media/ece/Graduate/Documents>

/GradDocs/Intro_to__Evolved_packet_core_network.pdf> . Acesso em 03 de Outubro de 2017.

SRIDHAR, Kamakshi. **INTRODUCTION TO EVOLVED PACKET CORE (EPC): EPC elements, protocols and procedures.** Disponível em: <<http://www.cvt-dallas.org/Aug12-Sridhar.pdf>>. Acesso em: 28 de Outubro de 2017.

TAKAKI R.; BAZZO, J. J. **voz para rede LTE:** Cad. CPqD Tecnologia, Campinas, v. 8, n. 2, p. 27-32, jul./dez. 2012.

TELECO. **Tutorial de telefonia celular: roaming.** 18 de Novembro de 2002. Disponível em: <<http://www.teleco.com.br/tutoriais/tutorialroam/>>. Acesso em: 11 de Setembro de 2017.

TOMOVIC, Slavica.; PEJANOVIC-DJURISIC, Milica.; RADUSINOVIC , Igor,; **SDN Based Mobile Networks: Concepts and Benefits.** Wireless Pers Commun DOI 10.1007/s11277-014-1909-6. 2014.

TU, Guan-Hua; LI, Yuanjie; PENG, Chunyi; LI, Chi-Yu. WANG, Hongyi; LU, Songwu. **control-plane protocol interactions in cellular networks. slide 03.** Disponível em: <<http://slideplayer.com/slide/4194409/>> . Acesso em: 30 de Setembro de 2017.