

Diogo António da Fonseca Mendes

Licenciado em Ciências da Engenharia



Análise Prescritiva do Desempenho de Equipamentos de Redes

Dissertação para Obtenção do Grau de Mestre em
Engenharia Eletrotécnica e de Computadores

Orientador: Pedro Miguel Figueiredo Amaral, Professor Auxili-
liar, FCT-UNL

Júri:

Presidente: Prof. Doutor João Francisco Alves Martins – FCT/UNL

Arguente: Prof. Doutor Rodolfo Alexandre Duarte Oliveira –
FCT/UNL

Vogal: Prof. Doutor Pedro Miguel Figueiredo Amaral –
FCT/UNL



FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Setembro, 2017

Análise Prescritiva do Desempenho de Equipamentos de Rede

Copyright © Diogo António da Fonseca Mendes, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa.

A Faculdade de Ciências e Tecnologia e a Universidade Nova de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor

À minha avó, mãe, irmão e amigos

Agradecimentos

Em primeiro lugar, gostaria de agradecer ao meu orientador, Professor Doutor Pedro Miguel Figueiredo Amaral, pela sua disponibilidade ao longo da realização desta dissertação e por me ter aceite como seu orientando. Queria agradecer ao Engenheiro João Pedro Neves, orientador representante da NOS Comunicações, S.A., por todo o seu empenho e por me ter ajudado na integração da sua equipa.

Agradeço ao Departamento de Engenharia Eletrotécnica e de Computadores da FCT-UNL, a minha formação durante estes cinco anos. Um agradecimento especial ao Departamento de Operações da NOS que me abriu as portas para o mundo empresarial. Foi uma experiência bastante enriquecedora.

Gostaria de agradecer a todos os meus amigos por me terem proporcionado momentos únicos. Quando tudo parecia mais difícil, eles mostravam-me o mais fácil.

A toda a minha família, em especial à minha avó Maria, mãe e irmão. A realização deste percurso não teria sido possível, se não fosse o amor incondicional que temos uns pelos outros. Por isso, queria agradecer à minha mãe por ser um exemplar exímio de profissionalismo e por todo o sacrifício que ela fez por mim. Ao meu irmão, por tudo o que ele significa para mim e por ser a pessoa mais presente na minha vida. À minha querida avó Maria gostaria de agradecer por estar sempre a olhar por mim.

A todos, um obrigado de coração.

Resumo

Para as empresas de telecomunicações as redes sem fios suportam múltiplos sistemas de comunicações móveis de elevada importância económica. Face à elevada percentagem de subscritores que desistem do serviço de Internet e TV residencial suportado na rede móvel denominado *Direct-to-Home* (DTH), houve necessidade de realizar uma análise prescritiva do desempenho de equipamentos de rede, onde se estudou a qualidade da internet fixa residencial na rede móvel.

Através dos sistemas de monitorização da NOS é possível recolher dados de toda a rede em si, mas também dos terminais que suportam o serviço DTH contratado pelo cliente. Assim, é possível verificar a qualidade de rede e do serviço prestado pela empresa, e também ter conhecimento da experiência de utilização do cliente.

Nesta dissertação são realizadas abordagens de *data mining*, de maneira a extrair informação dos dados que possam justificar a taxa de desistência do serviço em questão, e ainda são apresentados diversos resultados que fundamentam as conclusões do trabalho.

Palavras-chave: DTH; Data Mining; Throughput; UMTS; LTE; QoS; Internet Fixa; Rede Móvel

Abstract

For the telecommunication companies, the wireless networks support multiple systems of mobile communications of high economic importance. The high percentage of consumers subscribing the DTH service, demand for a detailed analysis of multiple key performance indicator values. These include a prescriptive analysis of the network equipment's performance, in order to measure the effective quality of the fixed internet over the mobile network.

Through different data gathering tasks in multiple network systems at NOS it was possible to obtain data from the network itself, as well as from the terminals serving the client. This was the main goal of this work, which focused not only on the quality of the service provided by NOS, as well as the quality of users' experience.

In this dissertation, several approaches of data mining were adopted, in order to extract information from the data that could justify the dropout rate of the service provided by NOS. At the final stage of the dissertation, several results identify multiple reasons for the service dropout rate, which are presented and discussed in detail.

Keywords: DTH; Data Mining; Throughput; UMTS; LTE; QoS; Fixed Wireless; Mobile Network

Índice Geral

Agradecimentos	vii
Resumo	ix
Abstract	xi
Índice Geral	xiii
Índice de Figuras	xv
Índice de Tabelas	xvii
Acrónimos	xix
1 Introdução	1
1.1 Motivação.....	1
1.2 Objetivos	3
1.3 Principais Contribuições.....	4
1.4 Organização da Dissertação	4
2 Estado de Arte	7
2.1 Introdução.....	7
2.2 Comunicações Móveis	7
2.2.1 <i>Arquitetura GSM (2G)</i>	8
2.2.2 <i>Arquitetura UMTS (3G)</i>	11
2.2.3 <i>Arquitetura LTE (4G)</i>	12
2.3 Big Data	17
2.4 Data Mining.....	18
2.4.1 <i>Classificação de Sistemas de Data Mining</i>	19
2.4.2 <i>Pré-Processamento dos Dados</i>	20

3	Descrição do Projeto	23
3.1	Recolha de dados	23
3.2	Análise de dados	28
3.2.1	<i>Dimensões de Análise</i>	28
3.2.2	<i>Processamento de dados</i>	29
4	Validação e Resultados	33
4.1	Universo da análise	33
4.2	Resultados Obtidos	36
4.3	Evolução horária de throughput	57
5	Conclusões	61
5.1	Síntese do Projeto Realizado	61
5.2	Sugestões para Trabalho Futuro	63
	Referências	65

Índice de Figuras

Figura 2.2.1 - Gerações dos sistemas de comunicação móvel[5].	8
Figura 2.2.2 - Arquitetura de rede GSM [10].	9
Figura 2.2.3 - Evolução do sistema GSM[9].	11
Figura 2.2.4 - Arquitetura da rede UMTS [12].	12
Figura 2.2.5 - Arquitetura da rede LTE [17].	13
Figura 2.2.6 - Arquitetura geral da rede LTE EPS[9].	14
Figura 2.2.7 - Arquitetura LTE E-UTRAN [9].	15
Figura 2.2.8 - Funcionalidades de E-UTRAN e EPC [18].	17
Figura 2.4.1 - Data Mining como etapas no processo de descoberta de conhecimento [28].	19
Figura 2.4.2 - Data Mining como influência de múltiplas disciplinas [28].	20
Figura 2.4.3 - Formas de pré-processamento de dados [28].	21
Figura 3.1.1- Diagrama de recolha de dados.	24
Figura 3.2.1 - Tabela representativa da capacidade da célula, em dBm.	31
Figura 4.1.1 - Universo do estudo e Amostra analisada.	35
Figura 4.1.2 - Distribuição das células estudadas no projeto.	36
Figura 4.2.1 - Distribuição por Produto, Tecnologia.	37
Figura 4.2.2 - Distribuição Produto, Processo.	37
Figura 4.2.3 - Distribuição por Tecnologia.	38
Figura 4.2.4 - Mapeamento por Tecnologia.	39
Figura 4.2.5 - Distribuição por Fornecedor, Tecnologia.	40
Figura 4.2.6 - Mapeamento por Fornecedor.	40
Figura 4.2.7 - Distribuição por Terminal, Tecnologia.	41
Figura 4.2.8 - Mapeamento por Terminal.	42

Figura 4.2.9 - Velocidade por Terminal, Tecnologia.	43
Figura 4.2.10 - Velocidade por Terminal, Horário [3G].	44
Figura 4.2.11 - Velocidade por Terminal, Horário [4G].	44
Figura 4.2.12 - Velocidade por Terminal, Tx. Ocupação [3G].	45
Figura 4.2.13 - Velocidade por Terminal, Taxa Ocupação [4G].	46
Figura 4.2.14 - Velocidade por Produto, Tecnologia.	47
Figura 4.2.15 - Velocidade por tipo Teste, Tecnologia.	48
Figura 4.2.16 - Velocidade por Processo, Horário [4G].	49
Figura 4.2.17 - Velocidade por Processo, Horário [3G].	49
Figura 4.2.18 - Velocidade por Fornecedor, Tecnologia.	50
Figura 4.2.19 - Velocidade por Fornecedor, Taxa Ocupação [4G].	50
Figura 4.2.20 - Velocidade por Fornecedor, Taxa Ocupação [3G].	51
Figura 4.2.21 - Velocidade por Fornecedor, KPIs [4G].	52
Figura 4.2.22 - Velocidade por Fornecedor, KPIs [3G].	52
Figura 4.2.23 - Velocidade por Fornecedor, Horário [3G].	53
Figura 4.2.24 - Velocidade por Fornecedor, Horário [4G].	53
Figura 4.2.25 - Velocidade por Banda, Tecnologia.	54
Figura 4.2.26 - Velocidade por Taxa Ocupação, KPIs [4G].	55
Figura 4.2.27 - Velocidade por Taxa Ocupação, KPIs [3G].	55
Figura 4.2.28 - Velocidade por KPIs, Taxa Ocupação [3G].	56
Figura 4.2.29 - Velocidade por KPI, Taxa Ocupação [4G].	56
Figura 4.2.30 - Velocidade por Tecnologia, Horário.	57
Figura 4.3.1 - Evolução horária em 3G e número de testes efetuados por hora.	58
Figura 4.3.2 - Evolução horária em 4G e número de testes efetuados por hora.	58
Figura 4.3.3 - Comparação das velocidades médias obtidas em 3G e 4G.	59
Figura 4.3.4 - Mapeamento das velocidades médias em Portugal.	60

Índice de Tabelas

Tabela 3.1.1 - Característica do conjunto de dados KPI.	25
Tabela 3.1.2 - Características do conjunto de dados dos testes de velocidade.....	26
Tabela 3.1.3 - Recolha de dados: fontes e respetivos dados.....	26
Tabela 3.2.1 - Análises de estudo.	28
Tabela 3.2.2 - Tabela representativa da Taxa de Ocupação da célula.....	32
Tabela 4.1.1 – Dimensões sobre o Universo de análise	34

Acrónimos

3GPP	Third Generation Partnership Project
ACS	Access Control System
AMPS	Advanced Mobile Phone System
AuC	Authentication Centre
BSC	Base Station Controller
BSS	Base Station Subsystem
BTS	Base Transceiver Station
CDMA	Code-Division Multiple Access
CN	Core Network
CPM	Customer Premise Management
DTH	Direct-To-Home
EN	External Network
EDGE	Enhanced Data for GSM Evolution, Enhanced Data Rates for Global
eNodeB	E-UTRAN NodeB
E-UTRAN	Evolved UTRAN
EPC	Evolved Packet Core
EPS	Evolved Packet System
FFTH	Fiber From The Home
FTH	File Transfer Protocol
GGSN	Gateway GPRS Support Node

GIS	Geographic Information System
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communications
GW	Gateway
HFC	Hybrid Fiber Coax
HLR	Home Location Register
HTTP	Hypertext Transfer Protocol
HSDPA	High Speed Downlink Packet Access
HSS	Home Subscriber Server
HSUPA	High Speed Uplink Packet Access
IMSI	International Mobile Subscriber Identity
IP	Internet Protocol
KDD	Knowledge Discovery from Data
KPI	Key Performance Indicator
LAN	Local Area Network
LTE	Long-Term Evolution
MME	Mobility Management Entity
MS	Mobile Station
MSC	Mobile Switching Center
MSISDN	Mobile Station Subscriber Directory Number
NodeB	NodeB, a logical node handling transmission/reception in multiple cells. Commonly, but not necessarily, corresponding to a base station
PCRF	Policy and Charging Rules Functions
PDC	Personal Digital Cellular
PDN	Packet Data Network
PDP	Packet Data Protocol
QoS	Quality-of-Service

RAT	Radio Access Technology
RF	Radio Frequency
RNC	Radio Network Controller
RSCP	Received Signal Code Power
RSRP	Reference Signal Received Power
S-GW	Serving Gateway
SAE	System Architecture Evolution
SAE GW	System Architecture Evolution Gateway
SDMA	Spatial Division Multiple Access
SIM	Subscriber Identity Module
TACS	Total Access Communication System
TD-SCDMA	Time-Division-Synchronous Code-Division Multiple Access
TE	Terminal Equipment
TV	Television
UE	User Equipment
UMTS	Universal Mobile Telecommunications System
UTRAN	UMTS Terrestrial Radio Access Network
VoIP	Voice over Internet Protocol
VLR	Visitor Location Register
WCDMA	Wideband Code-Division Multiple Access
WEB	World Wide Web

1 Introdução

1.1 Motivação

Atualmente, para as grandes operadoras de telecomunicações as redes sem fios suportam múltiplos sistemas de comunicações móveis de elevada importância económica. Com os avanços na tecnologia de comunicação, conectar-se com alguém, a qualquer hora e em qualquer lugar, tornou-se uma realidade da qual ninguém prescinde nos dias de hoje. Assim, fornecer um serviço de qualidade e resiliente tem-se revelado o principal desafio para qualquer operadora de comunicações móveis..

Nas últimas décadas, a Internet e as comunicações móveis, têm vindo a convergir para a definição de um novo paradigma, a Internet Móvel. A Internet Móvel combina a explosão do uso da Internet com a flexibilidade proporcionada pela mobilidade utilizados das telecomunicações móveis [1][2].

A possibilidade de aceder a informação e serviços a qualquer momento e em qualquer lugar está a moldar uma nova sociedade da informação, que vislumbra formas diferenciadas de acesso à Internet. Entretanto, a Internet não foi concebida para lidar com as características inerentes ao ambiente onde os vários utilizadores se movimentam estando conectados à rede sem fios com baixa largura de banda. A infraestrutura de comunicações móveis está a evoluir com o objetivo de fornecer uma melhor qualidade de serviço (QoS – *Quality of Service*) [3].

Neste projeto, proposto por uma operadora de telecomunicações Portuguesa, a NOS Comunicações, foi estudada a experiência de utilização da internet fixa dos clientes sobre a rede móvel (Internet fixa NOS e *Direct-to-Home* (DTH)). A NOS Comunicações oferece soluções fixas e móveis de última geração, de televisão, internet, voz e dados para

todos os segmentos de mercado – Residencial, Pessoal, Empresarial e *Wholesale*. Apresenta-se como um operador com ofertas convergentes multidispositivo que têm como objetivo uma elevada experiência de utilização em todo o País. Combina os interfaces de televisão com as velocidades de banda larga mais elevadas, cobertura superior e serviços móveis ilimitados para todas as redes.

Relativamente ao tema, o DTH é um serviço prestado pela NOS que inclui serviço TV (televisão) mais Internet residencial, em zonas do país que não está coberta por rede via cabo, como FFTH (*Fiber From The Home*) e HFC (*Hybrid Fiber Coax*).

A probabilidade de um cliente estar insatisfeito com o serviço contratado aumenta quando, sistematicamente, presencia velocidades de internet abaixo do contratado. Tal insatisfação, pode originar a perda de receitas por parte da operadora que presta o serviço e danos na imagem pública da empresa.

Verifica-se uma maior insatisfação dos clientes com serviço DTH em comparação ao serviço “WoW”, apesar de serem serviços idênticos. A diferença entre os dois serviços cinge-se na questão do serviço “WoW” ser apenas um pacote Internet, em vez de Televisão e Internet como no caso do DTH. Consequentemente, levantou-se a dúvida que deu origem a este estudo: “Qual a razão da elevada percentagem de subscritores que desistem do serviço DTH (*churn rate*)?”.

Como tal, o estudo incidiu sobre a rede móvel em que o serviço DTH oferece rede 3G e 4G, com velocidades máximas limitadas em 20Mbps e 40Mbps. Todos os operadores móveis usam indicadores de desempenho (KPIs – *Key Performance Indicators*) que são métricas estabelecidas para quantificar determinados aspetos do funcionamento da rede, neste caso, para avaliar o desempenho da qualidade de serviço (QoS) em relação à perspetiva do “consumidor” final [4]. Assim sendo, temas como a rede móvel e indicadores de desempenho tiveram um foco especial na concretização deste projeto.

A qualidade do sinal foi também um tema em foco neste estudo, uma vez que será, em teoria, um dos principais fatores para a qualidade do serviço. Em suma, analisaram-se vários parâmetros que tivessem influência, direta ou indireta, na qualidade do serviço oferecido pela operadora.

O desenvolvimento deste estudo destina-se ao departamento de Operações da NOS, podendo ser útil a muitas outras áreas (Engenharia para dimensionamento/planeamento, por exemplo).

Nesta tese é reunida informação sobre os temas que são partes nucleares no desenvolvimento deste estudo. Primeiro, foram introduzidos os sistemas da comunicação móvel para perceber como foram recolhidos os dados que são usados neste estudo e o seu

significado. Em seguida aborda-se o tema *Big Data*, onde se dá a conhecer o software em que os dados foram trabalhados. Mais à frente, é abordado o processamento de dados e, portanto, são apresentadas noções gerais de *Data Mining*.

1.2 Objetivos

Esta dissertação tem como objetivos propostos:

1. Verificar qualidade de rede e serviço;
2. Verificar experiência de utilização do cliente;

Neste sentido foram recolhidos dados de várias fontes. A partir da base de dados do *Access Control System* (ACS) consegue-se retirar os principais dados para esta análise. Os dados *Speed Tests* ou *Throughput* e os testes de qualidade de sinal, KPIs, dos clientes com serviço DTH da NOS permitem-nos fazer a primeira abordagem ao problema.

Após a extração destes dados é importante fazer o processamento dos mesmos. Os dados são compostos, entre muitos outros parâmetros, pelo número único que identifica o subscritor na rede móvel. Assim, consegue-se ter dados que não só dão a velocidade dada pelos testes de velocidade efetuados pelo router, bem como a potência do sinal aquando se deu o *speed test*.

Os testes de velocidade são efetuados em dois tipos de ocasião:

1. **Instalação** – São efetuados 3 *speed tests* num prazo de 24 horas. Dois dos três testes fora do horário mais congestionado (*off-peak*), que se define entre as 04:00h e as 08:00h e um teste em horário de maior congestionamento (*on-peak*), entre as 20:00h e as 24:00h.
2. **Suporte Cliente** – Usado pontualmente nas segundas linhas para despiste, testes executados pelos próprios clientes, etc.

Após extração de toda esta informação nas diversas fontes de dados, os dados são processados usando técnicas de *data mining*, desenvolvendo um estudo que identifique casos onde as degradações do serviço têm impacto e afetam os utilizadores com a finalidade de melhorar a experiência de utilização.

1.3 Principais Contribuições

O desenvolvimento do estudo descrito nesta dissertação inclui uma série de componentes de várias áreas das telecomunicações. Exigiu a interoperabilidade entre áreas como as Operações, departamento de Engenharia de Redes e Rádio, uma vez que os dados coletados estão sob supervisão de cada uma destas áreas e o correto entendimento dos dados foi discutido para que os resultados finais fossem coerentes e unânimes entre todos.

Com este estudo é possível ter conhecimento da experiência dos clientes num serviço que tem um impacto substancial na empresa, uma vez que, o número de clientes deste serviço é considerável. O estudo centra-se na problemática dos clientes apresentarem alguns sinais de insatisfação em relação ao serviço DTH, mais propriamente, na internet fixa na rede móvel.

Uma vez que este estudo ainda não tinha sido realizado na empresa, assume uma importância para o seu desenvolvimento. As diversas áreas abrangentes a este tema podem aceder a estes dados e aos respetivos resultados obtidos, permitindo assim que essas mesmas áreas consigam melhorar a qualidade do serviço prestado aos utilizadores. Visa, acima de tudo, encontrar pontos de partida que possam solucionar as velocidades experienciadas pelos utilizadores que estão abaixo do serviço contratado.

1.4 Organização da Dissertação

A presente dissertação está organizada em 5 capítulos. Após o capítulo introdutório, segue-se o capítulo 2, denominado de Estado de Arte, que para além de introduzir alguns conceitos relevantes para o estudo, refere como foram recolhidos os dados usados segundo algumas arquiteturas na rede móvel e o *software* usado para o processamento desses mesmos dados.

A descrição do estudo desenvolvido é feita no terceiro capítulo, começando por apresentar, conceptualmente, a recolha dos dados dos testes efetuados, introduzindo as características das diversas fontes de dados, bem como o dicionário correspondente aos campos dos dados retirados pelas fontes. Para além da recolha de dados, fala-se na análise dos dados, onde se apresenta as dimensões de análise. Ou seja, define-se um conjunto de dimensões que se achou uma mais valia para este estudo. De seguida, descreve-se, detalhadamente, o modo de processamento dos dados que anteriormente foram recolhidos (*data mining*).

No capítulo 4 são apresentados todos os resultados e testes efetuados sobre o estudo da qualidade da internet fixa da rede móvel com respetivas validações sobre os mesmos.

Por último, no capítulo 5 são referidas as conclusões do estudo e são apresentadas algumas sugestões futuras.

2 Estado de Arte

2.1 Introdução

Este capítulo destina-se à apresentação teórica dos temas abordados ao longo da dissertação. Incide sobre a topologia e principais características das diferentes arquiteturas de sistema móvel. Aborda ainda os temas relacionados com recolha e processamento dos dados, evidenciando uma série de conceitos considerados importantes para o projeto.

2.2 Comunicações Móveis

A comunicação móvel tornou-se uma comodidade diária. Nas últimas décadas, evoluiu de uma tecnologia cara para um nicho de pessoas para sistemas omnipresentes do dia de hoje e que são usados pela maioria da população mundial.

As tecnologias da comunicação móvel são divididas em quatro gerações de sistemas de comunicação móvel, cada uma associada a um conjunto específico de tecnologias e a um conjunto específico de casos de uso suportados, veja-se a Figura 2.2.1, com 1G a ser um sistema de rádio móvel analógico, 2G o primeiro sistema digital móvel, 3G o primeiro sistema móvel a suportar dados banda larga e 4G ou *Long-Term Evolution* (LTE) a fornecer um melhor suporte para banda larga móvel [5], [6].

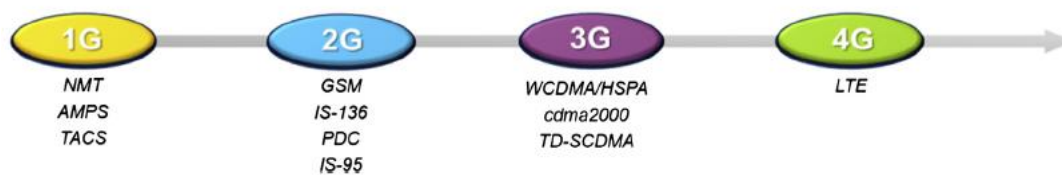


Figura 2.2.1 - Gerações dos sistemas de comunicação móvel[5].

Esta secção descreve, sucintamente, os sistemas de comunicação móvel 2G, 3G e 4G, caracterizando cada uma das arquiteturas fazendo o paralelismo entre as arquiteturas e os sistemas a que os dados estudados se referem, nomeadamente as arquiteturas 3G e 4G.

2.2.1 Arquitetura GSM (2G)

A arquitetura GSM (*Global System for Mobile Communication*) definiu a base principal e as diretrizes para todas as outras gerações de redes móveis vindouras. A rede rádio GSM consiste num conjunto de várias células de rádio controladas por uma BTS. Uma célula é uma representação geográfica da área de cobertura dentro da qual uma BTS pode enviar e receber dados. Cada estação base serve um número de *Mobile Stations* (MS) que representam os *users*, e um número de estações base são controladas pela *Base Station Controller* (BSC), que como o nome indica, controla funções como o controle de potência e entrega [7][4].

A Figura 2.2.2 mostra a arquitetura da rede GSM, onde se pode constatar que é dividida em quatro principais grupos funcionais, e são [9]:

- *Mobile Station* (MS): é também conhecida como *User Equipment* (UE), e esta entidade é composta pelo Módulo de Identidade do Subscritor (SIM) e pelo equipamento terminal;
- *Base Station Subsystem* (BSS): lida com funções de acesso rádio, como a gestão de recursos de rádio. Conecta os UEs com o núcleo da rede;
- *Core Network* (CN): inclui funções de transporte, gestão de mobilidade, base de dados dos subscritores com informações dos mesmos, funções de controle de serviço, pagamentos, etc.;
- *External Network* (EN): estas são as redes externas com que os UE podem comunicar e a que a rede móvel precisa de ser conectada. Pode ser, por exemplo, a rede telefónica pública ou qualquer outra rede GSM;

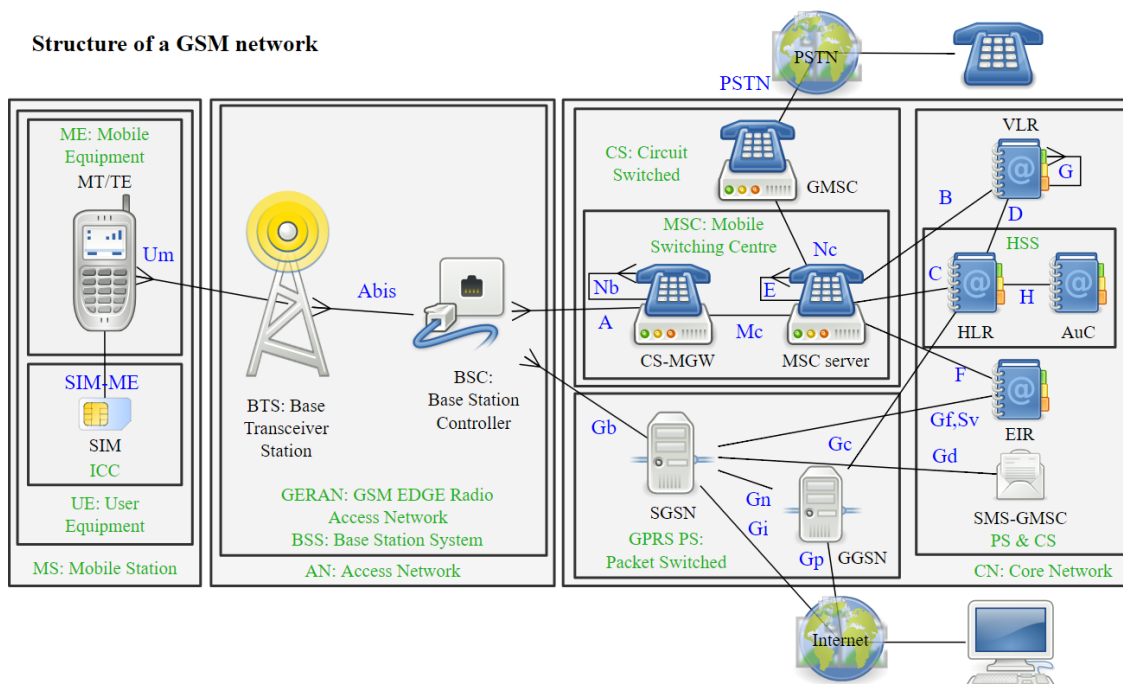


Figura 2.2.2 - Arquitetura de rede GSM [10].

Existem várias configurações BSC-BTS. Algumas configurações são mais adequadas para um elevado tráfego, e outras destinadas a atender áreas de tráfego moderado-baixo. A *Base Station Controller*, como o nome indica, controla funções como o controle de potência e entrega.

A *Base Station System* e a *Base Transceiver Station* juntas são conhecidas como BSS, que é visto pelo MSC, através de uma única interface, como a entidade responsável pela comunicação com a *Mobile Station* numa determinada área. A BSS, *Base Station System*, está associada a gestão de canal de rádio, funções de transmissão, controle de ligação de rádio e avaliação de qualidade e preparação para transferência, e garante ainda a cobertura de N células [8].

Uma rede de rádio móvel inclui uma pluralidade de BTS dispersas geograficamente, e cada uma fornece uma cobertura rádio de uma área geográfica ou célula. Cada BTS está, geralmente, no centro da respetiva célula. Assim, a área de cobertura geral é dividida em células e um dispositivo móvel, como um telemóvel, na área de coberta por uma rede de BTSs, pode em qualquer momento, trocar dados e sinalização por rádio com a BTS da célula que está atualmente localizado.

Dois principais tipos de dados podem ser “trocados” entre o dispositivo móvel e a BTS da célula em que o dispositivo está alocado:

- Dados de voz;

- Dados de Sinalização e controle, que incluem informações que permitem ao dispositivo móvel aceder à rede móvel ou manter uma chamada em progresso na mudança de uma célula para outra (“Transferência”).

Quando alguém pretende fazer uma chamada através de um telemóvel para outra pessoa, ou seja, para transmitir e receber voz, a respetiva BTS atua como uma estação retransmissora e, para este fim, tem pelo menos um transmissor e um recetor associado a, pelo menos, uma antena.

O nível de transmissão de dispositivos móveis é limitado, e entre muitas razões, se estiverem muito afastados do centro da célula, a receção da estação será fraca, pelo que o sinal está num nível muito fraco para se conseguir distinguir da interferência ou do ruído.

Em regiões rurais, onde as células, geralmente, cobrem uma grande área, e portanto, são divididas em sectores, muitos canais de tráfego estão disponíveis, e na prática, poucos são usados pois a frequência com que se realizam chamadas é muito baixa (por norma, as regiões rurais contam com uma população bastante mais reduzida se compararmos com cidades). [11].

O GSM usa técnicas de comutação de circuito, *circuit switched*, para suportar chamadas de voz. Devido às necessidades emergentes de maiores taxas de dados foi desenvolvido o *General Packet Radio Service* (GPRS). O GPRS é visto como um passo ao longo da segunda geração de comunicação móvel GSM para a terceira geração, *Universal Mobile Telecommunication System* (UMTS). Oferece maiores taxas de dados em comparação com as taxas de dados muito baixas que podem ser oferecidas pelo GSM. Permite assim, uma variedade de possibilidades e serviços oferecidos pelas operadoras móveis, por exemplo, navegação Web. Pela primeira vez, o sistema GSM consegue disponibilizar um sistema de uso de troca de pacotes na comunicação móvel.

Depois de GPRS, a evolução do sistema GSM continuou a suportar ainda maiores taxas de dados. Isto levou ao desenvolvimento de *Enhanced Data for GSM Evolution* (EDGE). A principal característica de EDGE era garantir maiores taxas de dados e esse aumento foi alcançado através da mudança do esquema de modulação. A Figura 2.2.3 mostra a evolução do sistema GSM com as respetivas taxas de dados[9].

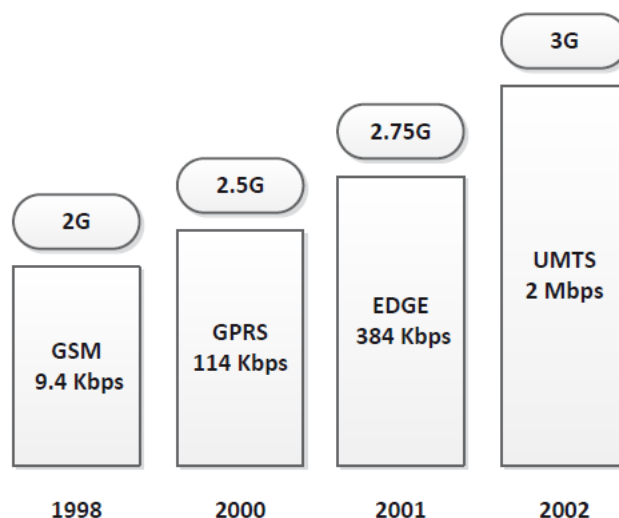


Figura 2.2.3 - Evolução do sistema GSM[9].

2.2.2 Arquitetura UMTS (3G)

A principal motivação por trás do UMTS foi definir um padrão de comunicação móvel universal que vise taxas de transmissão de dados mais altas, com a capacidade de adaptar, dinamicamente, as taxas de dados do cliente. Além disso, existem outros objetivos, como o suporte à diferenciação da Qualidade de Serviço (QoS) entre os diferentes serviços oferecidos pelo UMTS.

A arquitetura UMTS, ilustrada na Figura 2.2.4, está estruturada de forma semelhante ao GSM com algumas modificações. A rede de acesso rádio no sistema UMTS é chamada UMTS *Terrestrial Radio Access Network* (UTRAN) e consiste no *Radio Network Controller* (RNC) e vários NodeBs, que representam as estações base UMTS. A rede UMTS suporta conexões com comutação de circuitos (*circuit switched*) e comutação de pacotes (*packet switched*). As conexões com comutação de circuito são usadas para transportar serviços de voz, enquanto que as conexões de comutação de pacotes são usadas para outros serviços de dados, como navegação Web através de HTTP (*Hypertext Transfer Protocol*), e *downloads/uploads* de ficheiros por FTP (*File Transfer Protocol*) [9].

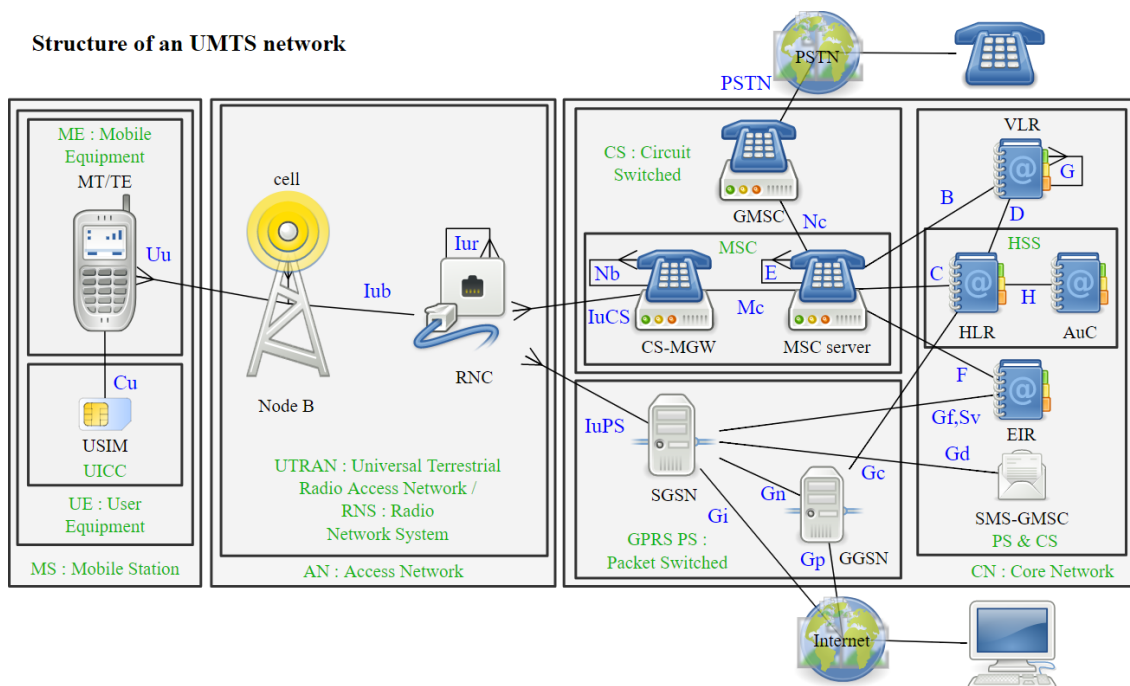


Figura 2.2.4 - Arquitetura da rede UMTS [12].

Semelhante ao GSM, existiram dois melhoramentos no sistema UMTS, aumentando as taxas de dados, a capacidade do sistema e a redução da latência do sistema. Estes são *High Speed Downlink Packet Access (HSDPA)* e *High Speed Uplink Packet Access (HSUPA)*.

O uso de ambos os melhoramentos (HSDPA e HSUPA) é muitas vezes referido como HSPA. Os operadores de rede implementaram o HSPA em coexistência com as redes UMTS R99 (primeira versão dos padrões de UMTS construído no final de 1990, e é por isso que, por vezes, UMTS é referido como *release 99* ou R99). O desempenho instantâneo do rádio pode variar ao longo do tempo, por vezes, alcançando um *throughput* muito alto nas células. No entanto, as operadoras móveis dimensionam o *backhaul* considerando o desempenho médio, de modo a conseguirem reduzir o custo [9], [13], o que causará congestionamentos de curto prazo na rede. De forma a mitigar a influência que causa na rede, esquemas de controle de congestão, bem como técnicas de separação de tráfego, são usados para superar problemas de congestão e fornecer uma diferença de QoS entre o tráfego HSPA e R99 [9], [13]–[16].

2.2.3 Arquitetura LTE (4G)

LTE, *Long Term Evolution*, é o lançamento mais recente da 3GPP (*3rd Generation Partnership Project*). O objetivo no 3GPP *Release 8* é melhorar a escalabilidade da rede

para o aumento do tráfego e minimizar a latência por redução do número de elementos da rede. Todos os protocolos rádio, gestão de mobilidade, compressão de cabeçalhos e todas as retransmissões de pacotes estão localizados nas estações base chamadas eNodeB.

eNodeB inclui todos os algoritmos localizados no *Radio Network Controller* (RNC) na arquitetura 3GPP *Release 6* (HSPA). O núcleo da rede também é simplificado por separação do *control plane* e *user plane* [17].

A Entidade de Gestão de Mobilidade (MME) é apenas o elemento do plano de controlo, enquanto o plano do utilizador ignora o MME e segue diretamente para o *System Architecture Evolution* (SAE) Gateway (GW). A evolução da arquitetura está ilustrada na Figura 2.2.5 [17].

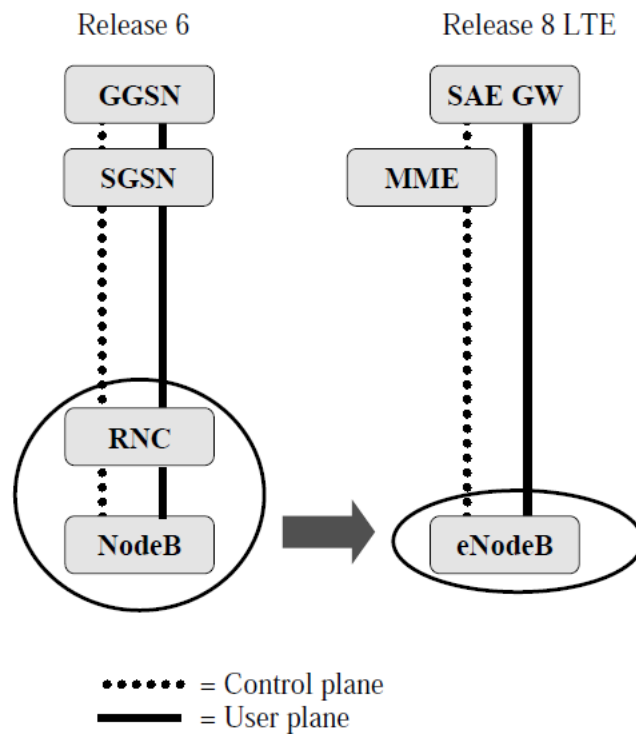


Figura 2.2.5 - Arquitetura da rede LTE [17].

O sistema LTE foi projetado para os serviços comutados de pacotes que fornecem conectividade por IP entre PDN (*Packet Data Network*) e UE sem interrupções de serviço. Pode ser dividido em duas partes principais: *Evolved Universal Terrestrial Radio Access Network* (E-UTRAN) e *System Architecture Evolution* (SAE). E-UTRAN é a evolução da rede de acesso rádio UMTS. SAE suporta a evolução do núcleo da rede de pacotes, também conhecida como *Evolved Packet Core* (EPC). A combinação de E-

UTRAN e SAE compõe o *Evolved Packet System* (EPS). A arquitetura geral da rede LTE é mostrada na Figura 2.2.6 [9].

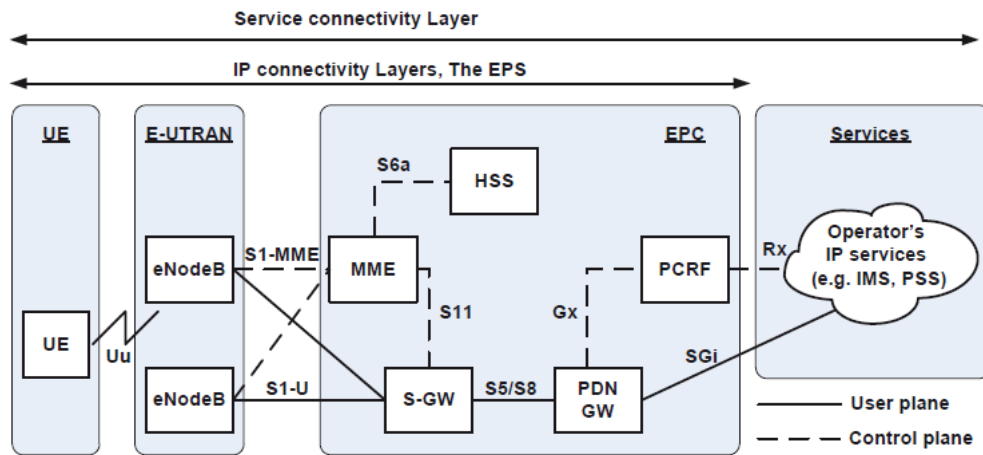


Figura 2.2.6 - Arquitetura geral da rede LTE EPS[9].

De acordo com [18], a conexão virtual em EPS é definida como um fluxo de pacotes IP entre PDN (*Packet Data Network*)-GW e UE com características pré-definidas de qualidade de Serviço (QoS). Tanto EPC como E-UTRAN são responsáveis pela configuração e libertação da tal conexão virtual dependendo dos requisitos de QoS da aplicação. É possível estabelecer diversas transmissões de sinal em LTE, a clientes com vários serviços, por exemplo, um cliente pode ter uma chamada de voz usando *Voice over Internet Protocol* (VoIP) e ao mesmo tempo estar a baixar um ficheiro usando *File Transfer Protocol* (FTP), ou a navegar na internet usando *Hypertext Transfer Protocol* (HTTP) [9].

2.2.3.1 User Equipment (UE)

Como sugere o nome, UE é o dispositivo que o consumidor LTE usa para se conectar à rede LTE e estabelecer conexão. Pode ser um telemóvel, *tablet* ou um cartão com dados usados por um computador/portátil. Semelhante aos outros sistemas 3GPP, o UE consiste em duas principais entidades: um cartão SIM, e o equipamento terminal (TE). O cartão SIM carrega a informação necessária fornecida pelo operador para identificação de *user* e procedimentos de autenticação. Já o equipamento terminal fornece os clientes com o hardware necessário (por exemplo, processamento, armazenamento, sistema operativo) para correr as aplicações e usar os serviços do sistema LTE [9].

2.2.3.2 E-UTRAN

E-UTRAN no LTE consiste na conectividade de eNodeBs através da interface X2 e na conexão com a rede nuclear através da interface S1. Assim, é eliminada uma das maiores desvantagens do antigo sistema 3GPP (UMTS/HSPA): a necessidade de conectar e controlar NodeBs através de RNC (*Radio Network Controller*) que tornam o sistema vulnerável a falhas RNC [9]. A arquitetura LTE E-UTRAN pode ser vista na Figura 2.2.7.

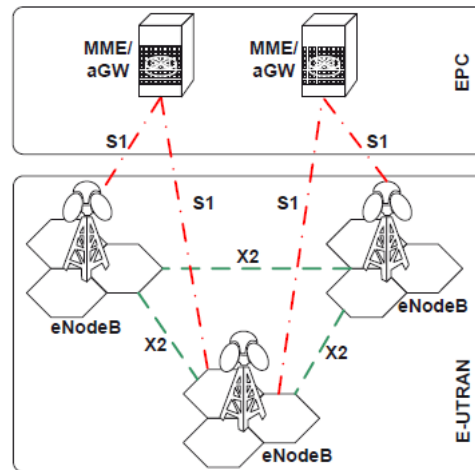


Figura 2.2.7 - Arquitetura LTE E-UTRAN [9].

A versão melhorada do NodeB, eNodeB, funciona como ponte entre o cliente e o EPC. Fornece ainda, os protocolos necessários de rádio ao UE, de modo a ser capaz de enviar e receber dados e canaliza (*tunneling*) os dados do cliente com segurança sobre o transporte LTE para o PDN-GW, e vice-versa.

O eNodeB é também responsável pelo agendamento de recursos, que é uma das mais importantes funções rádio, enquanto garante diferentes QoS para os consumidores finais. Adicionalmente, o eNodeB tem algumas funcionalidades de gestão de mobilidade, por exemplo, medições de ligações rádio e transferência de sinal para outros eNodeB [9].

2.2.3.3 Evolved Packet Core (EPC)

Como é mostrado na Figura 2.2.6, EPC consiste em três principais entidades: *Mobility Management Entity* (MME), *Serving Gateway* (S-GW) e *Packet Data Network Gateway* (PDN-GW). Além disso, existem outras entidades como *Home Subscriber Server* (HSS) e *Policy and Charging Rules Functions* (PCRF). O principal objetivo de EPC é fornecer as funcionalidades necessárias para apoiar os clientes [18]. As funcionalidades

de cada uma das entidades está ilustrada na Figura 2.2.8 conforme a descrição das mesmas descritas neste capítulo.

A entidade MME é equivalente ao *Home Location Register* (HLR) e ao *Visited Location Register* (VLR) na rede UMTS. MME lida com a sinalização e controle, e gestão de mobilidade. Isto facilita a otimização das redes implementadas e permite total flexibilidade na expansão da capacidade. Administra também o acesso do UE à rede através da interação com o HSS para autenticação de clientes, e providencia a função do plano de controle para permitir a mobilidade contínua entre redes móveis LTE e 2G/3G e ainda suporta a interceção legal de sinalização [19], [20].

O S-GW atua como o ponto terminal entre a rede de acesso rádio (E-UTRAN) e a rede central/nuclear. Envia pacotes de dados para e de eNodeB e PDN-GW e é responsável pela contabilidade e controle de dados do cliente. Serve também como âncora de mobilidade local para transferências entre eNodeBs ou passagem entre redes 3GPP e informa o tráfego do user [19]–[21]. O *Packet Data Network Gateway* (PDN-GW) atua como o ponto de conectividade do cliente para o tráfego do mesmo, é responsável por atribuir os endereços IP dos users, bem como classificar o tráfego em diferentes classes QoS. Adicionalmente, o PDN-GW atua como ponto de âncora de mobilidade para interagir com transferências entre sistemas de acesso 3GPP e tecnologias não-3GPP, como LAN (*Local Area Network*) sem fios [9].

PCRF dá permissão ou rejeita pedidos de multimédia. Cria e atualiza o contexto do protocolo de dados por pacotes (PDP) e controla a alocação de recursos. Fornece também as regras de cobrança com base no fluxo de dados do serviço para o PDN-GW, e coordena a qualidade de serviço. Já o HSS, inclui funcionalidades semelhantes ao HLR, ou seja, informações específicas do cliente. Base de dados permanente e central; Armazena a mobilidade e os dados de serviço para cada um dos assinantes e contém a funcionalidade do centro de autenticação (AuC) [19], [20].

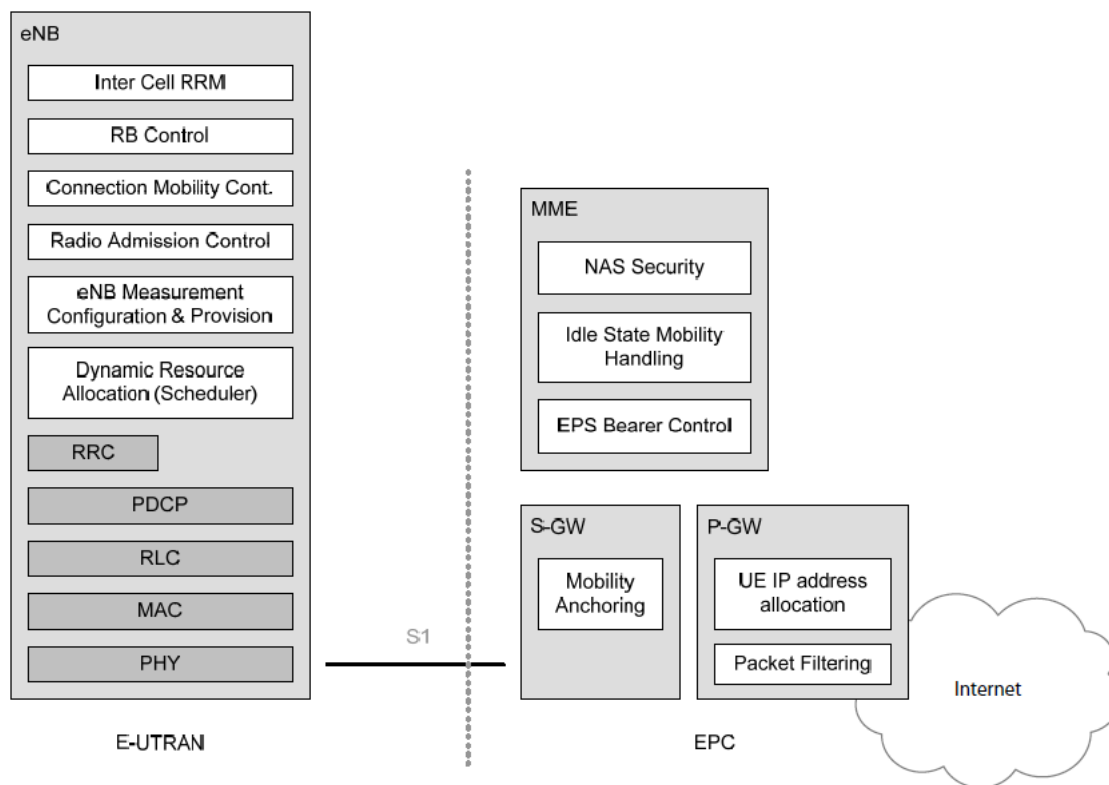


Figura 2.2.8 - Funcionalidades de E-UTRAN e EPC [18].

2.3 Big Data

Nos últimos anos, tem-se testemunhado um aumento exponencial na capacidade de gerar dados de várias fontes, em diferentes formatos. Esta inundação de dados ultrapassou a capacidade humana de processar, analisar, armazenar e entender estes conjuntos de dados [22]. O principal desafio para as aplicações de *Big Data* é explorar os grandes volumes de dados e extrair informações úteis ou conhecimento para ações futuras [23]. Em muitas situações, o processo de extração de conhecimento tem que ser muito eficiente e próximo ao tempo real, pois armazenar todos os dados registados é quase inviável [24].

Os dispositivos móveis são uma nova grande fonte de dados, e grandes empresas como a Google, Apple, Facebook, Yahoo, Twitter estão a começar a olhar cuidadosamente para esses dados para encontrar padrões úteis para melhorar a experiência do utilizador. São necessários novos algoritmos e novas ferramentas para lidar com esta quantidade de dados [22]. Doug Laney [25] foi o primeiro a mencionar os 3 V's do *Big Data* [26]:

- Volume: refere-se ao tamanho do *dataset* que se tem de processar;

- Variedade: os dados são produzidos por diversas fontes e existem muitos tipos diferentes de dados, como texto, dados de sensores, áudio, vídeo, gráficos e muito mais;
- Velocidade: os dados estão a chegar continuamente como fluxos de dados, e é importante obter informações úteis em tempo real;

Atualmente, existem mais dois V's [22]:

- Variabilidade: há mudanças na estrutura dos dados e como os utilizadores querem interpretar esses dados;
- Valor: valor de negócio que dá às organizações uma vantagem competitiva, devido à capacidade de tomar decisões baseadas em responder a perguntas que antes eram consideradas fora de alcance;

2.4 Data Mining

O *Data Mining* atrai muita atenção no setor da tecnologia da informação e na sociedade como um só todo, devido à ampla disponibilidade de enormes quantidades de dados e à necessidade iminente de transformar esses dados em informações e conhecimentos úteis. As informações e os conhecimentos adquiridos podem ser utilizados para as mais variadas aplicações. Contextualizando com o tema da dissertação, podem ser aplicados desde a análise de mercado e à retenção de clientes, até à produção e exploração científica.

Se os dados são caracterizados como factos registados, então a informação é um conjunto de padrões, ou expectativas, que estão subjacentes aos dados. Considera-se assim que *Data Mining* é nada mais que a extração de informações implícitas dos dados, previamente desconhecidas, e potencialmente úteis [27]. *Data Mining* é também conhecido como *Knowledge Discovery from Data* (KDD).

A descoberta de conhecimento, como um processo, é representada na Figura 2.4.1 e consiste numa sequência iterativa das seguintes etapas:

1. **Data Cleaning:** remove o ruído e dados inconsistentes;
2. **Data Integration:** onde várias fontes de dados podem ser combinadas;
3. **Data Selection:** onde dados relevantes para a tarefa de análise são recolhidos da base de dados;
4. **Data Transformation:** onde os dados são transformados ou consolidados em formas apropriadas para *mining*, por exemplo, por operações de agregação;

5. **Data Mining:** processo essencial onde métodos inteligentes são aplicados de maneira a extrair padrões de dados;
6. **Pattern Evaluation:** identifica os padrões verdadeiramente interessantes que representam o conhecimento com base em algumas medidas de interesse;
7. **Knowledge Presentation:** onde a visualização e as técnicas de representação do conhecimento são usadas para apresentar o conhecimento extraído dos utilizadores

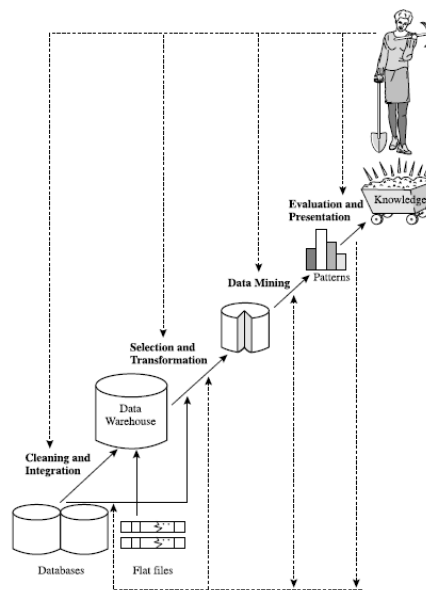


Figura 2.4.1 - *Data Mining* como etapas no processo de descoberta de conhecimento [28].

As etapas 1 a 4 são diferentes formas de pré-processamento de dados, onde os dados são preparados para *mining*. A etapa de *data mining* pode interagir com o utilizador, o que é o caso do nosso estudo, em que os padrões interessantes são apresentados ao utilizador e podem ser armazenados como novos conhecimentos. Note-se que, de acordo com esta visão, o *data mining* é apenas uma etapa em todo o processo, embora seja essencial pois está na base da descoberta de padrões ocultos à primeira vista [28], [29].

2.4.1 Classificação de Sistemas de Data Mining

Data Mining é um campo interdisciplinar com influência de um conjunto de disciplinas, como sistemas de bases de dados, estatística, aprendizagem automática, visualização e ciências de informação (Figura 2.4.2).

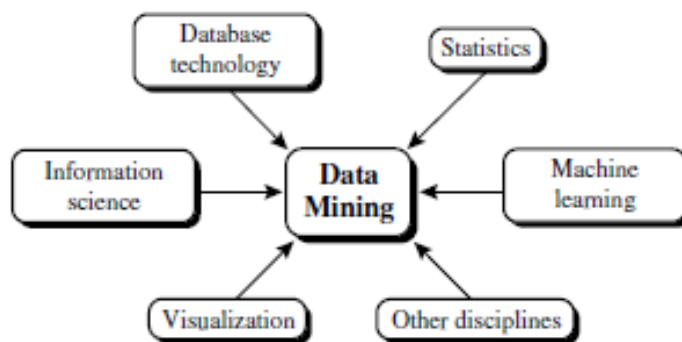


Figura 2.4.2 - *Data Mining* como influência de múltiplas disciplinas [28].

2.4.2 Pré-Processamento dos Dados

Dados incompletos e inconsistentes são ocorrências comuns em grandes bases de dados. Os atributos com interesse nem sempre podem estar disponíveis, outros dados podem não ser, simplesmente, incluídos porque não são considerados importantes à primeira vista. Os dados inconsistentes podem e devem ser excluídos dos outros dados gravados.

Em resumo, os dados do mundo real tendem a ser incompletos e inconsistentes. As técnicas de pré-processamento podem melhorar a qualidade dos dados, ajudando a melhorar a precisão e a eficiência do subseqüente processo de *mining* [28]. A Figura 2.4.3 sumariza as etapas de pré-processamento de dados que foram realizadas neste trabalho e que são explicadas mais ao detalhe na secção 3.2.2.

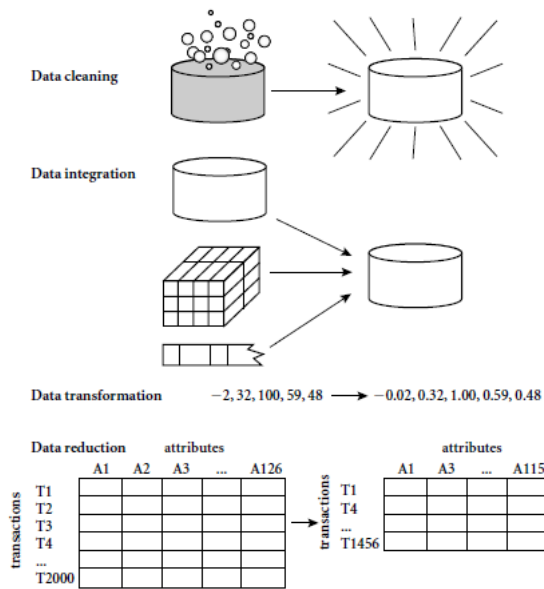


Figura 2.4.3 - Formas de pré-processamento de dados [28].

O pré-processamento de dados é um passo bastante importante no processo de descoberta de conhecimento, porque as decisões de qualidade devem basear-se em dados de qualidade. Detetar anomalias de dados, corrigi-las antecipadamente e reduzir os dados a serem analisados contribuem com grandes ganhos no processo de tomada de decisão [28].

3 Descrição do Projeto

O capítulo 3 descreve todo o estudo realizado sobre a qualidade da Internet fixa na rede móvel do serviço DTH da NOS Telecomunicações. O objetivo é, como referido anteriormente, verificar a qualidade de rede e do serviço prestado pela empresa e a experiência de utilização do cliente.

Este projeto baseia-se nos dados recolhidos no sistema de monitorização que permitem obter um vasto leque de informação a respeito dos clientes que têm este serviço. Sabendo que a experiência do utilizador acaba por ser o fator mais importante para qualquer empresa, especialmente, quando falamos de uma empresa inserida num mercado tão competitivo como as telecomunicações. No caso deste estudo, a experiência de utilização do consumidor está relacionada com a velocidade da Internet (“*Throughput*”).

Portanto, ao longo deste capítulo são apresentados os *datasets* recolhidos, alguns dos pré-processamentos efetuados para trabalhar os referidos dados e quais os algoritmos usados para, de alguma maneira, “aprender” através da base de dados.

3.1 Recolha de dados

O *data mining* e *data science* são dos principais temas para qualquer estudo do mundo empresarial. Para este foi identificado um problema em concreto que tem uma

parte importante a nível económico, uma vez que enfrentamos um problema que afeta diretamente a experiência do cliente.

Antes de passar para o processamento de dados, é importante referir de que forma e que tipos de dados foram recolhidos para contruir os *datasets* que foram usados para estudar o problema em questão. Tal como se encontra ilustrado na Figura 3.1.1, é monitorizado todo o tráfego de um dado terminal (ACS - *Access Control System*). Tendo em conta que a NOS é uma companhia de telecomunicações em Portugal e qualquer cliente com serviço DTH com internet fixa terá a produzir tráfego, é possível ter uma quantidade astronómica de dados.

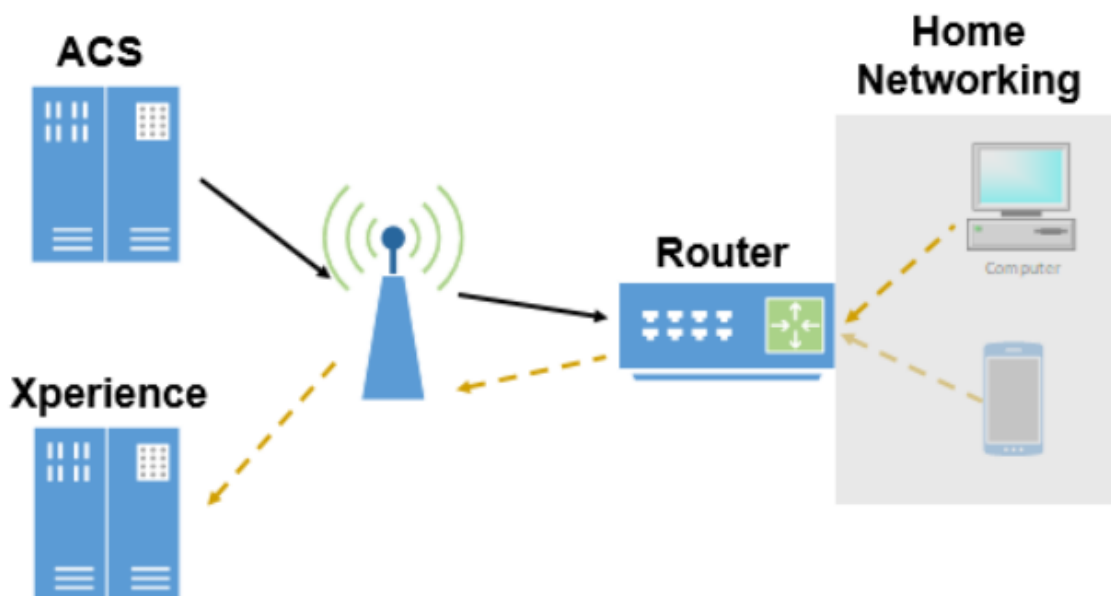


Figura 3.1.1- Diagrama de recolha de dados.

Como mostrado pela Tabela 3.1.3, tem-se diversas fontes de recolha de dados, e cada conjunto de dados recolhidos incide tanto em dados da célula como dados relacionados com a experiência do utilizador [8] - Número que identifica de forma exclusiva uma assinatura na rede móvel)

Os conjuntos de dados recolhidos é apresentado na Tabela 3.1.1 e Tabela 3.1.2.

Tabela 3.1.1 - Característica do conjunto de dados KPI.

Característica	Descrição
CPEID	Serial <i>number</i> que executou a leitura dos KPI radio
READ_STAMP	<i>Unixtimestamp</i> da leitura dos KPI radio
CPE_VERSION	Versão do equipamento
IMEI	Imei do equipamento
IMSI	IMSI do MSISDN do cartão de dados
MSISDN	MSISDN do cartão de dados
RAT	RADIO ACCESS TYPE (4G/3G/2G)
CELL_ID	<i>Cell id</i> onde o equipamento estava ligado ("Antena")
RSSI	Métrica 3G - <i>Received Signal Strength Indication</i> – Nível total do sinal recebido, ou seja, é a soma do nível de sinal rádio RSCP ao qual se junta o ruído EcNo. A medida RSSI no <i>Uplink</i> é utilizada para detectar interferências externas ou provocadas pelo próprio sistema de antenas.
ECIO	Métrica 3G - é a razão entre o nível de sinal e o ruído
RSCP	Métrica 3G - <i>Received Signal Code Power</i> - Nível de sinal recebido por um equipamento móvel, reporta um determinado valor RSCP em dBm, Valores entre -50 e -80 dBm considerados bons, entre -80 e -100 razoáveis e menos de -100 dBm valores fracos de cobertura
RSRP	Métrica 4G - <i>Reference Signal Received Power</i> – Nível de sinal recebido por um equipamento, reporta um determinado valor RSRP em dBm, Valores entre -50 e -80 dBm considerados bons, entre -80 e -100 razoáveis e menos de -100 dBm valores fracos de cobertura.
RSRQ	Métrica 4G - <i>Reference Signal Received Quality</i> – Qualidade do sinal, relação entre o nível de sinal RSRP e RSSI.
LTE_BAND	Banda LTE: 3 - 1800MHz 7 - 2600MHz 20 - 800MHz
TOTAL_BYTES_SENT	Número de Bytes enviados pelo router (<i>reset no factory</i>)
TOTAL_BYTES_RCVD	Número de bytes recebidos pelo router (<i>reset no factory</i>)
UPTIME	Tempo desde último <i>boot</i>
EXTRACTED_STAMP	NA (NULL)
ANTENNA_STATUS	Se a antena usada é a que vem no equipamento ou externa (antena externa ligada ao equipamento)

Tabela 3.1.2 - Características do conjunto de dados dos testes de velocidade

Característica	Descrição
Master_ID	
Internal_ID	
Status_CODE	0 - NOK 1-OK
TEST_TYPE	<i>Scheduled</i> - agendado <i>On-demand</i> - executar na altura
SCHEDULE_STAMP	<i>Unixtimestamp</i> de quando foi pedido ou agendado
DEADLINE_STAMP	<i>Unixtimestamp</i> de fim
CPEID	Serial <i>Number</i> do <i>device</i> que executou o teste
IMSI	IMSI do MSISDN do cartão de dados
MSISDN	MSISDN do cartão de dados
IMEI	Imei do equipamento
RAT	RADIO ACCESS TYPE (4G/3G/2G)
CELL_ID	<i>Cell id</i> onde o equipamento estava ligado ("Antena")
DST_TRG_SIZE_B	Número de bytes alvo DS
DST_TRF_SIZE_B	Número de bytes efetivo DS
DST_START_STAMP	Início de download
DST_END_STAMP	Fim de download
UST_TRG_SIZE_B	Número de bytes alvo US
UST_TRF_SIZE_B	Número de bytes efetivo US
UST_START_STAMP	Início de download
UST_END_STAMP	Fim de download
INSERT_STAMP	<i>Export time</i>
EXTRACTED_STAMP	NA (NULL)

Na tabela seguinte encontra-se as fontes e os respetivos tipos de dados.

Tabela 3.1.3 - Recolha de dados: fontes e respetivos dados

Fonte	Característica	Observação
ACS (célula/msisdn)	KPIs de sinal	Potência de sinal
	<i>Throughput</i>	

	Banda	4G – 800; 1800; 2600 3G – 2100;
	RAT	3G/4G
Spiro (célula)	Utilização	Verde – Boa Amarela – Normal Vermelha – Congestionada
	Capacidade	Excelente; Muito bom; Bom; Razoável; Mau; Muito Mau
CPM (msisdn)	Produto	WOW/DTH
	Velocidade	20/40 Mbps
	PTs	Queixas de clientes
GIS (célula)	Localização das células	
	Freguesias/Conselhos/Região	
	Fornecedor	

O ACS é uma das várias fontes de dados do sistema de monitorização onde são recolhidas medições no equipamento do utilizador. Aqui é medida a velocidade do serviço e a potência de sinal no router, no momento em que a instalação do serviço DTH foi efetuada na casa do cliente, expurgando assim efeitos do *home networking*. São, então, efetuados três testes de velocidade no momento da instalação, num prazo de 24 horas, onde dois testes são efetuados *off-peak* [04:00h-08:00h] e um *on-peak* [20:00h-24:00h]. Todos os restantes testes estão no âmbito de suporte ao cliente e são usados pontualmente nas segundas linhas para despiste, campanhas, etc.

Sabendo que, na rede móvel os testes realizados são em 3G ou 4G, o ACS agrega essa informação e a respetiva banda de cada RAT, à velocidade da internet móvel.

O Spiro é uma plataforma de *performance* e gestão da rede NOS. Isto é, uma base de dados que associa toda a informação da rede em si, em que existem métricas associadas à célula, estatísticas de performance, informações de QoS, taxas de utilização, contadores, etc. Como mostra na Tabela 3.1.3, dois tipos de dados importantes foram usados: a Utilização/Ocupação e a Capacidade.

Em cada um dos tipos de dados foram subdivididos conforme o estado da célula. Na parte do processamento de dados é justificado essa subdivisão para ambos os casos. Por enquanto, pode-se afirmar, em termos de estudo do *dataset*, que a taxa de utilização é dividida por cores. No caso de um *speed test* ocorrer numa situação em que a célula está congestionada, um campo no *dataset* marca o teste como uma taxa de ocupação vermelha.

Numa situação em que a célula estiver livre de congestão, tem uma taxa de ocupação verde. Caso não seja ambas as situações anteriores, ou seja, não estejam nem congestionadas nem completamente livre de congestão, a taxa de ocupação é assinalada como amarela.

A Capacidade é outra característica dos dados retirados no Spiro que é importante para este estudo. Tal como a Taxa de Ocupação, a Capacidade é subdivida em escalões relativas ao estado da célula e também são explicados na secção Processamento de dados.

Na mesma linha que o Spiro, surge o GIS (Geographic Information System). Como o nome indica, é um sistema de informações geográficas das células que permite ter conhecimento da localização das mesmas, bem como o fornecedor correspondente.

Já o CPM, *Customer Premise Management*, faz a gestão do cliente. Assim, pelo identificador único do cliente podemos ter acesso as informações diretas do cliente relacionadas com este estudo. Ou seja, podemos ter informação do tipo de serviço contratado (produto), a velocidade que o cliente paga para ter em sua casa e eventuais participações técnicas (queixas de clientes ao serviço de apoio).

3.2 Análise de dados

3.2.1 Dimensões de Análise

Antes de qualquer processamento dos dados recolhidos e para um melhor seguimento de todo o projeto, foram previamente definidas, as análises de estudo como ilustrado na Tabela 3.2.1.

Tabela 3.2.1 - Análises de estudo.

Análise	Fase	Observação
Horário	1	<i>On-peak</i> [20:00h-24:00h] <i>Off-peak</i> [04:00h-08:00h]
Terminal	1	Router E5172/B310
Cliente	3	Identificação do cliente
PT's	3	Queixa dos clientes
Célula	1	

Capacidade da Célula	1	
Utilização da Célula	1	
Evolução diária da Célula	1	Evolução da célula (<i>throughput</i> de cliente)
RAT	1	3G/4G
Largura Banda Target	4	Com larguras de banda diferentes tem capacidade e alcances diferentes
Número de Dispositivos na célula	2	
Fornecedor	2	Huawei/Ericsson
KPI sinal	1	3G – RSCP 4G – RSRP
Produto	1	DTH

Tendo conhecimento de quais as características pretendidas para descrever os dados recolhidos pelas diversas fontes já descritas anteriormente, e sabendo de forma objetiva os principais temas de estudo, é a altura para o processamento dos dados – *data mining*.

3.2.2 Processamento de dados

O processamento dos dados recolhidos é feito recorrendo à linguagem de programação R, no ambiente RStudio. A linguagem R surge como uma implementação da linguagem S e é um sistema para computação estatística e gráfica. É uma linguagem de programação que fornece gráficos de alto nível e muito útil para manipulação e análises de grandes volumes de dados. Para além disso, é possível aplicar algoritmos de *machine learning* [30],[31],[32].

No que diz respeito ao estudo, sempre que há novos clientes são recolhidas 3 medições de velocidade do ACS até ao router do cliente durante o próprio dia e dia seguinte à instalação. Uma medida no horário “on peak” e duas em “off peak”. Com estas medidas é possível perceber a qualidade da internet dada ao cliente, assim como os sinais ao longo do tempo nas células.

Deve ainda ser analisado a amostra a trabalhar e se os dados recolhidos são suficientes para representar o universo. Ou seja, no caso de uma célula com apenas 1 cliente novo por dia não deve ser representativo para este estudo.

Como visto anteriormente na secção da Recolha de dados, através do ACS consegue-se extrair dados relacionados com os testes de velocidade e testes de potência de sinal

no momento em que foram feitos os *speedtests*. Sendo que são dados correlacionados, mas que estão em *datasets* distintos, é importante correlacioná-los de algum modo. Uma vez que o ACS contém informação tanto da célula como do cliente, foi possível relacionar ambos *datasets* através do identificador único.

Com o corelacionamento dos conjuntos de dados dos testes de velocidade e da potência de sinal, reparou-se que alguns parâmetros ou não estavam preenchidos (supostamente, deviam estar) ou estavam a *null*. Para o caso do estudo do *throughput*, por exemplo, ter o campo do número de bytes efetivo de download e *upload* como *null*, não faz qualquer sentido e pode levar a conclusões a este estudo que não correspondem à verdade. Portanto, acrescentou-se um campo ao *dataset* definido por “TEST_RESULT”, em que definia um “TEST_OK” quando os parâmetros do “CELL_ID”, “DST_TRF_SIZE_B” e “UST_TRF_SIZE_B” estivessem não *null*.

Caso contrario, considera-se que o teste nem sequer tinha ocorrido se não se associasse a uma célula ou é um teste “não ok” no caso de se associar a uma célula, mas tivesse o campo do número de bytes efetivo de download e *upload* como *null*. Apesar de se ter caracterizado estes testes, os mesmos não entraram no estudo.

Relativamente ao *dataset* da potência de sinal, a Tabela 3.1.1 mostra duas métricas importantes de 3G e 4G, RSCP (*Received Signal Code Power*) e RSRP (*Reference Signal Received Power*) respetivamente, para caraterizar o nível de sinal recebido por um equipamento móvel. É dito ainda que, para ambos, reporta um determinado valor RSCP ou RSRP em dBm, Valores entre -50 e -80 dBm são considerados bons, entre -80 e -100 razoáveis e menos de -100 dBm consideram-se valores fracos de cobertura.

Assim, foi criado um campo para a métrica 3G (RSCP_KPI_3G) e outro para a métrica 4G (RSRP_KPI_4G) que caracteriza o nível de sinal recebido por um equipamento:

- OK – [-50,-80] dBm;
- REASONABLE –]-80, -100] dBm;
- NOK –]-100, -∞[dBm;

Na introdução do projeto foi admitido dois tipos de serviços, WOW e DTH, e o principal objetivo deste estudo seria perceber o elevado número problemas relatados pelos clientes do serviço DTH em relação ao serviço WOW, uma vez que ambos os serviços são idênticos, diferindo apenas na adição de internet fixa pela rede móvel ao serviço DTH. Como tal foi excluído todos os testes e análises que feitas com o serviço WOW.

Além do serviço, pelos dados fornecidos pelo CPM podemos ter acesso ao terminal, ou router, usados quando foram realizados os testes, e são: E5172 e B310. A velocidade

contratada pelo cliente é outra informação útil que se pode retirar, com vista a comparação que o cliente realmente tem em sua casa com a velocidade que paga.

Como representado na Tabela 3.2.1, o projeto foi dividido em fases. Como tal, inicialmente, foram recolhidos os dados do ACS, os testes de velocidade e de potência de sinal. Por si só, estes testes já continham bastante informação relativa às pressupostas análises pré-dimensionadas, como o “IMSI”, “MSISDN”, “RAT”, “CELL_IDM”, “DST_END_STAMP”, “DST_START_STAMP” e “DST_TRF_SIZE_B”.

O cálculo do *throughput* não era diretamente dado nos testes de velocidade. Estes continham vários campos que permitiu o cálculo do mesmo.

$$throughput = \frac{DST_TRF_SIZE_B}{DTS_END_STAMP - DST_START_STAMP} * \frac{8}{1000} * \frac{1}{1000} [Mbps] \quad (3.1)$$

Por (3.1) percebe-se que o cálculo da velocidade é dado pela divisão do número de bytes efetivos pelo tempo que se realizou o *download*. Multiplicando por uma constante, o *throughput* é dado em Mbps.

O cruzamento com SPIRO enriqueceu o *dataset* e, conseqüentemente, a análise com a congestão das células. Em conferência com o departamento de rádio e de gestão da operação e qualidade, definiram-se valores que pudessem caracterizar a capacidade e taxa de utilização da célula, como mostra a Figura 3.2.1 e a Tabela 3.2.2, respetivamente.

Measurement Threshold (full Compensation)	Serviço	Deep Indoor Muito Bom	Indoor Daylight Bom	Indoor Window Razoável	Incar Mau	Outdoor Muito mau
2G (RxLev)	Voz	-71	-77	-86	-87	-95
3G (RSCP)	HSDPA 4Mbps	-65	-70	-79	-82	-90
	HSDPA 1Mbps	-77	-82	-91	-94	-103
	Voz	-79	-84	-94	-93	-104
	PS64/384	-79	-84	-93	-97	-105
	HSDPA 400kbps	-80	-85	-94	-97	-105
	PS64/128	-82	-87	-97	-100	-108
4G (RSRP)	PS64/64	-84	-88	-98	-101	-109
	LTE800MHz UL 512kbps	-96	-101	-111	-112	-119
	LTE1800MHz UL 1Mbps	-96	-101	-113	-114	-121
	LTE800MHz DL 40Mbps	-78	-83	-93	-94	-101
	LTE1800MHz DL 50Mbps (desenhado 45Mbps)	-84	-89	-101	-102	-109

Figura 3.2.1 - Tabela representativa da capacidade da célula, em dBm.

Para a métrica 3G, RSCP, serviço HSDPA 1Mbps, definiu-se a capacidade em seis estados:

- Excelente – valores até -77 dBm;
- Muito bom –] -77; -82] dBm;

- Bom –] -82; -91] dBm;
- Razoável –] -91; -94] dBm;
- Mau –] -94; -103] dBm;
- Muito mau – valores inferiores a -103 dBm;

Já para o 4G, RSRP, serviço LTE 40 Mbps, foi definido a capacidade da célula em seis estados que diferem com os valores anteriormente vistos para 3G:

- Excelente – valores até -78 dBm;
- Muito bom –] -78; -83] dBm;
- Bom –] -83; -93] dBm;
- Razoável –] -93; -94] dBm;
- Mau –] -94; -101] dBm;
- Muito mau – valores inferiores a -101 dBm;

Para a taxa de ocupação, a célula pode ser caracterizada, em 4G, em estado “vermelho” se, por semana, 20% do tempo a velocidade for inferior a 3Mbps; “amarela” se estiver entre 1% e 20% do tempo, por semana, inferior a 3Mbps; e “verde” se for inferior a 3Mbps, menos 1% do tempo. Para 3G, os intervalos de tempo mantêm-se, alterando para uma velocidade inferior a 1Mbps.

Tabela 3.2.2 - Tabela representativa da Taxa de Ocupação da célula

QoE 4G	VERMELHO	AMARELO
4G: VEL. DL < 3 Mbps/10 Mbps (10 Mbps LTE1800)	Mais do que 20% tempo (por semana)	De 1% –20% do tempo (por semana)
3G: VEL. DL < 1 Mbps		

Com os dados retirados do GIS, é possível localizar os locais das células em que todos os testes foram realizados através da latitude e longitude. Na secção 4.2 são apresentados alguns mapas de Portugal onde é possível verificar a utilidade destes dados. Adicionalmente, é referido a marca do fornecedor da célula.

4 Validação e Resultados

Neste capítulo são mostrados todos os testes efetuados sobre o estudo da qualidade da Internet fixa na rede móvel e posteriormente, uma análise dos mesmos.

O processo de validação ocorreu sempre em paralelo ao processo de implementação para que, à medida que o projeto fosse avançando, os requisitos especificados e planeados inicialmente fossem cumpridos. Assim, o projeto foi dividido em quatro principais fases: recolha dos diferentes *datasets*, a fase da junção dos *datasets*, processamento e validação dos resultados do processamento.

4.1 Universo da análise

Definidos os principais pontos de análise, é importante fazer um levantamento do Universo do estudo. A validação do *dataset* de estudo diz respeito à fase inicial, em que, a partir dos cortes referidos na Tabela 4.1.1, foi possível construir um Universo de dados que consistente e capaz de corresponder as dimensões de análise definidas na secção 3.2.1.

Tabela 4.1.1 – Dimensões sobre o Universo de análise

Horário	Terminal	Tech	KPI Sinal	Banda
On-Peak [20h – 24h]	B310	3G	3G – RSCP	3G– 2100 MHz
Off-Peak	E5172	4G	4G - RSRP	4G – 800 MHz 4G– 1800 MHz 4G– 2600 MHz

Tx. Ocupação	Processo	Produto	Fornecedor	Evolução diária
Células Verde	Instalação	DTH	Ericsson	Evolução das velocidades ao longo do tempo.
Células Amarelo	2 ^{as} linhas (despiste)	WOW	Huawei	Evolução das velocidades 3G e 4G ao longo do tempo.
Células Vermelho				

Assim, numa análise da performance dos testes efetuados no período entre dia 1 de janeiro de 2017 a 28 de junho de 2017, tem-se uma amostra de, aproximadamente, quinze mil clientes distintos e cerca de trinta e nove mil testes efetuados.

Dentro dessa amostra foram retirados os dados reais para a análise. Veja-se a Figura 4.1.1, dos 15 mil clientes totais, cerca de 77% foram válidos, o que dá perto de onze mil clientes, com um total de vinte e cinco mil testes válidos. Portanto, foram eliminados aproximadamente três mil clientes (23% dos quinze mil clientes distintos do Universo total) e catorze mil testes efetuados (35% dos trinta e nove mil testes do Universo total).

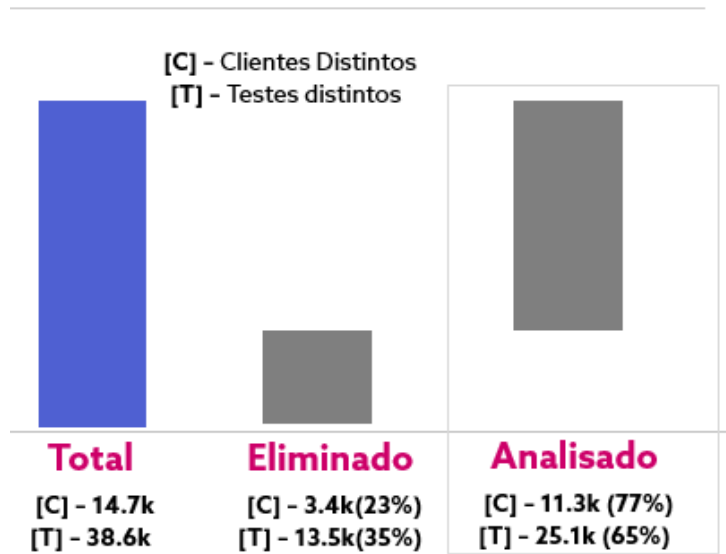


Figura 4.1.1 - Universo do estudo e Amostra analisada.

Esta é uma análise com detalhe os testes inválidos, e como foi dito e justificado na secção 3.2.2, alguns dados não faziam sentido permanecer no Universo de análise. Portanto, considera-se que esta análise inclui testes programados que não foram devidamente executados (por exemplo, terminal desligado, testes falhados por falta de cobertura, por *timeout*, etc.), bem como dados com campos não preenchidos (por exemplo, campos da célula e *download/upload* ambos a *null*).

Referindo, especificamente, o universo das células analisadas neste estudo, obteve-se a distribuição das mesmas por Portugal continental e ilhas (Açores e Madeira) e está ilustrado na Figura 4.1.2. Observa-se que, perante a Figura 4.1.2, tem-se uma completa distribuição das células por Portugal inteiro, o que indicia um estudo que cobre as diferentes regiões do país.

De referir que esta primeira abordagem inclui testes dos serviços DTH e WOW. Na próxima secção 4.2 irá ser introduzido esse corte, distinguindo todos os testes efetuados pelos diferentes serviços, sabendo que o único serviço em estudo será o DTH.

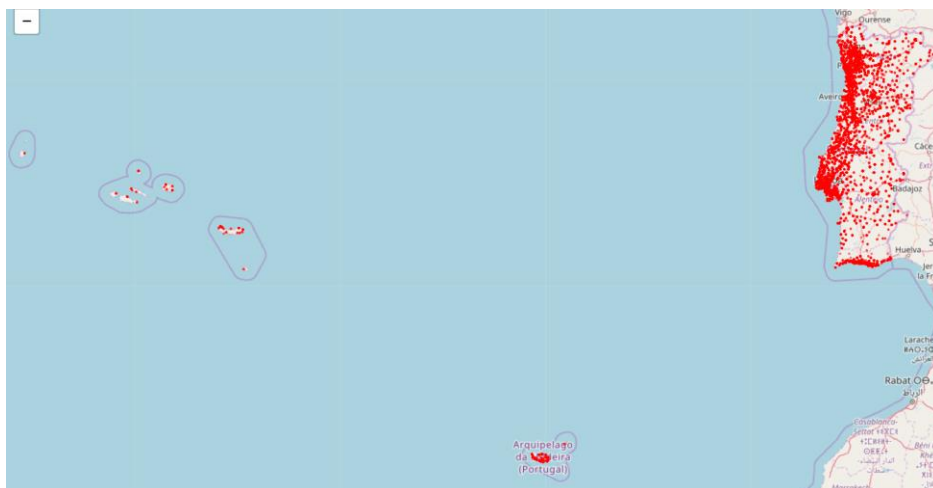


Figura 4.1.2 - Distribuição das células estudadas no projeto.

4.2 Resultados Obtidos

Como se tem vindo a falar ao longo deste documento, pretende-se estudar a experiência de utilização da internet fixa dos clientes sobre a rede móvel (Internet fixa NOS e DTH). Através dos dados recolhidos pelas diversas fontes de dados, foi validado a maior presença do DTH em detrimento do serviço WOW. Portanto, como esperado, a Figura 4.2.1 mostra a distribuição por Produto (DTH), Tecnologia (3G/4G) dos testes recolhidos.

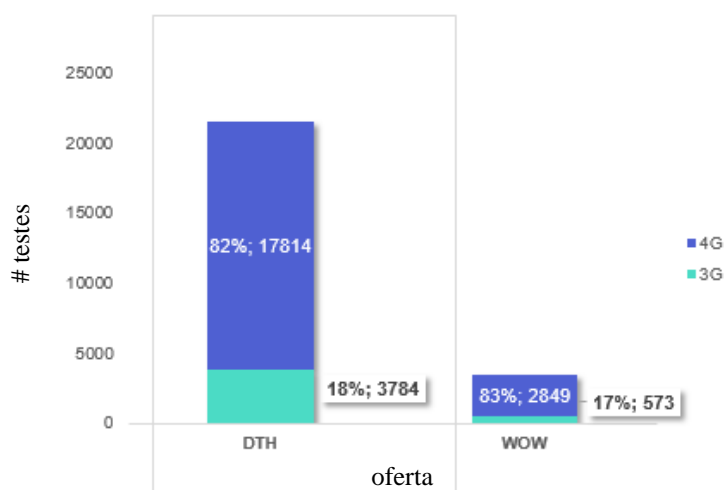


Figura 4.2.1 - Distribuição por Produto, Tecnologia.

Assim, é possível verificar que os dados representam significativamente o serviço DTH, com 86% dos testes (21 mil testes) e 87% de IMSI, ou seja, próximo de 10 mil clientes com IMSIs diferentes. Ainda é representado a percentagem da quantidade de amostras de cada tecnologia por serviço. Em ambos os casos, os testes efetuados em 4G representam, maioritariamente, o Universo com 82% no DTH e 83% no WOW. Já os testes realizados em 3G, 18% no DTH e 17% no WOW.

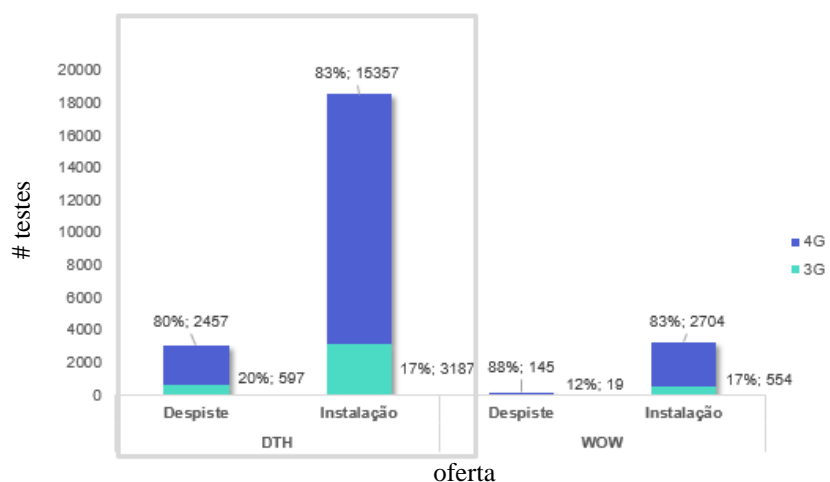


Figura 4.2.2 - Distribuição Produto, Processo.

Na Figura 4.2.2 foi estudado a distribuição do Produto (DTH/WOW), Processo (dados dos testes de velocidade no processo de instalação e para despiste). Pode confirmar-se que a fonte de dados ACS coleciona mais registos no processo de instalação. Na prática pode-se dizer que ACS não representa efeitos de *home networking*.

No que diz respeito à distribuição da tecnologia pelos diferentes produtos e processos, uma vez mais, é notório o domínio do 4G, igualando o mesmo rácio de testes 4G/3G em DTH/WOW no processo de instalação que a Figura 4.2.1 representa.

Posteriormente, foi estudado a distribuição por tecnologia, pelo IMSI e pelos testes realizados. Ver Figura 4.2.3

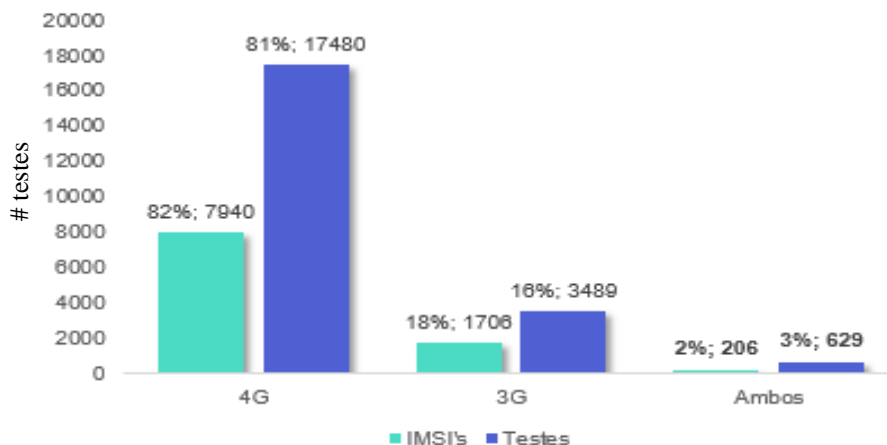


Figura 4.2.3 - Distribuição por Tecnologia.

Dos *speedtests* observados, não existe alternância entre 3G e 4G, o que quer dizer que grande parte dos clientes está em 4G e aí permanecem. Em relação aos clientes que estão em 3G representam 18% dos clientes distintos (18% de IMSIs).

Valida-se que apenas 2% dos clientes sofreu mudanças de tecnologia, mas visto ser uma percentagem tão pouca de clientes considera-se um dado praticamente irrelevante.

O domínio dos testes de velocidade feitos em 4G, visto na figura acima, é confirmado pela Figura 4.2.4, onde estão mapeadas as regiões de Portugal Continental e a respetiva tecnologia dominante. A distribuição por cores corresponde ao número de testes efetuados nessa mesma tecnologia.

Apenas Beja apresenta um domínio de testes efetuados em 3G, apesar do número de testes serem bastante poucos, menos de 500 testes.

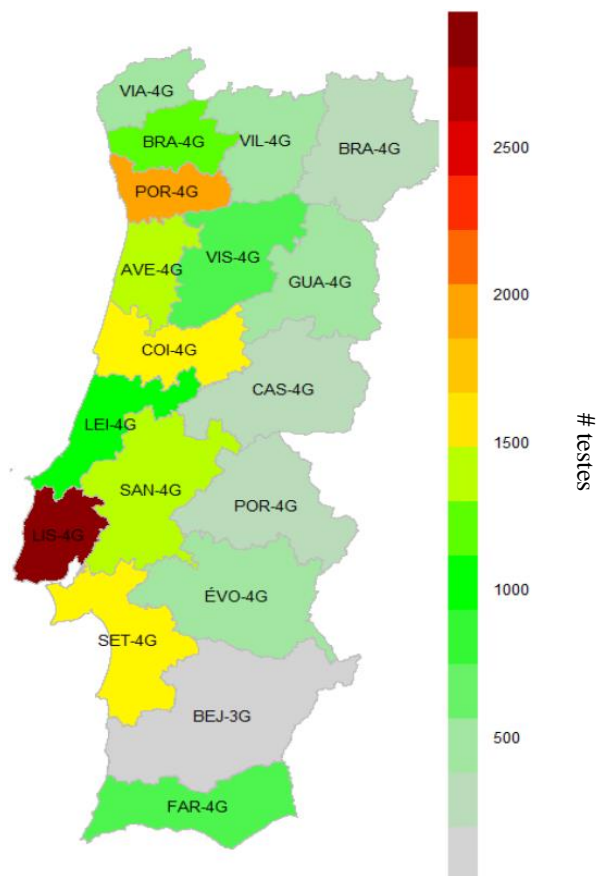


Figura 4.2.4 - Mapeamento por Tecnologia.

Em análise da distribuição por fornecedor, Figura 4.2.5, reparou-se que havia uma maior incidência de testes com equipamentos Huawei. Deve-se, muito provavelmente, a novas instalações com maior concentração na zona norte. Isto é, existe maior número de equipamentos Huawei na zona norte de Portugal, já a sul de Portugal os equipamentos predominantes são Ericsson, visível na Figura 4.2.6.

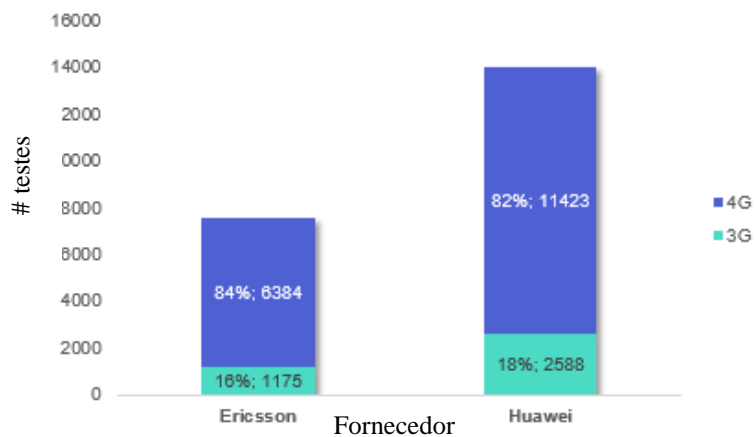


Figura 4.2.5 - Distribuição por Fornecedor, Tecnologia.

A distribuição por cores nos diferentes distritos indica-nos a quantidade de teste que corresponde à tecnologia indicada nessa mesma região.

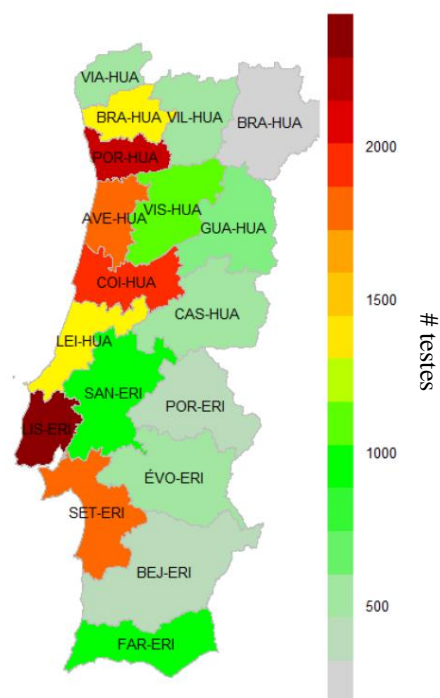


Figura 4.2.6 - Mapeamento por Fornecedor.

Observa-se ainda que existe um rácio semelhante dos testes efetuados em 3G/4G nos fornecedores Ericsson/Huawei. Relativamente aos terminais (B310 e E5172) usados nos testes, pode-se dizer que o router B310 é o terminal mais recente e que veio substituir

o E5172, pelo que era de esperar que houvesse uma maior distribuição de testes no router E5172, como mostra a Figura 4.2.7.

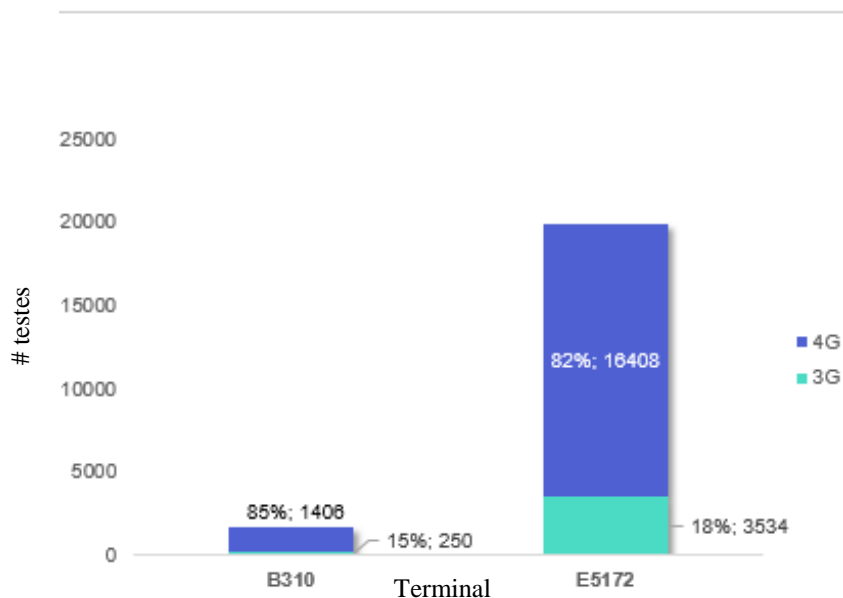


Figura 4.2.7 - Distribuição por Terminal, Tecnologia.

Pelo gráfico acima, pode-se constatar que existe, aproximadamente, a mesma distribuição de testes 3G e 4G para ambos os terminais:

- B310: 15% testes 3G e 85% testes 4G;
- E5172: 18% testes 3G e 82% testes 4G;

Confirmando a distribuição dos terminais usados neste estudo, fez-se o mapeamento por Terminal ao longo de Portugal Continental com preenchimento colorido consoante o número de testes por terminal correspondentes às regiões (Figura 4.2.8).

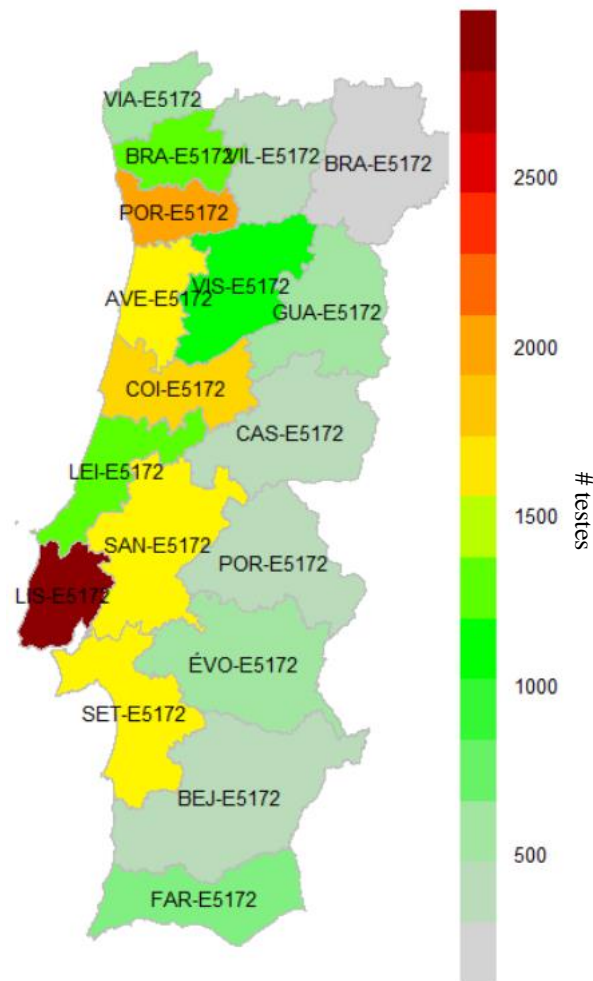


Figura 4.2.8 - Mapeamento por Terminal.

Passou-se, então, para os testes relacionados com a experiência do utilizador, ou seja, onde a variável principal, *throughput*, está presente nas análises. Observe-se a Figura 4.2.9.

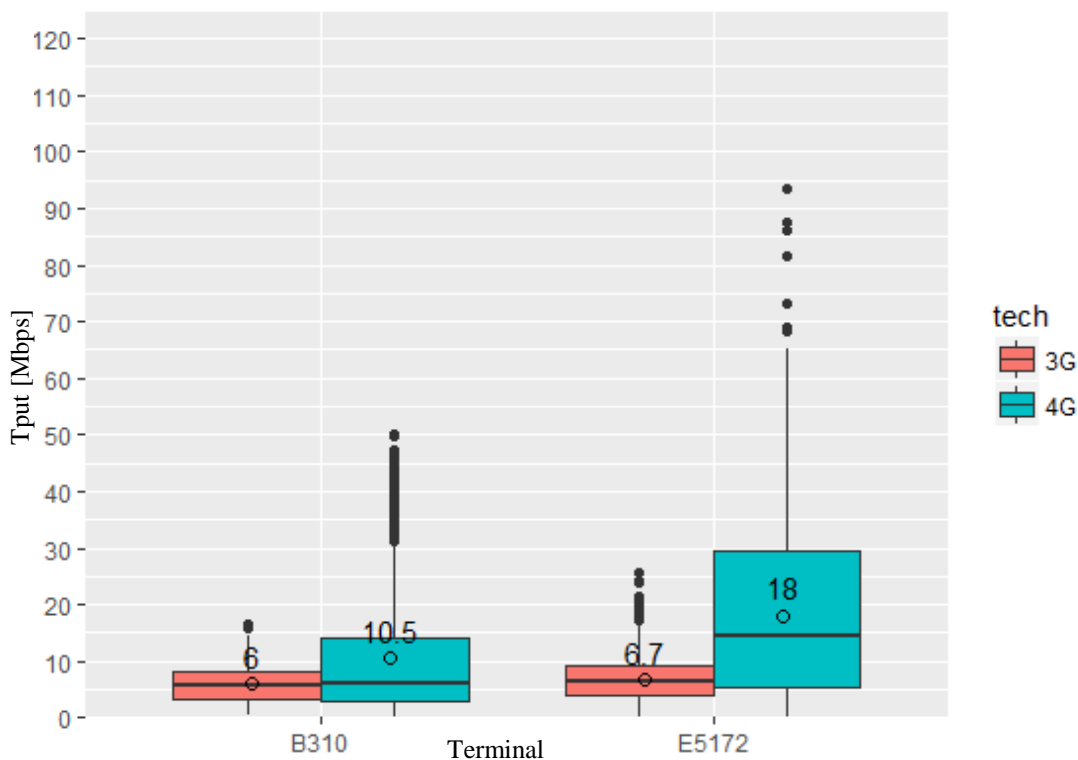


Figura 4.2.9 - Velocidade por Terminal, Tecnologia.

Numa primeira análise, pode-se retirar que os routers B310 exibem uma pior performance em 4G comparado com o terminal E5172 e, em 3G com desempenhos semelhantes. Uma vez que, o terminal B310 representa apenas 13% dos nossos testes pode ter influência nos resultados. Assim, será feita uma análise mais intensiva no que toca a todas as variantes das velocidades nos terminais.

Como foi detalhado na secção 3, no processo de instalação são definidos dois horários: o horário com maior congestionamento, dito “on-peak”, entre as 20:00 horas e as 24:00 horas; e em horas que, supostamente, o congestionamento é pouco, “off-peak”, definiu-se um horário das 04:00 horas até 08:00 horas. Os testes que ocorreram, fora desse intervalo de tempo, designam-se “offpeakI”.

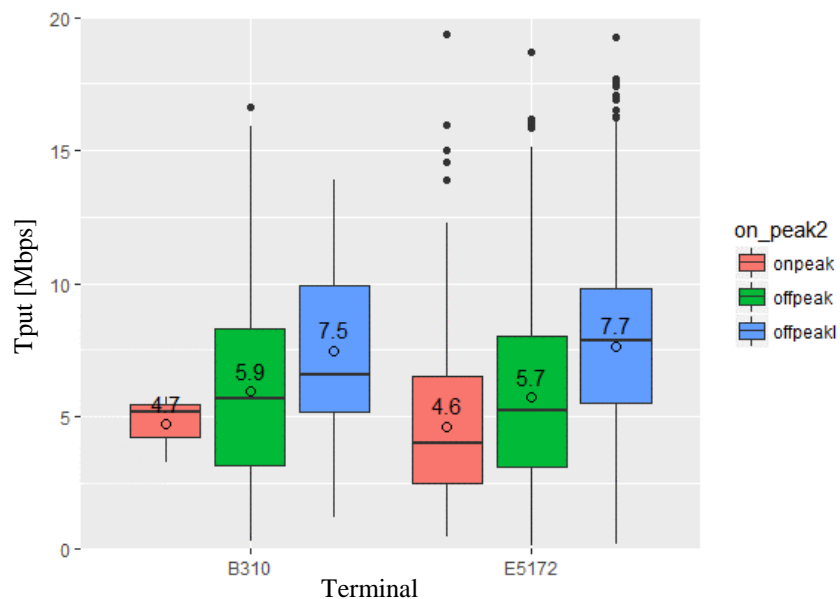


Figura 4.2.10 - Velocidade por Terminal, Horário [3G].

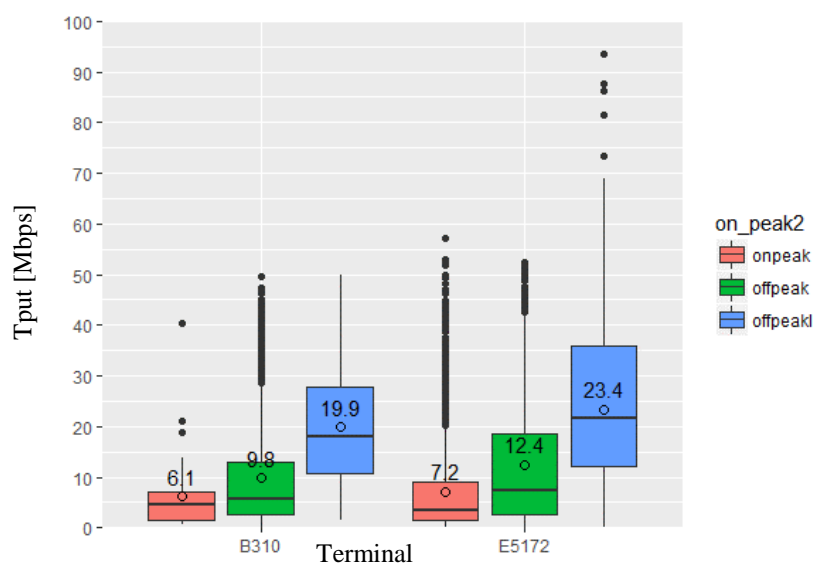


Figura 4.2.11 - Velocidade por Terminal, Horário [4G].

Como era de esperar, pela Figura 4.2.10 e Figura 4.2.11, em horário de maior congestão, a velocidade medida em ambos os terminais nas diferentes redes móveis (4G e 3G) é mais baixa comparando com os demais. Em 3G, é observável um equilíbrio das velocidades nos terminais nos diferentes horários do dia. Já em 4G, B310 continua com pior performance nos cenários em períodos de utilização semelhante. Uma vez mais, é importante referir que a amostra reduzida pode reduzir a confiança da análise.

Introduzindo a taxa de ocupação no estudo, através da base dados Spiro, representado na Figura 4.2.12 e Figura 4.2.13, permitiu ter conhecimento das velocidades praticadas consoante o estado de congestão das células em função dos terminais. Visto ser uma amostra reduzida do terminal B310 e, adicionalmente, existirem poucos testes em 3G é compreensível que não exista amostras com uma taxa de ocupação vermelha (Figura 4.2.12). Pode-se observar ainda que as velocidades médias recolhidas em 3G para o terminal E5172 são semelhantes.

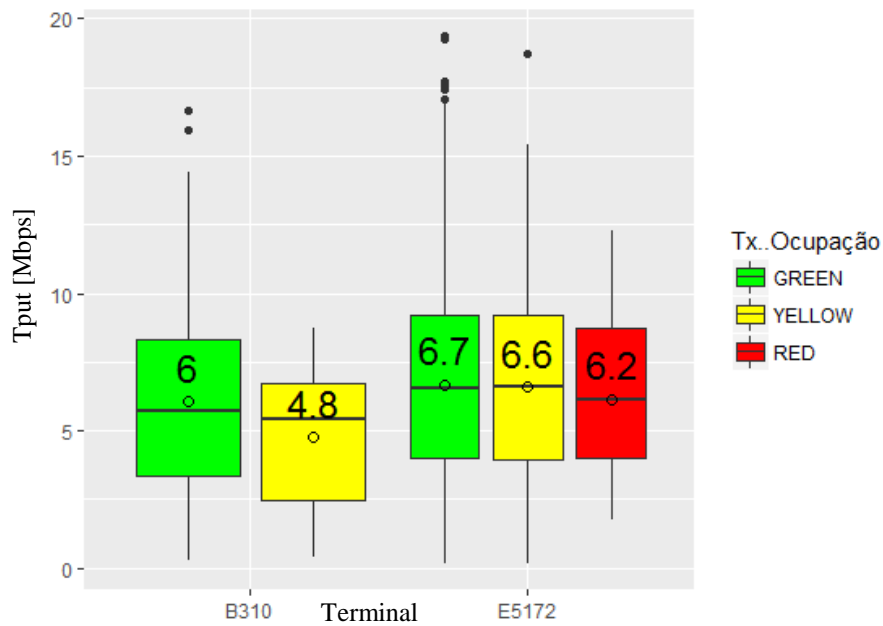


Figura 4.2.12 - Velocidade por Terminal, Tx. Ocupação [3G].

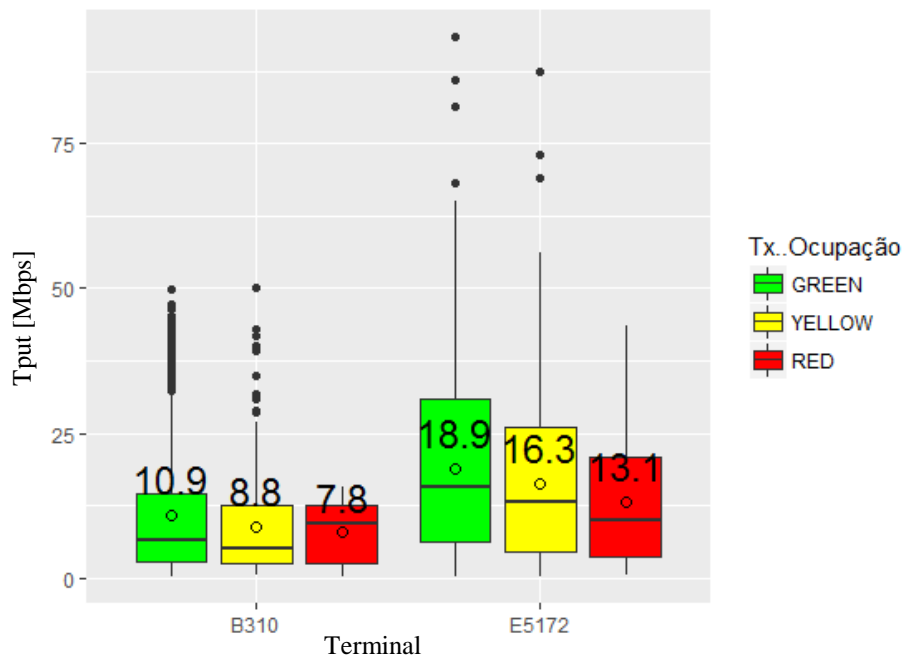


Figura 4.2.13 - Velocidade por Terminal, Taxa Ocupação [4G].

Em 4G, Figura 4.2.13, é perceptível que existe diferença de velocidades no mesmo terminal para as diferentes taxas de ocupação, o que valida a premissa de menores velocidades quanto maior for a taxa de ocupação da célula. Ainda no 4G, B310 continua com pior performance em células com a mesma taxa de ocupação.

Com o objetivo de perceber se o elevado número de *churn* na oferta DTH está relacionado com velocidades menores comparando à oferta WoW, estudou-se as velocidades em 3G e 4G em ambos os serviços. Veja-se a Figura 4.2.14.

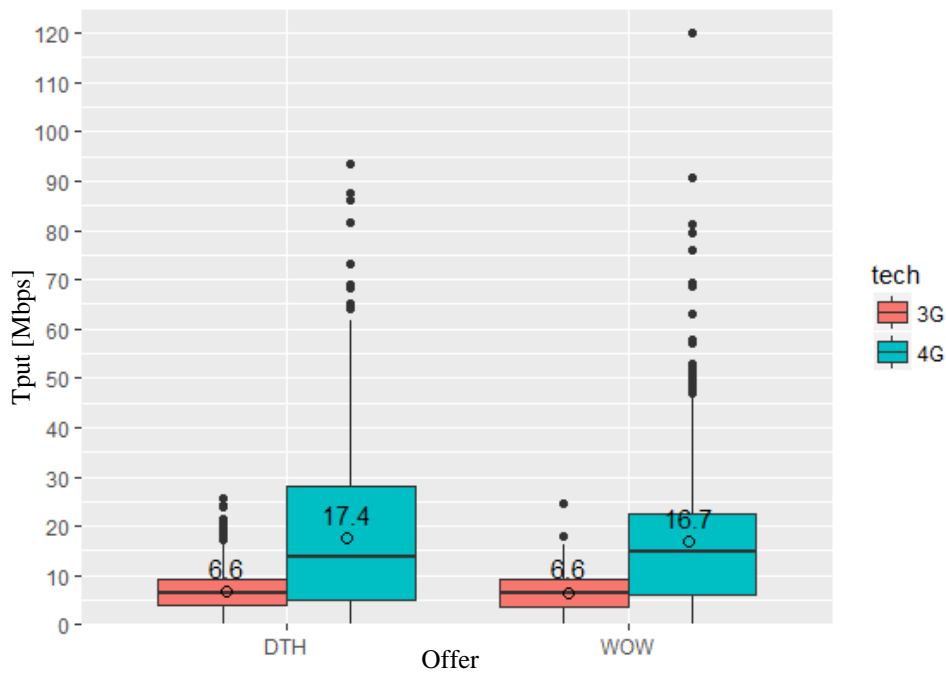


Figura 4.2.14 - Velocidade por Produto, Tecnologia.

Então, pela Figura 4.2.14, não se consegue justificar o elevado número de *churn* pela velocidade de oferta. Apesar de em 3G as velocidades serem idênticas, em 4G o serviço DTH tem um *throughput* superior a oferta WoW. Ainda assim, as velocidades encontram-se a baixo do esperado ou contratado pelo cliente, pelo que deviam estar mais próximas das velocidades oferecidas, 20 Mbps ou 40 Mbps.

Outro ponto importante validado é mostrado na Figura 4.2.15, em que as velocidades obtidas no processo de despiste são inferiores comparadas as velocidades obtidas no processo de instalação.

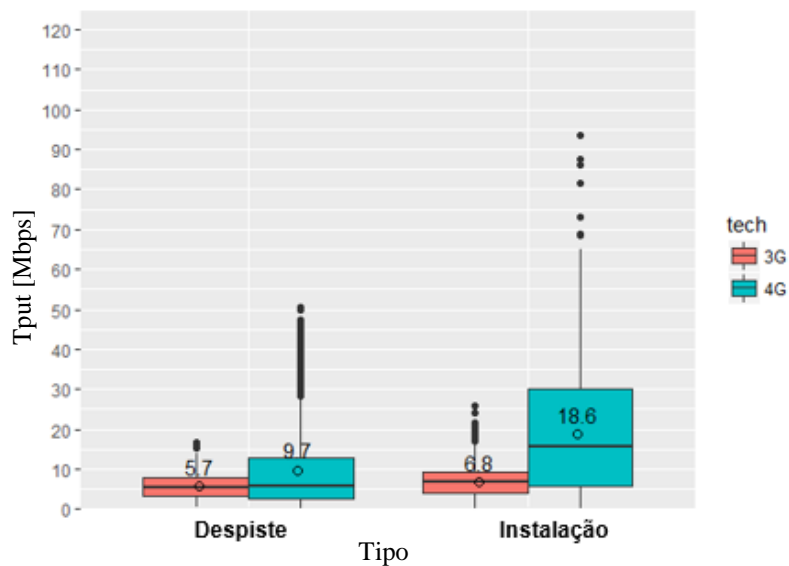


Figura 4.2.15 - Velocidade por tipo Teste, Tecnologia.

Intuitivamente, essa diferença de velocidades nos diferentes tipos de processo é justificada nos testes, maioritariamente, em 4G. Os testes efetuados no processo de instalação não são representativos da experiência de utilização do cliente, mas é o que mais se aproxima de tal.

Pelo esperado, em regra geral, as velocidades obtidas ao longo dos diferentes intervalos do dia no processo de instalação são superiores as do processo de despiste. No horário de menor congestão, “offpeak” e “offpeakI”, as velocidades obtidas no processo de despiste são inferiores quando comparado com as de instalação. Apenas num caso, a velocidade nos despistes é superior as da instalação, e está ilustrado na Figura 4.2.17.

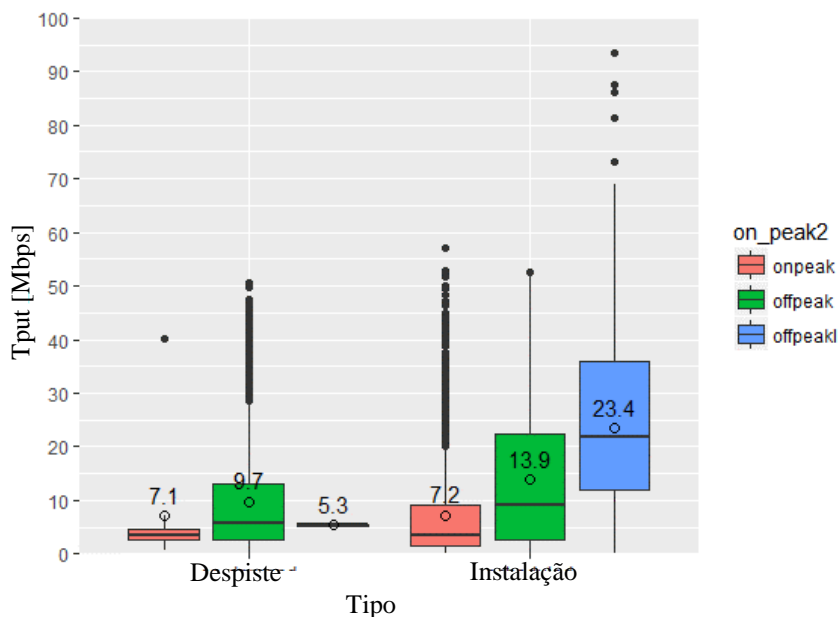


Figura 4.2.16 - Velocidade por Processo, Horário [4G].

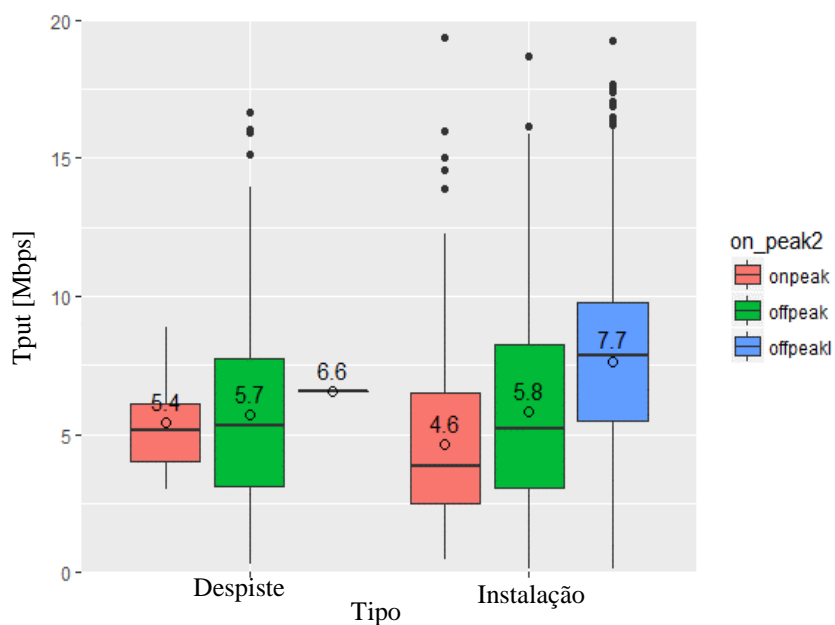


Figura 4.2.17 - Velocidade por Processo, Horário [3G].

E m 3G, na hora de maior afluência (“onpeak”), o processo de despiste tem uma ligeira vantagem em relação ao outro tipo de processo. Já em 4G a Figura 4.2.16 corrobora o facto de as velocidades obtidas no processo de despiste sejam inferiores, ou seja, com uma pior performance.

Observa-se, agora, os resultados obtidos pelo processamento dos dados pelos fornecedores. A Figura 4.2.18 mostra-nos as velocidades obtidas nas redes 3G e 4G nos equipamentos Huawei e Ericsson.

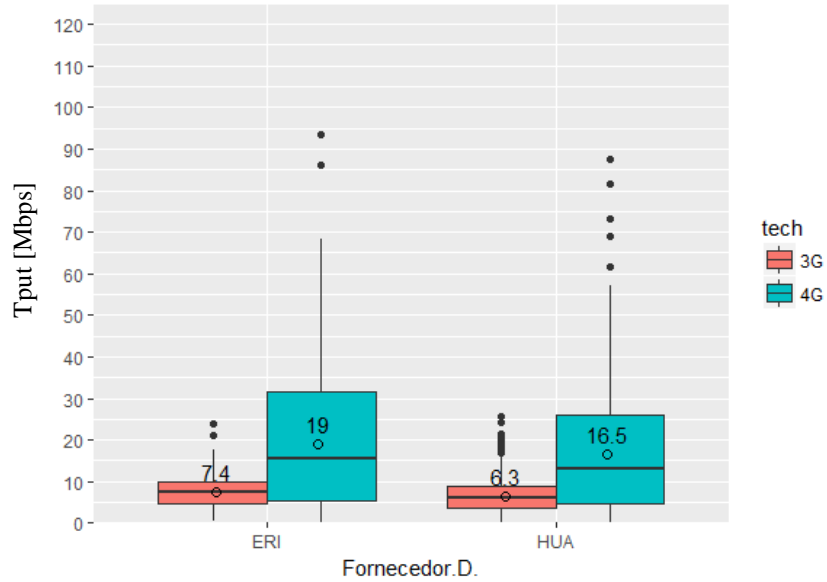


Figura 4.2.18 - Velocidade por Fornecedor, Tecnologia.

Pode-se constatar que a velocidade no fornecedor Ericsson é ligeiramente superior à Huawei, tanto para 3G como 4G. Para uma análise mais consistente, foi necessário efetuar o estudo da velocidade do fornecedor por Taxa de ocupação da célula, potência do sinal e em diferentes horários ao longo do dia.

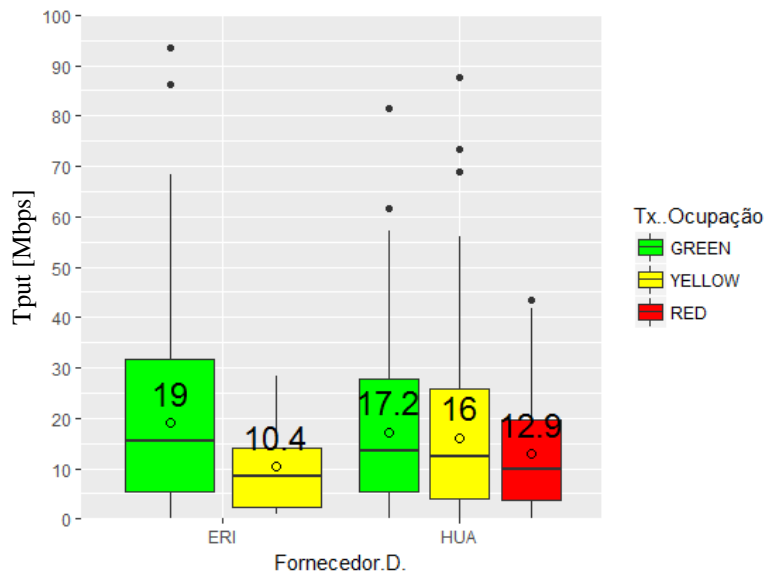


Figura 4.2.19 - Velocidade por Fornecedor, Taxa Ocupação [4G].

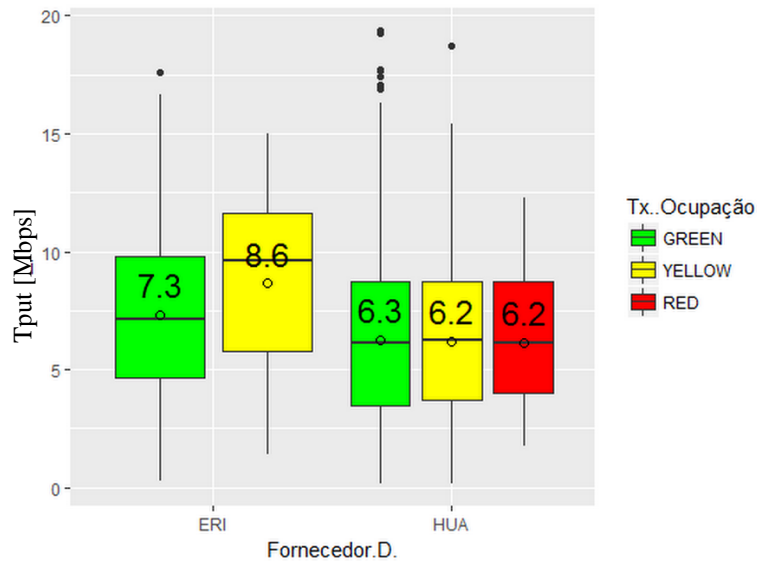


Figura 4.2.20 - Velocidade por Fornecedor, Taxa Ocupação [3G].

A Figura 4.2.19 mostra a velocidade por fornecedor e por taxa de ocupação em rede 4G e a Figura 4.2.20 em 3G. No geral, nota-se um domínio das velocidades obtidas em Ericsson nas diferentes taxas de ocupação da célula. Um dado curioso é, a Huawei em rede 3G tem *throughput* bastante idêntico nos três tipos de utilização da célula.

Relativamente, à potência do sinal em que é dividida por seis características diferentes explicadas na secção 3.2.2, é analisado através da Figura 4.2.21 e Figura 4.2.22. Os gráficos continuam a provar que, independentemente da potência de sinal caracterizada no momento do teste de velocidade, a Ericsson mantém-se com velocidades ligeiramente superiores relativamente à Huawei, quer seja em 3G ou 4G.

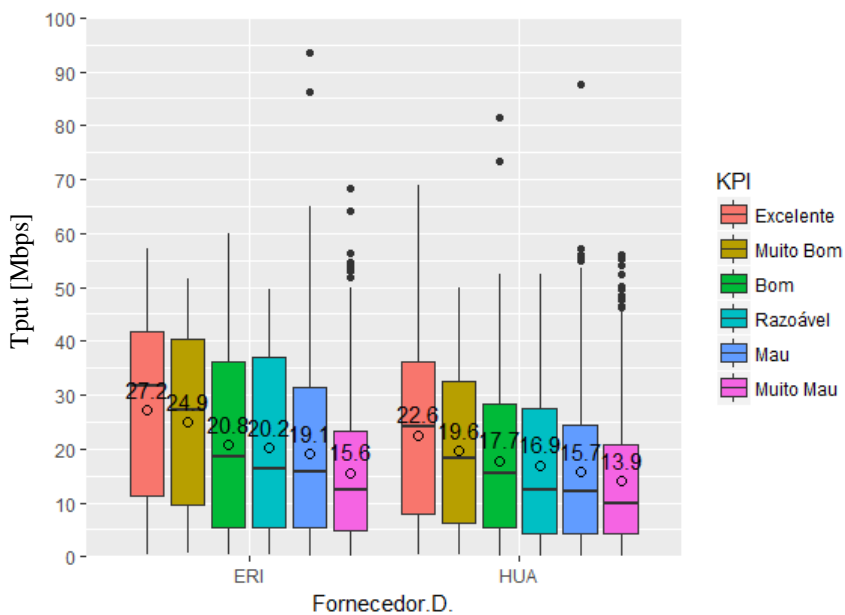


Figura 4.2.21 - Velocidade por Fornecedor, KPIs [4G].

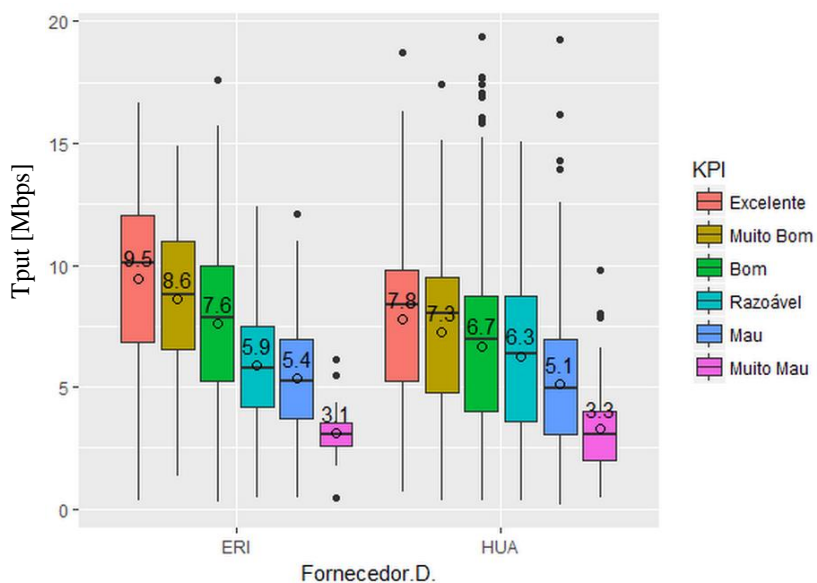


Figura 4.2.22 - Velocidade por Fornecedor, KPIs [3G].

Os gráficos Figura 4.2.24 e da Figura 4.2.23 confirmam todas as anteriores análises relativamente à ligeira superioridade da Ericsson. Seja em 4G ou 3G, Ericsson tem um melhor desempenho no que toca ao tema das velocidades em qualquer período do dia.

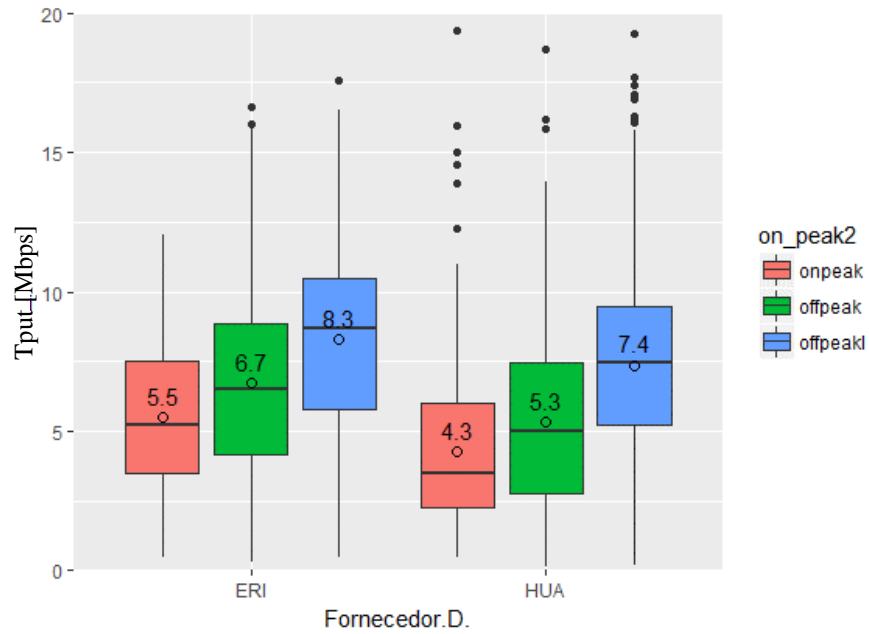


Figura 4.2.23 - Velocidade por Fornecedor, Horário [3G].

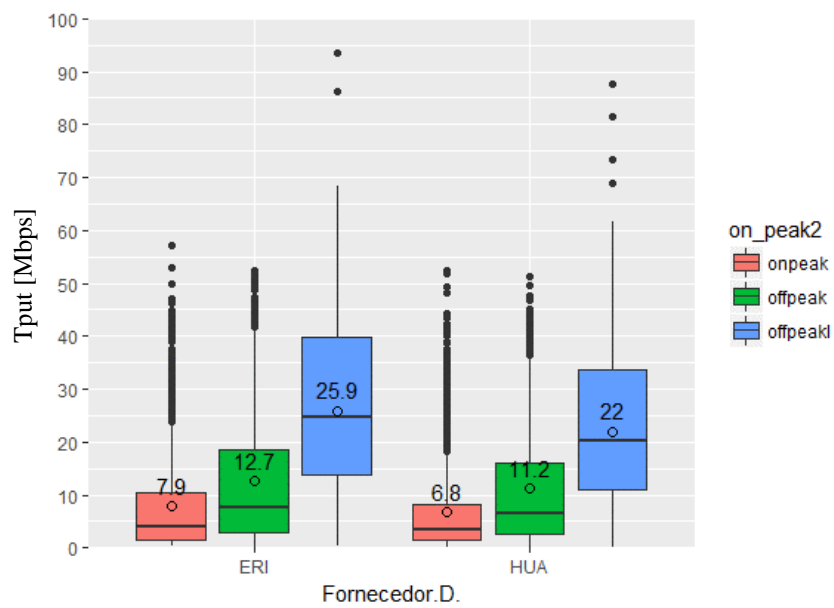


Figura 4.2.24 - Velocidade por Fornecedor, Horário [4G].

Ora, um dos principais pontos deste estudo é representado na Figura 4.2.25, onde se consegue ver as diferentes bandas usadas nos testes de velocidade.

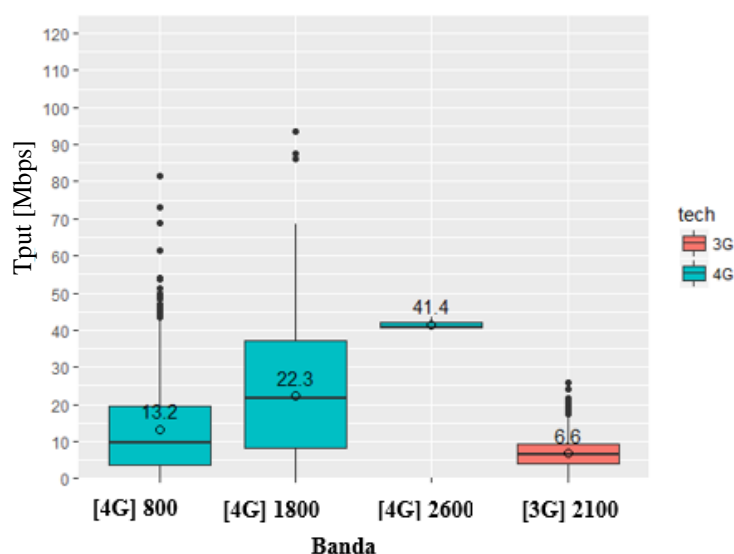


Figura 4.2.25 - Velocidade por Banda, Tecnologia.

Existem três bandas em rede LTE (4G):

- LTE 800 MHz;
- LTE 1800 MHz;
- LTE 2600 MHz;

E uma banda no sistema UMTS (3G), que é 2100 MHz. Vendo a Figura 4.2.25 confirma-se que os testes de velocidade em 3G têm um *throughput* inferior aos testes de velocidade em 4G, independentemente da banda LTE usada. Refira-se, também, que a velocidade é superior à medida que banda aumenta, ou seja, em LTE 2600 MHz tem-se um *throughput* médio de 41.4 Mbps, em LTE 1800 MHz um *throughput* médio de 22.3 Mbps e em LTE 800 MHz um *throughput* médio de 13.2 Mbps.

Assim, pode-se concluir que as médias das velocidades obtidas nas bandas LTE 1800 MHz e 2600 MHz estão alinhadas com o produto oferecido, que são velocidades até 20 Mbps e até 40 Mbps.

Sabendo que as velocidades atingidas pelos clientes podem variar pela largura de banda, da qualidade do canal e da sobrecarga da rede [33], foi analisado as velocidades obtidas segundo os critérios pré-estipulados, na secção 3.2.2, em relação a taxa de ocupação e da qualidade e potência do sinal.

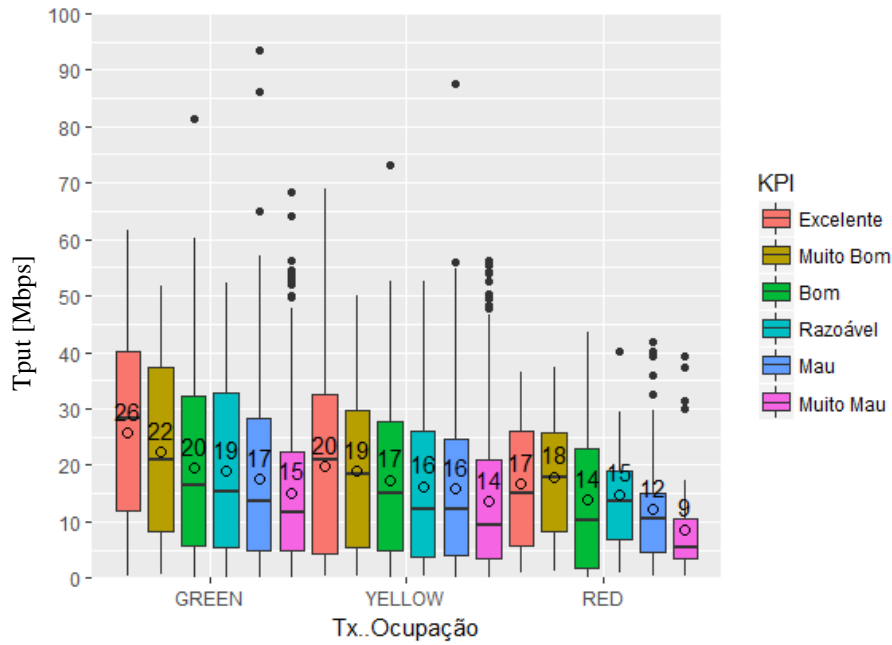


Figura 4.2.26 - Velocidade por Taxa Ocupação, KPIs [4G].

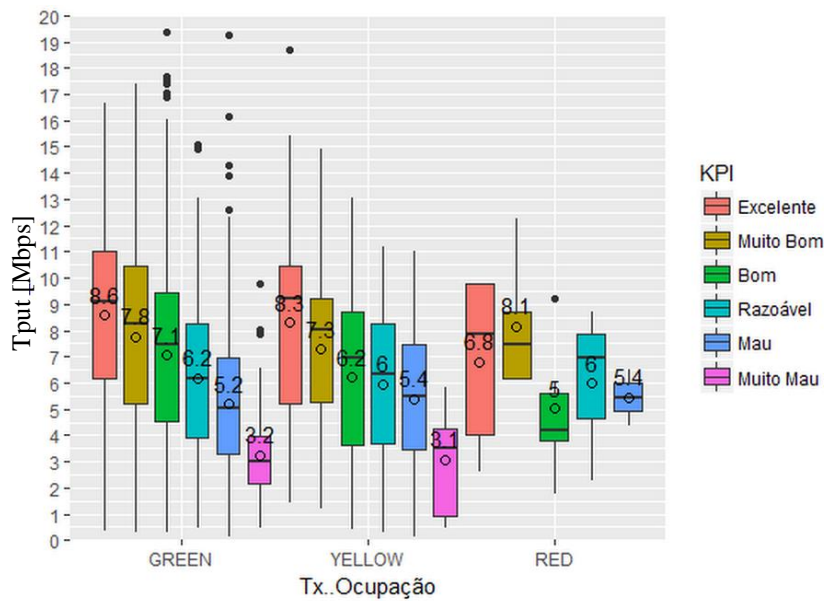


Figura 4.2.27 - Velocidade por Taxa Ocupação, KPIs [3G].

Tanto em 3G como em 4G, Figura 4.2.27 e Figura 4.2.26 respectivamente, é visível que quanto melhores forem as condições, isto é, quanto melhor for a potência do sinal e estando a célula menos congestionada, as velocidades obtidas são sempre superiores. Uma outra maneira de ver os resultados obtidos foi, relacionar a velocidade por potência do sinal (KPI), taxa de Ocupação. Veja-se a Figura 4.2.28 e Figura 4.2.29.

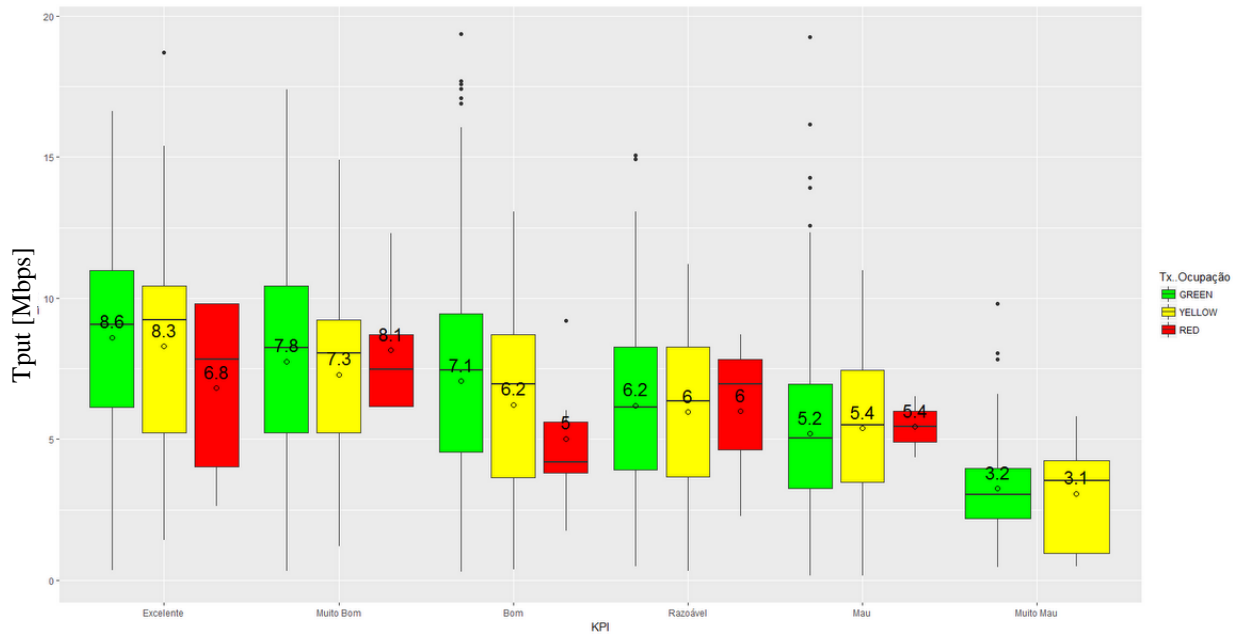


Figura 4.2.28 - Velocidade por KPIs, Taxa Ocupação [3G].

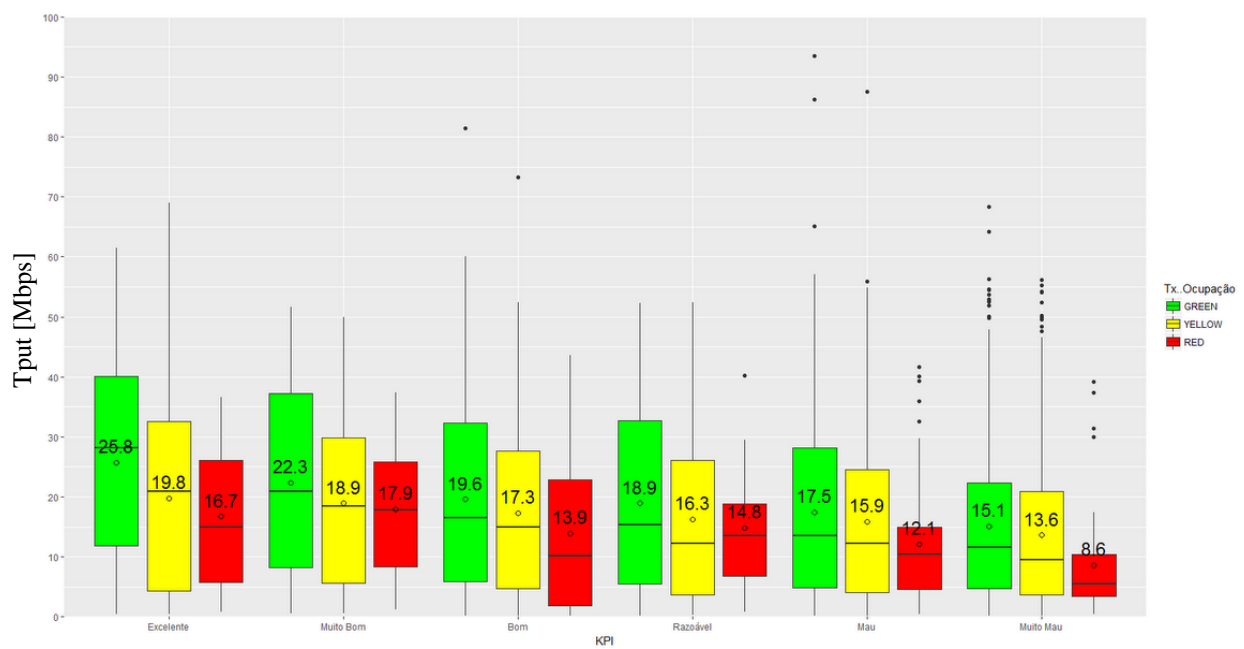


Figura 4.2.29 - Velocidade por KPI, Taxa Ocupação [4G].

Apesar de, em determinadas situações, o sinal se apresentar como “Excelente” e com uma taxa de ocupação “green” (célula pouco congestionada), as velocidades encontram-se abaixo do oferecido, tanto em 3G como em 4G. No entanto, é possível verificar que para situações idênticas em relação à qualidade do sinal e à taxa de ocupação da

célula, o 4G apresenta valores de velocidades sempre superiores aos valores de velocidades no 3G.

Para reforçar a conclusão acima referida, calculou-se as velocidades, nas diferentes horas do dia, por tecnologia (3G e 4G), como é ilustrado na Figura 4.2.30. Nas condições em que há um maior congestionamento, ou seja, em “onpeak”, o 3G e o 4G apresenta velocidades maiores nas horas entre “onpeak” e “offpeak”.

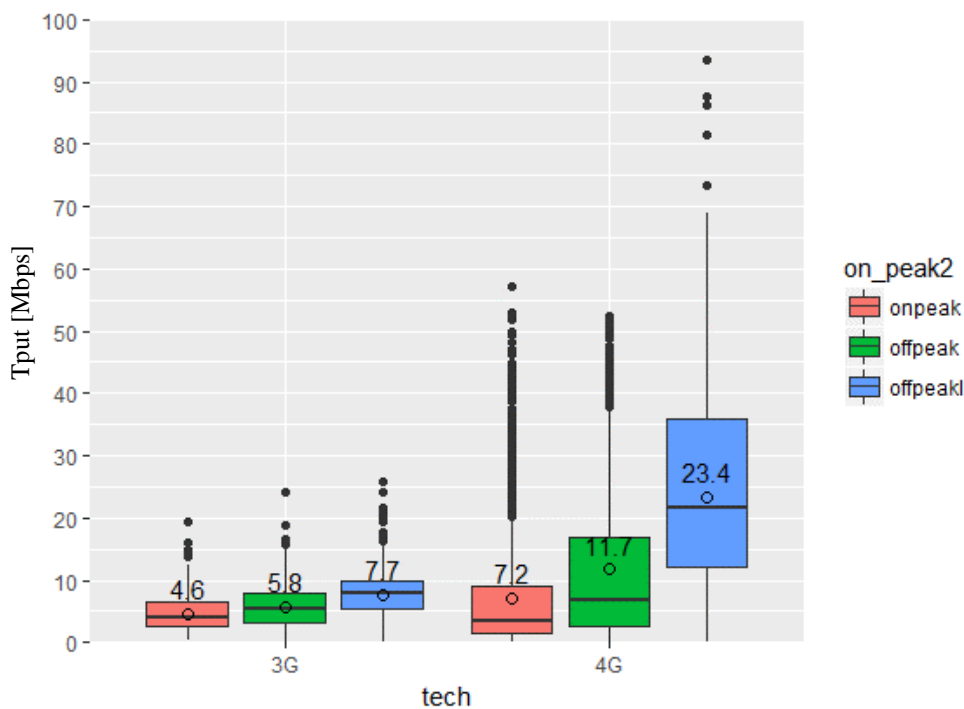


Figura 4.2.30 - Velocidade por Tecnologia, Horário.

Também pode-se evidenciar velocidades maiores devido ao intervalo de tempo em “offpeakI” ser maior e apresentar um maior número de testes efetuados. Comparando agora, apenas os intervalos de tempo “onpeak” e “offpeak”, confirma-se o que era esperado. Em “offpeak” tem-se velocidades bastante mais favoráveis para a experiência do utilizador, em ambas as tecnologias.

4.3 Evolução horária de *throughput*

Para uma visão genérica sobre a experiência de utilização, calculou-se a evolução horária ao longo do dia, através dos dados colecionados para este estudo. Veja-se agora como a velocidade da Internet fixa do serviço DTH em 3G varia ao longo do dia.

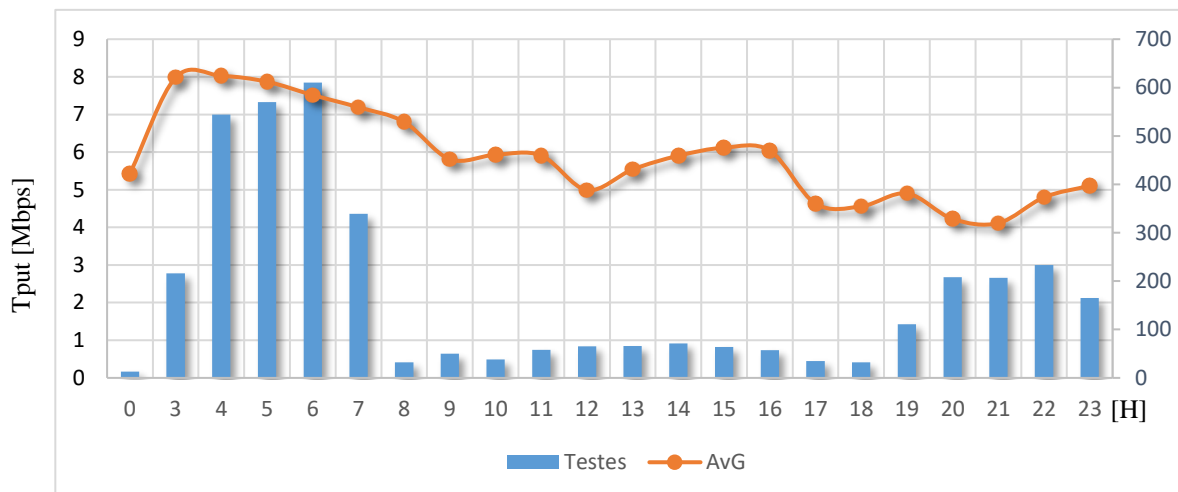


Figura 4.3.1 - Evolução horária em 3G e número de testes efetuados por hora.

Pela Figura 4.3.1, exatamente como esperado, observa-se um maior número de testes de velocidade em horário de menor congestão, pois no processo de instalação são efetuados dois testes em “offpeak”. Relativamente à velocidade média do horário “onpeak” e “offpeak” é possível verificar que, em “offpeak” atinge-se velocidades médias maiores comparadas as velocidades médias atingidas em horário de maior congestão.

Em 4G, a evolução horária é representada na Figura 4.3.2, e tal como se sucedeu em 3G, existe um número substancialmente maior de testes efetuados em “offpeak”, ou seja, entre as 04 horas e as 08 horas, apresentado também velocidades médias superiores neste respetivo horário. Existe uma maior diferença entre velocidades médias em “onpeak” e “offpeak” em 4G, do que em 3G como se pode confirmar na Figura 4.3.1 e Figura 4.2.30.

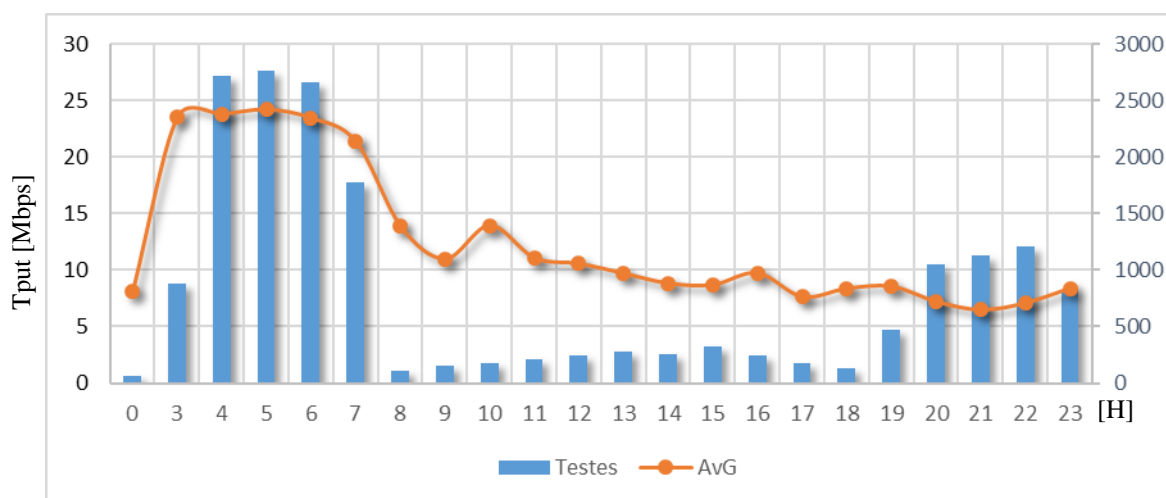


Figura 4.3.2 - Evolução horária em 4G e número de testes efetuados por hora.

Em título de comparação, juntou-se os *throughputs* médios na Figura 4.3.3. Como era de esperar, as velocidades em 4G apresentam-se francamente superiores que as velocidades em 3G. É visível ainda, que existe uma maior oscilação de velocidades em 4G, especialmente, no horário de menor congestionamento, entre as 04 da manhã e as 08 da manhã, para o resto do dia. Em 3G, presencia-se pouca oscilação no que diz respeito às velocidades médias por hora ao longo do dia.

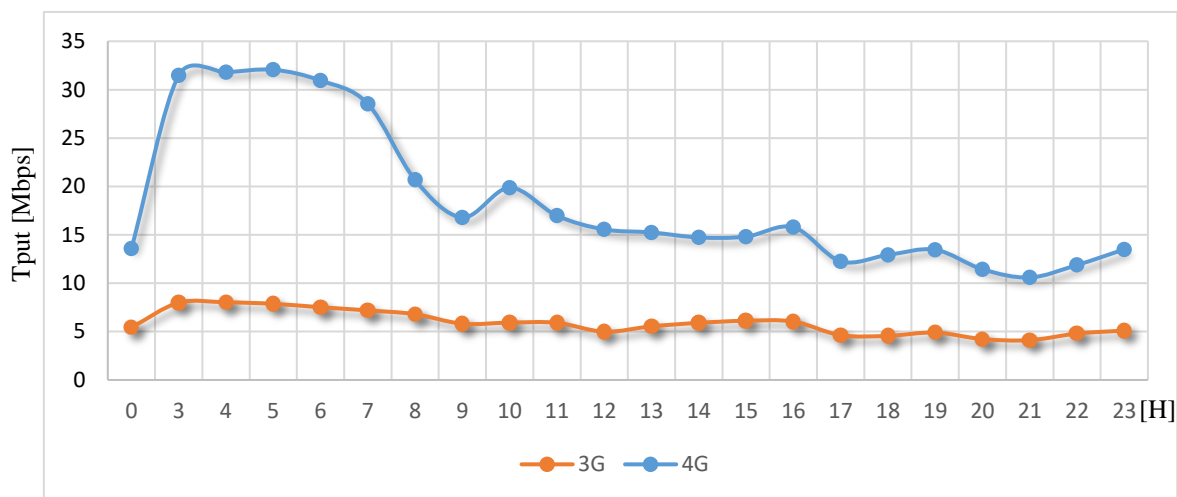


Figura 4.3.3 - Comparação das velocidades médias obtidas em 3G e 4G.

Num registo de mapeamento das velocidades médias por região, obteve-se a Figura 4.3.4. Guarda, Lisboa e Faro, foram as regiões de Portugal Continental que conseguiram obter as três melhores velocidades médias com 18.86 Mbps, 17.45 Mbps e 17.38 Mbps, respetivamente. Por sua vez, Leiria, Viseu e Coimbra, tiveram o pior registo com, apenas, 13.41 Mbps, 13.7 Mbps e 13.81 Mbps.

Conclui-se ainda que, as regiões de Portugal, num modo geral, têm velocidades médias algo idênticas, com uma diferença de, aproximadamente, 5 Mbps entre a maior e a menor velocidade média obtida (Guarda e Leira, respetivamente).

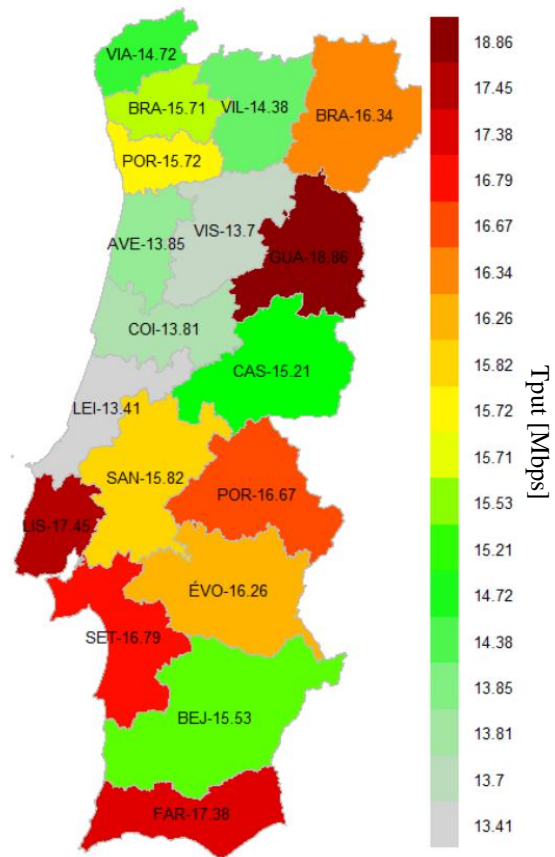


Figura 4.3.4 - Mapeamento das velocidades médias em Portugal.

5 Conclusões

As conclusões circunscrevem o último capítulo desta dissertação. Neste capítulo será elaborada uma síntese de todo o trabalho realizado, em que serão referidos os aspetos mais importantes ocorridos durante projeto, as conclusões referentes aos resultados obtidos e algumas sugestões futuras.

5.1 Síntese do Projeto Realizado

O principal objetivo desta dissertação era averiguar o elevado *churn* do serviço DTH comparando o com o serviço WoW, uma vez que são dois serviços bastante idênticos, diferindo apenas na inclusão do pacote TV no serviço DTH. Existe uma grande disponibilidade de dados e recursos computacionais com capacidade para processar esses mesmo dados, possibilitando extrair conhecimento de dados retirados nas diversas fontes existentes na rede utilizando *software* de *data mining*. Extrair conhecimentos é uma ideia muito geral em que, a extração de conhecimento dos dados pode ser a mais variada possível.

A partir do momento que se conseguiu reunir todos os dados disponíveis pelas diferentes fontes, é necessário explorar e estruturar os dados para que, através deste estudo, permitisse a tal extração de conhecimento para responder à questão que conduziu esta dissertação: “Qual a razão do elevado *churn* do serviço DTH?”

Ao longo deste projeto procedeu-se ao desenvolvimento de um estudo sobre a qualidade da Internet fixa na rede móvel que será útil no contexto empresarial, visando a

experiência de utilização dos clientes que contratam este tipo de serviço. O estudo permite ter uma ideia geral das velocidades obtidas em detrimento de diversas variantes, desde a tecnologia que os testes foram efetuados (3G e 4G); o horário, em horas de maior congestão ou horas com menor afluência de tráfego na rede; a taxa de ocupação das células; a qualidade do sinal, entre muitas outras vistas na secção 4.2.

O grande contributo desta dissertação deve-se com a novidade deste estudo proposto pela empresa de telecomunicações NOS, uma vez que, nunca antes tinha sido feito qualquer estudo referente ao tema desta dissertação. Este estudo será importante para ter uma ideia de como a qualidade da internet fixa via rede móvel varia, porque permitiu encontrar uma panóplia de dimensões que, de uma maneira, direta ou indireta, têm influência para qualidade do serviço DTH, especificamente, a internet fixa.

Tendo sido efetuado no departamento de Operações da NOS, estes resultados podem e devem ser úteis para as demais áreas dentro da empresa. É bastante útil para questões de melhoramento de qualidade de serviço, o que inclui um conjunto de diversas áreas, como áreas de engenharia (rádio, redes, etc.), e também para questões de marketing e campanhas, uma vez que se trata de um serviço que inclui pacote de TV e Internet fixa com velocidades de 20 Mbps e 40 Mbps, dependendo do que os clientes estiverem dispostos a contratar e da sua necessidade.

Relativamente a problemas encontrados, começa-se por referir o primeiro impacto empresarial. Sendo que se trata de um estudo de uma das principais operadoras de telecomunicações em Portugal foi necessário mudar o estigma académico e conseguir apontar para estudos mais analistas que são essenciais para o crescimento empresarial.

Segunda dificuldade foi definir as dimensões de análise. Isto é, tendo noção da grande quantidade de dados produzidos por todas as fontes consideradas para este projeto, foi importante conseguir definir as principais características dos dados para criar conjuntos de dados coesos que possam garantir extrair informação útil.

Uma das dificuldades que mais demorou a ultrapassar terá sido o software R. Visto ser uma ferramenta com funcionalidades matemáticas requer alguma prática e como tal, retirou algum tempo neste projeto.

Contextualizado com o resultado final, o autor considera que a solução encontrada é uma grande mais-valia para a empresa pois permite que seja um objeto de estudo para futuros projetos que envolvam este tipo de serviço, ou para qualquer outra área na empresa, desde marketing a engenharia.

5.2 Sugestões para Trabalho Futuro

A ideia por detrás deste projeto foi construir um programa que possibilitasse a análise os dados retirados e tirar conclusões e respetivas análises dos dados em questão. Portanto, sugere-se que, no futuro se construa um programa que percorra, automaticamente, as bases de dados e que consiga tirar alguma informação (*machine learning*). Também era útil um programa que se focasse à procura de padrões que possam ser úteis, e que provavelmente, fossem generalizados para fazer previsões precisas sobre dados futuras, tendo em conta o objetivo que se pretende. Neste caso, era prever falhas das comunicações móveis para ter algumas garantias de qualidade de serviço, nas mais diversas zonas do País.

Podia também ser interessante conseguir associar estes dados com fontes que nos forneçam mais feedback por parte do cliente. Ou seja, que se consiga igualar a quantidade de dados feitos no ato da instalação e dados produzidos diretamente pelo utilizador (*speedtests* efetuados por ordem do utilizador).

Referências

- [1] C. Andersson and P. Svensson, “Mobile Internet - An industry-wide paradigm shift?,” *Ericsson Rev. (English Ed.)*, vol. 76, no. 4, pp. 206–213, 1999.
- [2] C. Keramane, “The wireless world web,” *IEEE Multimed.*, vol. 7, no. 2, pp. 10–14, 2000.
- [3] K. L. Dias and D. F. H. Sadok, “Internet Móvel: Tecnologias, Aplicações e QoS,” *XIX Simpósio Bras. Redes Comput.*, p. 51, 2001.
- [4] M. K. I. Bilal, Haider, M. K. Islam, “Radio Frequency Optimization & QoS Evaluation in Operational GSM Network,” *World Congr. Eng. Comput. Sci.*, vol. I, pp. 2–5, 2009.
- [5] E. Dahlman and J. Sko, *4G LTE/LTE-Advance for Mobile Broadband, Second Edition*. 2014.
- [6] E. Dahlman, S. Parkvall, and J. Sko, *The Road to 5G Third Edition*. 2016.
- [7] T. Halonen, J. Romero, and J. Melero, *GSM, GPRS and EDGE Performance: Evolution Towards 3G/UMTS*. 2003.
- [8] A. Mehrotra, *GSM systems engineering*. 1997.
- [9] O. Herzog and M. Lawo, *Future Mobile Communications: LTE Optimization and Mobile Network Virtualization*. 2012.
- [10] “<https://en.wikipedia.org/wiki/GSM>.” [Online]. Available: <https://en.wikipedia.org/wiki/GSM>. [Accessed: 15-Aug-2017]
- [11] R. No, T. D. Compagnia, P. Italiana, S. Rapporto, and M. I. Roma, “METHOD OF EXCHANGING DATA BETWEEN A BASE TRANSCIEVER STATION OF A MOBILE RADIO NETWORK AND A MOBILE IN THE NETWORK,” vol. 92, no. 2, pp. 141–151, 1996.
- [12] “<https://en.wikipedia.org/wiki/UMTS>.” [Online]. Available: <https://en.wikipedia.org/wiki/UMTS>. [Accessed: 15-Aug-2017]
- [13] X. Li, Y. Zaki, T. Weerawardane, A. Timm-Giel, and C. Goerg, “HSUPA backhaul

- bandwidth dimensioning,” in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC*, 2008.
- [14] T. Weerawardane, Y. Zaki, A. Timm-Giel, G. C. Malafrente, S. Hauth, and C. Görg, “Impact of the transport network congestion control on the HSUPA performance,” *IEEE Veh. Technol. Conf.*, 2009.
- [15] X. Li, T. Weerawardane, Y. N. Zaki, C. Görg, and A. Timm-Giel, “Shared Transport for Different Radio Broadband Mobile Technologies (Chapter 9),” *Recent Adv. Broadband Integr. Netw. Oper. Serv. Manag.*, pp. 135–139, 2011.
- [16] X. Li, Y. N. Zaki, T. L. Weerawardane, A. Timm-Giel, C. Görg, and G. C. Malafrente, “Use of Traffic Separation Techniques for the Transport of HSPA and R99 Traffic in the Radio Access Network with Differentiated Quality of Service (Chapter 56),” *Netw. Telecommun. Concepts, Methodol. Tools, Appl.*, vol. 5, no. 2, pp. 863–878, 2010.
- [17] H. Holma and A. Toskala, *LTE for UMTS - OFDMA and SC-FDMA based radio access*. 2009.
- [18] S. Sesia, I. Toufik, and M. Baker, *LTE – The UMTS Long Term Evolution from Theory to Practice*. 2009.
- [19] F. Sandu, S. Cserey, and E. Mile-Ciobanu, “Simulation of LTE signaling,” *Adv. Electr. Comput. Eng.*, vol. 10, no. 2, pp. 108–114, 2010.
- [20] Tekelec, “Monitoring LTE Networks - A new paradigm for monitoring mobile broadband networks and services,” 2010.
- [21] 3GPP, “3GPP TS 36.300: Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN),” 2011. [Online]. Available: <http://www.3gpp.org/dynareport/36300.htm>.
- [22] W. Fan and A. Bifet, “Mining Big Data : Current Status , and Forecast to the Future,” *ACM SIGKDD Explor. Newsl.*, vol. 14, no. 2, pp. 1–5, 2013.
- [23] A. Rajaraman and J. D. Ullman, “Mining of Massive Datasets,” *Lect. Notes Stanford CS345A Web Min.*, vol. 67, p. 328, 2011.
- [24] X. Wu, X. Zhu, G.-Q. Wu, and W. Ding, “Data Mining with Big Data,” *Knowl. Data Eng. IEEE Trans.*, vol. 26, no. 1, pp. 97–107, 2014.
- [25] D. Laney, “META Delta,” *Appl. Deliv. Strateg.*, vol. 949, no. February 2001, p. 4, 2001.
- [26] I. Udacity, “Lesson 1 Notes,” 2014.
- [27] I. H. Witten, E. Frank, and M. a. Hall, *Data Mining: Practical machine learning tools and techniques*. 2011.
- [28] F. Gorunescu, *Data Mining: Concepts and Techniques*, vol. 12. 2011.
- [29] J. Han, “Data Mining: Concepts and Techniques,” *Univ. Illinois Urbana-Champaign*, pp. 1–136, 2007.
- [30] A. Wilson and N. Norden, “The R Project for Statistical Computing The R Project for Statistical Computing,” vol. 3. 2015.

- [31] R Development Core Team, “R Language Definition,” *Web*, vol. 0, p. 62, 2011.
- [32] “R: The R Project for Statistical Computing.” [Online]. Available: <https://www.r-project.org/>. [Accessed: 28-Aug-2017].
- [33] “How to calculate LTE throughput.” [Online]. Available: http://anisimoff.org/eng/lte_throughput.html. [Accessed: 08-Sep-2017].

