CZECH TECHNICAL UNIVERSITY IN PRAGUE

Faculty of Electrical Engineering
Department of Telecommunication Engineering

Master Thesis

Data analysis from the mobile network

Bc. David Blagodárný

Study Programme:	Communication, Multimedia and Electronics
Branch of study:	Electronic Communication Networks

Thesis adviser: Ing. Robert Bešťák, Ph.D. Prague, May 2018

Čestné prohlášení

Prohlašuji, že jsem zadanou diplomovou práci zpracoval sám s přispěním vedoucího práce a konzultanta a používal jsem pouze literaturu v práci uvedenou. Dále prohlašuji, že nemám námitek proti půjčování nebo zveřejňování mé diplomové práce nebo její části se souhlasem katedry.

Datum: 25. 5. 2018

....

ČVUT
ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

ZADÁNÍ DIPLOMOVÉ PRÁCE

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Blagodárný** Jméno: **David** Osobní číslo: **406073**

Fakulta/ústav: **Fakulta elektrotechnická**

Zadávající katedra/ústav: **Katedra telekomunikační techniky**

Studijní program: **Komunikace, multimédia a elektronika**

Studijní obor: **Sítě elektronických komunikací**

II. ÚDAJE K DIPLOMOVÉ PRÁCI

Název diplomové práce:

Analýza dat z mobilní sítě

Název diplomové práce anglicky:

Data analysis from the mobile network

Pokyny pro vypracování:

Navrhněte vhodný formát vstupních dat získaných z páteřní sítě mobilního operátora pro účely jejich dalšího zpracování a jejich následnou analýzu zaměřenou na detekci anomálii v síti s využitím nástrojů pro zpracování velkých dat (např. Hadoop).

Seznam doporučené literatury:

[1] Walke, B. H.: Mobile Radio Networks, Networking and Protocols. John Wiley & Sons, Stuttgart 1999. ISBN: 0-471-97595-8.
[2] EMC Education Services: Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data. Wiley, 1st Edition, January 2015. ISBN: 1-118-87613-X.
[3] Baesens, B.: Analytics in a Big Data World: The Essential Guide to Data Science and its Applications. Wiley, 1st Edition. April 2014. ASIN: B00JR5LAC6.

Jméno a pracoviště vedoucí(ho) diplomové práce:

Ing. Robert Bešťák, Ph.D., katedra telekomunikační techniky FEL

Jméno a pracoviště druhé(ho) vedoucí(ho) nebo konzultanta(ky) diplomové práce:

Datum zadání diplomové práce: **04.01.2018** Termín odevzdání diplomové práce: **25.05.2018**

Platnost zadání diplomové práce: **30.09.2019**

Ing. Robert Bešťák, Ph.D.
podpis vedoucí(ho) práce

podpis vedoucí(ho) ústavu/katedry

prof. Ing. Pavel Ripka, CSc.
podpis děkana(ky)

III. PŘEVZETÍ ZADÁNÍ

Diplomant bere na vědomí, že je povinen vypracovat diplomovou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací. Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v diplomové práci.

.
Datum převzetí zadání

Podpis studenta

Anotace:

Tato diplomová práce se zabývá analýzou dat získaných z páteřní sítě mobilního operátora, konkrétně z části zabezpečující síťová pravidla a účtování. Cílem práce je vypracování postupů a následné analýzy k určení problémů v implementované technologii VoLTE. Analýza se zaměřuje na základní ukazatele v síti, množství přerušených spojení a hlášených chyb protokolu DIAMETER. Je navržen způsob získání vybraných dat pomocí nástrojů pro velká data, konkrétně Hadoop a Spark. Uvedené nástroje jsou následně použity pro analýzu dat. Analyzovaná data jsou mimo jiné vizualizována v mapě pomocí vytvořené webové aplikace. Dále je navrženo řešení využívající výkonnostní indikátor pro odhalení anomálií v síti. Toto řešení je na datech vyzkoušeno a popsáno.

Klíčová slova: LTE, VoLTE, Analysis, Anomaly Detection, Hadoop, Big Data, Spark

Summary:

This diploma thesis deals with the analysis of data obtained from the mobile network operator, specifically from the Policy and Charging Control subsystem. The aim is to develop procedures and subsequent analyses to identify problems in implemented VoLTE technology. The analysis focuses on the basic network indicators, the number of aborted sessions, and reported DIAMETER errors. A method is proposed, how to obtain the selected data with tools for Big Data, specifically Hadoop and Spark. These tools are then used for data analysis. Analyzed data is visualized on the map using the created web application. Furthermore, a solution using a performance indicator for anomaly detection is proposed. This solution is further tested and described using the provided data.

Index Terms: LTE, VoLTE, Analysis, Anomaly Detection, Hadoop, Big Data, Spark

Poděkování

V první řadě bych chtěl poděkoval vedoucímu mé diplomové práce Ing. Robertu Bešťákovi, Ph.D. za poskytnutí věcných připomínek k mé práci včetně cenných rad, jak práci vylepšit. Velmi si cením času, který mi věnoval během konzultací. Rovněž patří můj dík rodině za podporu při studiu a tvorbě této práce.

Acknowledgement

First and foremost, I would like to thank my diploma thesis supervisor Ing. Robert Bešťák, Ph.D. for giving me factual comments of my work and precious advice how to improve it. I sincerely appreciate the time he took to give me consultation. Furthermore, I would like to pass on my sincere thanks and gratitude to my family for giving me support during my studies and creation of this thesis.

Contents

List of Figures

List of Tables

List of Abbreviations

3GPP	**3**rd **G**eneration **P**artnership **P**roject
AF	**A**pplication **F**unction
AS	**A**pplication **S**erver
AVP	**A**ttribute-**V**alue **P**air
BTS	**B**ase **T**ransceiver **S**tation
CLI	**C**ommand **L**ine **I**nterface
CS	**C**ircuit **S**witched
CSCF	**C**all **S**ession **C**ontrol **F**unction
E-UTRAN	**E**volved **U**niversal **T**errestrial **RAN**
eNB	**e**volved **N**ode **B**
EPC	**E**volved **P**acket **C**ore
EPS	**E**volved **P**acket **S**ystem
GBR	**G**uaranteed **B**it **R**ate
GIS	**G**eographic **I**nformation **S**ystem
GSM	**G**lobal **S**ystem for **M**obile Communications
HDFS	**H**adoop **D**istributed **F**ile **S**ystem
HSPA	**H**igh **S**peed **P**acket **A**ccess
HSS	**H**ome **S**ubscriber **S**erver
ICSCF	**I**nterrogating **CSCF**
IM-MGW	**IM**S **M**edia **G**ate**W**ay
IMEI	**I**nternational **M**obile **E**quipment **I**dentity
IMS	**I**P **M**ultimedia **S**ubsystem
IMSI	**I**nternational **M**obile **S**ubscriber **I**dentity
IP-CAN	**IP** **C**onnectivity **A**ccess **N**etwork
KPI	**K**ey **P**erformance **I**ndex
LAU	**L**ocal **A**dministrative **U**nit
LTE	**L**ong **T**erm **E**volution
MGCF	**M**edia **G**ateway **C**ontrol **F**unction
MME	**M**obility **M**anagement **E**ntity
MS	**M**obile **S**tation
NUTS	**N**omenclature of **T**erritorial **U**nits for **S**tatistics
OCF	**O**nline **C**harging **F**unction
PCC	**P**olicy and **C**harging **C**ontrol
PCEF	**P**olicy and **C**harging **E**nforcement **F**unction
PCRF	**P**olicy and **C**harging **R**ules **F**unction
PCSCF	**P**roxy **CSCF**
PDN	**P**acket **D**ata **N**etwork
PLMN	**P**ublic **L**and **M**obile **N**etwork
PS	**P**acket **S**witched
PSTN	**P**ublic **S**witched **T**elephone **N**etwork

QCI	**Q**oS **C**lass **I**dentifier
QoS	**Q**uality **o**f **S**ervice
RAN	**R**adio **A**ccess **N**etwork
RDBMS	**R**elation **D**ata**B**ase **M**anagement **S**ystem
RDD	**R**esilent **D**istributed **D**ataset
RFC	**R**equest **F**or **C**omments
RTP	**R**eal-time **T**ransport **P**rotocol
S-GW	**S**erving **G**ate**W**ay
SCSCF	**S**erving **CSCF**
SDF	**S**ervice **D**ata **F**low
SDP	**S**ession **D**escription **P**rotocol
SIP	**S**ession **I**nitiation **P**rotocol
SPR	**S**ubscriber and **P**rofile **R**epository
TA	**T**racking **A**rea
UDR	**U**nified **D**ata **R**epository
UE	**U**ser **E**quipment
UMTS	**U**niversal **M**obile **T**elecommunication **S**ystem
VoLTE	**V**oice **o**ver **LTE**
VoIP	**V**oice **o**ver **IP**
WCDMA	**W**ideband **C**ode **D**ivision **M**ultiple **A**ccess
YARN	**Y**et **A**nother **R**esource **N**egotiator

CHAPTER 1

Introduction

Cellular mobile networks are undergoing a rapid development every year and the networks have come a long way in the last twenty-five years from circuit-switched Global System for Mobile Communications (GSM) to packet-switched Long-Term Evolution (LTE) or awaited 5G networks. However, there is one service that relies on GSM, the voice service. Still, most of the mobile networks worldwide use GSM to provide voice services.

Motivation

There is a need for convergence of mobile networks because the demand for mobile services is increasing and the number of frequency bands is limited. The only way to increase the throughput is to use the radio spectrum efficiently, thus using newer and more efficient technologies. The Voice over LTE (VoLTE) seems to be the most common approach for voice service in packet-switched networks. Unlike voice service in GSM, VoLTE is not clearly specified, enabling various implementations of it which pose a problem of interoperability.

This work focuses on analyses of various issues in VoLTE and finding possible reasons for it. The analyses are accomplished using Big Data approach, analyzing message flow in Policy and Charging Control (PCC) subsystem, specifically Rx and Gx interface. Anomaly detection solution using performance indicator is proposed. This proposed performance indicator can identify cells, eventually devices in the network experiencing problems with VoLTE.

Structure

This paper is divided into seven chapters as follows:

Chapter 1 *(Introduction)* introduces the purpose of this work and briefly describes the content of each subsequent chapter.

Chapter 2 *(Introduction to the mobile networks, LTE and VoLTE)* describes the basic architecture of LTE with EPS bearers. Then the Policy and Charging Control subsystem in LTE is introduced. At the end of chapter, VoLTE and IP Multimedia Subsystem (IMS) are described.

Chapter 3 *(Description of VoLTE procedures)* is dealing with message flows in PCC subsystem related to VoLTE. Examples of various message flows are described, including EPS bearer and VoLTE session establishment.

Chapter 4 *(Hadoop ecosystem)* is introducing Big Data approach for undertaking the analyses. Used frameworks and tools are described with reasons why they were used.

Chapter 5 *(VoLTE analyses)* contains results of the analyses. The first type of analyses is more general whereas the second type is dealing with CS fallbacks and third with DIAMETER errors.

Chapter 6 *(Anomaly detection)* explains what is meant by anomaly and introduces proposed performance indicator for anomaly detection. Results from selected cells and devices are later discussed within this chapter.

Chapter 7 *(Conclusion)* is the last chapter and concludes this work with achieved results and possible continuation.

CHAPTER 2

Introduction to the mobile networks, LTE and VoLTE

A mobile phone network is a cellular network used for transmitting voice, data and other services. The official term for this network is a Public Land Mobile Network (PLMN). This network is created and maintained by mobile network operators, worldwide or local companies such as Vodafone or Orange. The mobile network comprises three main components, the core network, the radio access network and the mobile device.

The core network is Circuit Switched (CS) or Packet Switched (PS) but can contain both domains. The CS domain is using a similar approach to traditional telecommunication systems with fixed lines. The network creates logical circuit with allocated resources for each phone call. Each call is then connected to the Public Switched Telephone Network (PSTN) to enable communication with subscribers on landlines or CS networks belonging to other operators. A PS domain network transports encapsulated data streams from the subscriber to external Packet Data Networks (PDN) such as internet. These encapsulated data streams can be web pages, emails or other services which are being sent over the IP networks.

The radio access network implements various radio access technologies and resides between the core network and the mobile device, interconnecting each other. Therefore, it secures radio communication of the core network with the subscriber. The mobile device is a subscriber's device such as a smartphone or laptop with broadband capabilities and communicates with the radio access network over the air interface. It is known as a User Equipment (UE) or a Mobile Station (MS) depending on the generation of the used mobile network.

Long-Term Evolution (LTE) is a standard for cellular wireless communication for mobile devices designed by the Third Generation Partnership Project (3GPP). The main task was to surpass and later replace earlier 3GPP systems, the Universal Mobile Telecommunication System (UMTS) and its predecessor, the Global System for Mobile Communications (GSM). The aim was to improve performance and capability of networks such as spectral efficiency and latency. LTE is also known as 4th generation network, UMTS as 3rd generation and GSM as 2nd.

Nowadays, the number of subscriptions of LTE networks is increasing at the expense of other technologies and the trend of replacing its predecessors such as 2nd generation (GSM, CDMA) and 3rd generation (WCDMA/HSPA) networks is evident, as shown in figure 2.1. This chart contains the number of subscriptions of each cellular mobile network from the last eight years. LTE is the world's most used cellular mobile communication technology, and it is likely to be dominant for some years to come.

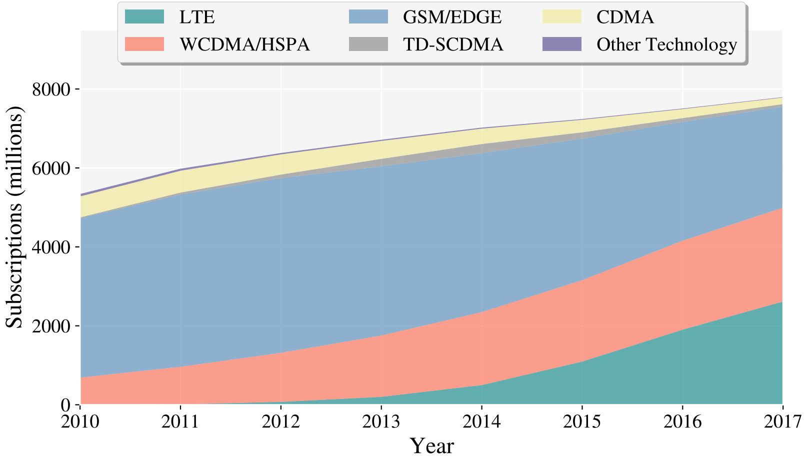


Figure 2.1: Number of subscriptions per mobile network technology [1].

2.1 Basic architecture of LTE

LTE comprises three basic components already introduced in the previous section: core network, radio access network and UE. The basic architecture of LTE is shown in figure 2.2. The radio access network in LTE is called Evolved Universal Terrestrial Radio Access Network (E-UTRAN). The only mandatory node of it is the E-UTRAN Node B, in an abbreviated form also known as eNB. It is equivalent to Base Transceiver Station (BTS) communicating with UE using the air interface in previous generations of mobile networks. Unlike BTS, eNB has extended functionality and performs many more tasks by itself, such as handover. Multiple eNBs can even directly communicate with each other without the need of network controller. These facts contribute to lower response time in the network. For more information about the radio access network and the UE see [2, 3].

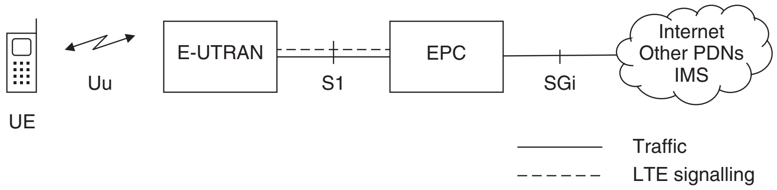


Figure 2.2: Basic architecture of LTE [2].

The core network in LTE is called Evolved Packet Core (EPC). Whereas previous generations of mobile networks were using both packet switching and circuit switching, LTE completely omitted circuit switching and relies entirely on packet switching. The reason behind this decision is to make the core network more simple and robust, operating only as an imaginary pipe for any payload between UE and PDN such as internet. All additional services including voice service are then left to different components outside of the EPC. The EPC comprises multiple components which are shown in figure 2.3.

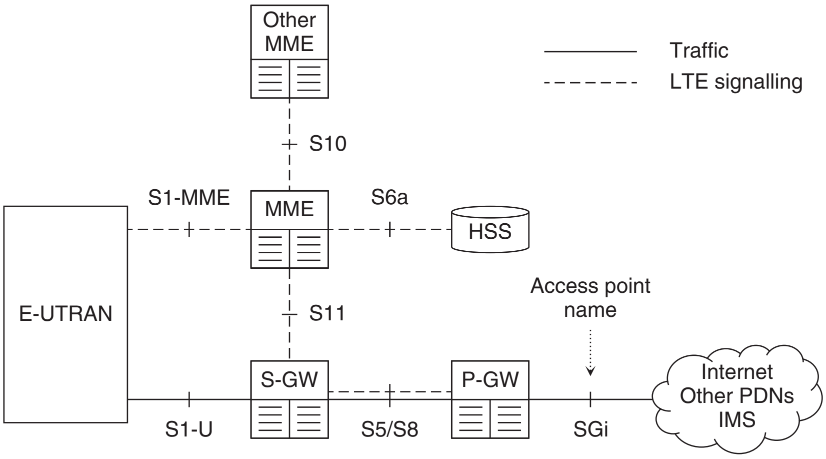


Figure 2.3: Main components of EPC [2].

The Home Subscriber Server (HSS) is a central database containing information about all the network subscribers. HSS is one of the few components that are also present in previous generations of mobile networks. The Mobility Management Entity (MME) is the main signaling node in EPC. MME is responsible for authentication of UE, retaining its location information and selecting an appropriate gateway for each subscriber. MME plays a vital role in handover procedure between LTE and previous generation networks. Additional MMEs can be created to handle a higher load in the network, typically for different geographical regions.

The serving gateway (S-GW) acts as a high-level router, forwarding data between eNB and PDN gateway. The typical mobile network contains multiple S-GWs and a single S-GW is usually assigned to a certain geographical region. Each UE can be handled by only one S-GW at the time but can be changed if the subscriber moves sufficiently far. The PDN gateway (P-GW) is the point of contact with the outside world. Each P-GW exchange data with external PDNs such as IP Multimedia Subsystem (IMS) and internet. Default P-GW is assigned to every UE after it registers to the network, giving it always-on connectivity to default PDN. Additional P-GWs are assigned when UE requires connections to another PDN such as IMS.

2.2 EPS Bearer

The mobile networks must address two major issues when establishing the connection to PDN. First, the mobility of subscribers, thus the ability to maintain the connection when the subscriber moves from one location to another. Then the quality of service (QoS), a term describing parameters of data streams, such as Guaranteed Bit Rate (GBR), maximum error rate, and maximum delay. To address these issues, data from one part of the LTE system (UE) to another (P-GW) are transported using EPS bearers. They are virtual bi-directional data pipes that transfer data with pre-defined QoS.

As already mentioned, a default EPS bearer is established when the UE is switched on and registered to the mobile network and remains active until UE is detached from the network. A default bearer comes with an IP address and provides always-on IP connectivity. UE can also receive additional bearers known as dedicated bearers enabling more specific traffic, such as voice or video. The dedicated bearer is established on top of default bearer and shares the same IP address but enables required QoS, such as GBR. Dedicated bearers are created when the requested service requires it and canceled when no longer needed by the service.

LTE networks with the voice over LTE (VoLTE) implementations have usually two default bearers and one dedicated bearer. The first default bearer is dedicated to signaling messages related to IMS network enabling Session Initiation Protocol (SIP) signaling. In the event of a VoLTE call, a dedicated bearer with required QoS is created and at the end of the call, this bearer is cancelled. Second default bearer is dedicated for all other traffic, such as web browsing, email or chat. Two default bearers are needed because IMS network is a separate network from the internet.

2.2.1 QoS in LTE

The most important parameter defining IP level characteristic in LTE is the QoS Class Identifier (QCI) [4]. The QCI is a single number specifying four other characteristics. The first is resource type defining whether GBR is required or not. Bearers with GBR are suitable for real-time services such as voice, in which case GBR value corresponds to the lowest bit rate of the voice codec. A default bearer is always a non-GBR bearer. The next characteristics are priority, packet delay, and packet error loss rate. Important QCIs with their characteristics are shown in table 2.1.

Table 2.1: The example of QCI characteristics [4].

QCI	Resource Type	Priority Level	Packet Delay Budget	Packet Error Loss Rate	Example Services
1	GBR	2	100 ms	10^{-2}	Conversational Voice
2	GBR	4	150 ms	10^{-3}	Conversational Video (Live Streaming)
3	GBR	3	50 ms	10^{-3}	Real Time Gaming
4	GBR	5	300 ms	10^{-6}	Non-Conversational Video (Buffered Streaming)
5	Non-GBR	1	100 ms	10^{-6}	IMS Signalling
6	Non-GBR	6	300 ms	10^{-6}	Video (Buffered Streaming), TCP-based (e.g., www, e-mail, chat)
7	Non-GBR	7	100 ms	10^{-3}	Voice, Video (Live Streaming), Interactive Gaming

2.3 Policy and Charging Control

LTE is purely packet-switched network and mobile network operator must have an option how to enforce charging and various policies for offered services. The task of the Policy and Charging Control (PCC) is to enable adding and re-configuring policies to dynamically manage and control QoS and to apply operator-predefined charging rules to each Service Data Flow (SDF). SDF is a term for multiple data streams with the same QoS in a single EPS bearer, and it is associated with a single PCC rule [2]. PCC in LTE consists of two main elements, the Policy and Charging Rules Function (PCRF) and the Policy and Charging Enforcement Function (PCEF). The basic PCC architecture is shown in figure 2.4.

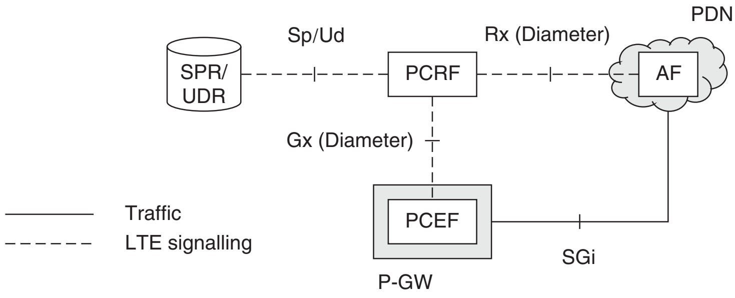


Figure 2.4: Basic policy and charging control architecture [2].

The main element of PCC is the PCRF, authorizing that each SDF in the network will be treated accordingly to its PCC rule. The PCEF is included in the P-GW and its role is to implement decisions from PCRF for each SDF. There are two types of PCC rules in LTE networks: pre-defined and dynamic rules. Pre-defined rules are already pre-configured in the PCEF and dynamical rules are dynamically provisioned by PCRF at the time of request. The Application Function (AF) shown in figure 2.4 provides environment for running an external application such as IMS in case of VoLTE implementation or video streaming server. AF may be controlled by another party [4]. PCRF uses DIAMETER protocol for most of its signalling interfaces, such as Gx [5] and Rx [6], as shown in figure 2.4. LTE networks can even operate without the use of the PCRF, with only a single pre-defined PCC rule used for all subscribers and networks. In this case, there is no control over QoS.

The Subscriber Profile Repository (SPR) is the database keeping the subscribers' policies and charging profiles, for the QoS management in PCRF. The Unified Data Repository (UDR) in LTE is the back-end database keeping all the information, whereas HSS and SPR are front-end databases for the access to the UDR. Some vendors include all of these databases into a single database [2].

2.4 Voice Calls over LTE

Voice applications are not an integral part of LTE because EPC is designed to only transport data to, and from the subscriber and has no concern with the content of data. Voice calls are very important part of telecommunication networks, therefore there are two main techniques how to enable them. The first option is called Circuit Switch Fallback (CSFB) and the latter option is by using an external network that includes the required signalling functions such as IMS. These approaches can be combined which is the case of most networks today and it will be further described in the later text. Additionally, there are other approaches, such as dual radio devices [7] or VoLGA [8], but these options are not further developed in this text. Instead of the abbreviation CSFB, the term CS fallback will be used to ensure good readability through the text.

The CS fallback approach is commonly used technique by early rollouts of LTE because of the use of already built legacy mobile networks. When a voice call is required, the mobile network transfers UE from LTE cell to a legacy 2G/3G cell and the voice call is done traditionally in the CS domain of these networks. The advantage of this approach is in using of already working solution, therefore most mobile operators adopted this technique in their early LTE rollouts. However, there are multiple disadvantages of this approach such as the need to maintain legacy networks, delay in the setup of a voice call and many other complications and restrictions [2]. Thus, a reason why operators should move on to the next approach.

Second approach is using an external network that includes the required signalling functions and take care of the call procedures. The simplest option is to offer a Voice over IP (VoIP) service through a third-party suppliers such as Skype and Viber. However, this technique is not advisable because all of the signalling messages are for the EPC indistinguishable from other data. Main advantage is in low deployment costs for the operators, nevertheless the network operator no longer owns and controls the voice service. Other options include proprietary VoIP systems developed by operators which are problematic in interconnection with other networks. Therefore, a technique of using partially standardized IMS as an external network for voice calls became the most used approach for enabling voice calls in LTE networks [9]. This method is commonly known as Voice over LTE (VoLTE).

2.5 IP Multimedia Subsystem

The IMS is not an integral part of LTE network but it is a separate PDN such as internet. However, this network is crucial for making voice calls in LTE networks because LTE network itself is not capable of it. The IMS have a signalling functionality that manage VoIP calls. The used signalling protocol is SIP. IMS shares similarities with third-party VoIP servers, unlike them, it belongs to the network operator and has enhanced functionality with support of QoS and emergency calls which are not implemented in third-party VoIP servers but are required by the regulators in mobile network. The most important components are shown in figure 2.5.

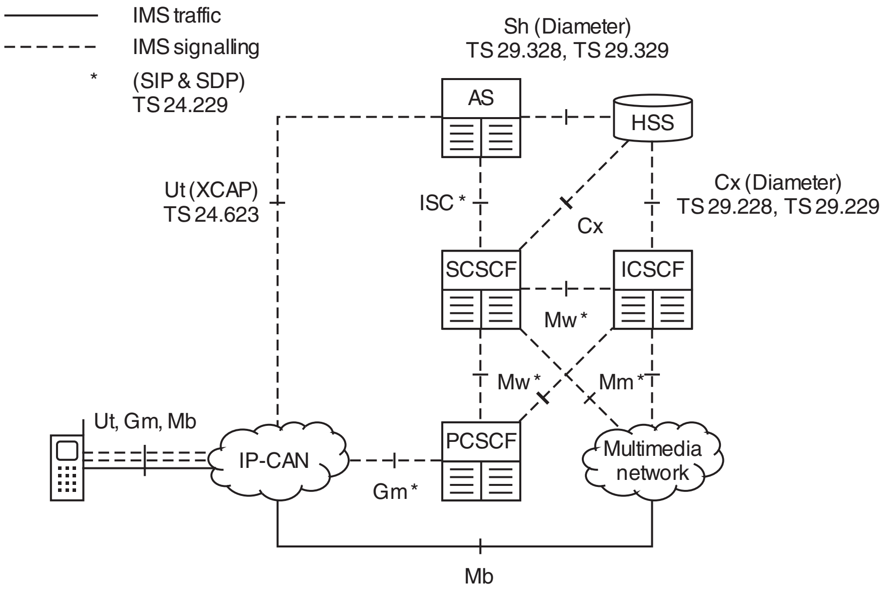


Figure 2.5: Basic components of IMS [2].

The IMS can be reached through an IP Connectivity Access Network (IP-CAN), which can be LTE network, wireless LAN or even 3G network. IMS is independent of the network if it complies with IP network standards. Basically, IMS can communicate with any IP network, even another IMS. The IMS is mainly concerned with control plane communication using SIP and Session Description Protocol (SDP), whereas user plane represented by Real-time Transport Protocol (RTP) media stream mostly bypasses the IMS. The basic components of any IMS are known as Call Session Control Functions (CSCFs). There are three types of CSCF: Serving CSCF (SCSCF), Proxy CSCF (PCSCF) and Interrogating CSCF (ICSCF) [10].

The PCSCF is the first contact point within the IMS and behaves like a normal proxy, accepting requests and forwarding them on to ensure that SIP messages contain correct and required information.The PCSCF is connected with the PCRF via standardized Rx interface, acting from outside as an IMS application function. The SIP messages are forwarded to the SCSCF. The SCSCF performs the session control services, therefore acts as a SIP server. The ICSCF is the first point of contact for the messages arriving from other IP networks such as another IMS. The Application Server (AS) supplies additional services including voicemail and SMS. The HSS in IMS is a database containing subscribers' profiles and can be either same or different than HSS in operator's EPC. Most of the interfaces between components within the IMS are included in figure 2.5 with information about associated technical specifications from 3GPP [2].

IMS can communicate not only with another IP networks but also with a CS domain in 2G/3G networks or the PSTN using the IMS Media Gateway (IM-MGW). The IM-MGW is a user plane interface and utilize transcoding and media streaming functions such as echo cancellation or tone generator. The Media Gateway Control Function (MGCF) is a control plane interface that controls and translates signalling messages from CS domain to the IM-MGW. All of these components are shown in figure 2.6.

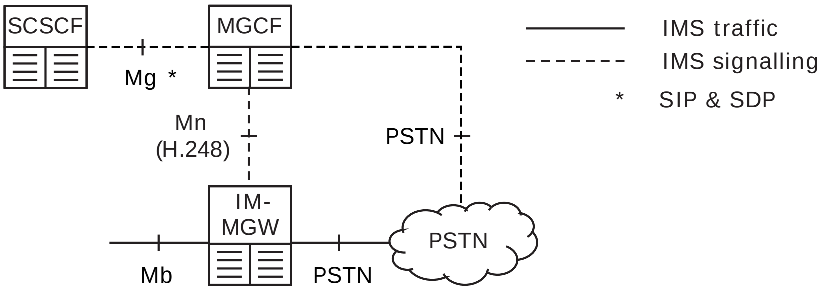


Figure 2.6: Media gateway function of IMS [2].

CHAPTER 3

Description of VoLTE procedures

The scope of this text is to analyze the data that are generated by PCC subsystem inside of the EPC. Therefore, only procedures taking place in this subsystem are described in this chapter. More information about all the other procedures required for establishing and maintaining VoLTE voice call are comprehensively described in [2, 3, 7, 11].

More precisely, only two interfaces of PCRF subsystem are analyzed and described in this chapter: Rx and Gx interface. The Gx interface is located between PCEF and PCRF and its role is to enable control over allocation of network resources for the UE. The Rx interface is located between AF and PCRF delegating communication with SIP server inside of IMS. Both interfaces use DIAMETER protocol. RX is specified in [6] and Gx interface in [5]. Both of these interfaces were described in the previous chapter.

The PCEF is part of the PGW, therefore instead of the PCEF, the term PGW will be used in this text. It is similar with AF, which is basically SIP server and it is a functionality of S-CSCF in the IMS and that is the reason why the term IMS will be used instead of AF. These terms are replaced to more general terms because this text is dealing only with LTE mobile networks and it will make this problematic easier for understanding.

First part of this chapter is dealing with an explanation of specific DIAMETER messages and their Attribute-Value Pairs (AVPs). Description of each type of used DIAMETER message is included, and it is explained when and why they occur. The following part contains DIAMETER message flows during various events such as subscriber registration, regular call or interrupted call. The last part of this chapter is dealing with DIAMETER errors, such as what are the errors, when and where the errors usually occur and what can be their cause.

The SIP message control plane will be completely omitted from this text because it is not part of the PCC subsystem of the LTE network. The PCRF has no information about ongoing SIP message flow and the SIP server must initiate DIAMETER command to carry out EPS bearer change. More information about utilization of the SIP and SIP message flow associated with voice calls in VoLTE can be found in [11].

3.1 DIAMETER messages used in VoLTE

DIAMETER is an AAA (Authentication, Authorization, Accounting) protocol facilitating authentication of subscribers and authorization of their access to services, such as VoLTE. It is also used for collecting of information on resource usage, which are necessary for charging of subscribers. It is defined in Request For Comments (RFC) 7155 [12].

DIAMETER messages comprise multiple AVPs and are divided by command-code AVP based on their role. The DIAMETER message can be also called a command. Each command starts with a request and then requires an answer in return. This request/answer pair always share the same command code. Commands with different command-code AVP usually contains different AVPs, but it is not a necessary condition. For example, the session-name AVP is included in each command identifying it to the session. DIAMETER command codes used on Rx and Gx interface in VoLTE are listed in table 3.1 below. Re-Auth-Request/ Re-Auth-Answer command pair can be also implemented on Rx interface, but it is not required for VoLTE.

Table 3.1: Command codes used on Rx and Gx interface in VoLTE.

Command Name	Abbreviation	Code	Interface
Re-Auth-Request	RAR	258	Gx*
Re-Auth-Answer	RAA	258	Gx*
AA-Request	AAR	265	Rx
AA-Answer	AAA	265	Rx
Credit-Control-Request	CCR	272	Gx
Credit-Control-Answer	CCA	272	Gx
Abort-Session-Request	ASR	274	Rx
Abort-Session-Answer	ASA	274	Rx
Session-Termination-Request	STR	275	Rx
Session-Termination-Answer	STA	275	Rx

3.1.1 Re-Auth-Request/ Re-Auth-Answer

The RAR command is sent by the PCRF to the PCEF in PGW using Gx interface. It is also possible to use RAR commands over Rx interface, but it is beyond the scope of this text. Nevertheless, main purpose of the RAR commands is in provisioning of unsolicited PCC rules, event triggering and event reporting of indications for the ongoing session. The RAA command is a response to the RAR command from PCRF to the PGW.

The AVP for PCC rule installation, update and removal in RAR/RAA commands are named charging-rule-install and charging-rule-remove respectively, and they contain all important information for creating or removing unsolicited PCC rule in the PGW. This AVP also contain information about charging rule name and other charging identifiers.

3.1.2 AA-Request/ AA-Answer

The abbreviation AA stands for authorize and authenticate. The AAR command is sent by an AF, which is basically SIP server inside of the IMS, to the PCRF to provide session information. More specifically, IMS is sending media-component-description AVP containing service information about the media component within the Rx session. This information is intended for PCRF to determine PCC rule selection, authorized QoS and IP flow classifiers.

This command notifies PCRF to send RAR command to create dedicated bearer on top of the default EPS bearer. This AVP contains information about the codec and bitrate for the RTP media stream, IP addresses of the flow route and charging identifiers. In the initial AAR, IMS provide specific-action AVP with selected events when PCRF should contact IMS and notify it about bearer status. For example, failed resource allocation or indication of loss of bearer to terminate the session.

3.1.3 Credit-Control-Request/ Credit-Control-Answer

The CCR command is always sent by the PGW to the PCRF, thus using the Gx interface. It is sent either to request PCC rule for the bearer and provision IP flow mobility routing rules or to indicate event related changes of bearer, PCC rule and IP flow mobility routing rule including termination of the bearer. The CCA command is sent in the opposite direction in response to the CCR command.

Each CCR/CCA command contains cc-request-type AVP with the number identifying the type of the CCR/CCA command. The CCR/CCA commands are divided into four main types with their corresponding numbers as follows:

1. INITIAL_REQUEST
2. UPDATE_REQUEST
3. TERMINATION_REQUEST
4. EVENT_REQUEST

When the default EPS bearer is established, INITIAL_REQUEST type is being sent with session identifier to be able to identify all upcoming messages within this bearer. This message initiates a credit-control session and contains all required information for the initiation. All the upcoming CCR/CCA commands are UPDATE_REQUEST type, updating information about an already created credit-control session. This command should be sent when re-authorization is needed, such as at the expiry of the allocated quota or after expiration of validation timer. The TERMINATION_REQUEST type is sent at the end of credit-control session to terminate it. For example, when the default EPS bearer is no longer available. The EVENT_REQUEST type is used when there is a request of service from PGW. However, this last-mentioned command does not need to be always implemented. All the commands within one session are numbered, starting with INITIAL_REQUEST as number zero and ending after the TERMINATION_REQUEST, when the session is terminated.

When UE change Tracking Area (TA) and location update is being sent to EPC, the credit-control session is updated via CCR. This message contains location information and all the other related information about the subscriber's identity, such as International Mobile Subscriber Identity (IMSI), routing address or International Mobile Equipment Identity (IMEI). All this information is sent because of the IP flow mobility rule in LTE.

3.1.4 Abort-Session-Request/ Abort-Session-Answer

The ASR command is sent by the PCRF to inform the AF that the bearer for established session is no longer available and the session should be terminated. It occurs mainly when subscriber with established dedicated bearer, in other words with ongoing voice call, gets out of LTE network coverage and is either transferred to CS 2G network or the call is interrupted completely. However, ASR/ASA command informs only about the loss of bearer and subsequent STR/STA command is always required to terminate the session.

3.1.5 Session-Termination-Request / Session-Term.-Answer

The STR/STA command is sent by the AF to inform the PCRF that an already established session shall be terminated. In other words, this command notifies the PCRF and subsequently PGW to release dedicated bearer, which was created for the RTP media stream, thus voice call. This command is always sent at the end of successful voice call or after the AST/ASA command. Otherwise, the session is still active when no STR/STA command is being sent. The number of STR or STA commands can be used to determine the number of voice calls because this command is being sent only once for each dedicated bearer, at the end.

3.2 Default EPS bearer setup

When UE is switched on or within reach of the eNB for the first time, UE registers its location to MME and establish necessary connection over the air interface. In case of LTE cell, UE also obtains IP address providing connectivity to default PDN, usually internet. This connectivity to any PDN is called Evolved Packet System (EPS) bearer, which is also default bearer. This procedure of establishing an EPS bearer is repeated when the connectivity to IMS is required. There is a need for additional bearer for SIP message signaling because IMS is a different PDN than internet. Nevertheless, even this EPS bearer for SIP signaling is default bearer and the procedure of establishing is almost the same. This procedure is also repeating every time UE gets out of LTE coverage or after periodical session termination. The DIAMETER message flow for this procedure including PCRF and PGW is shown in figure 3.1.

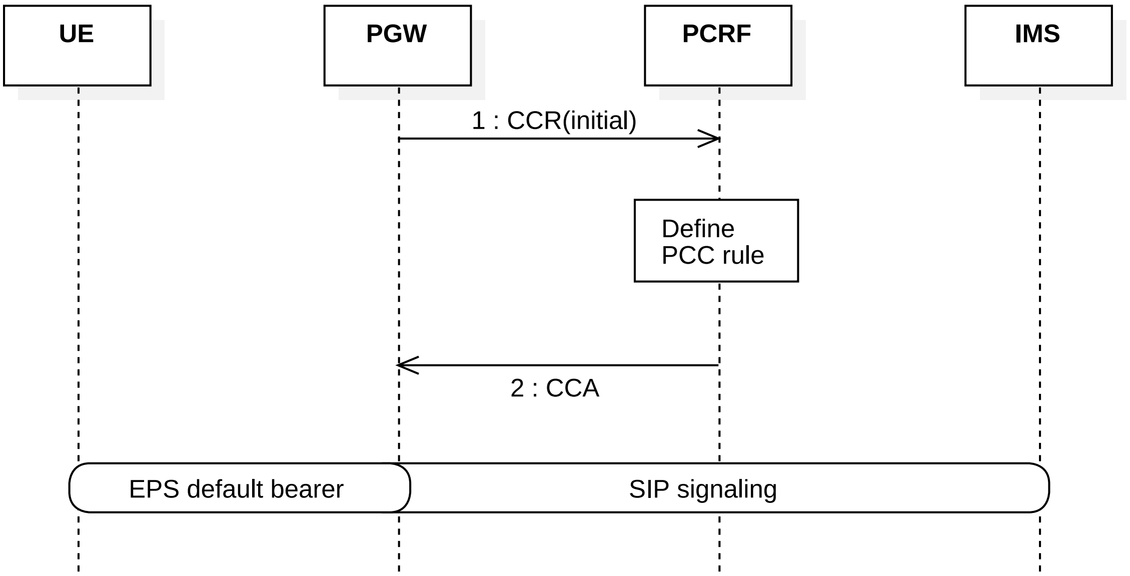


Figure 3.1: DIAMETER message flow of EPS default bearer establishment.

When the UE is connected to the network over the air interface to the serving MME, PGW receives a request to establish default EPS bearer to target PDN with pre-defined QoS, in this case with IMS to be able to receive and send SIP messages to the SIP server, which is part of the IMS.

PGW sends credit-control-request with a subscriber's identifiers and other required information for authentication and authorization. Then the PCRF contacts subscriber database, usually HSS, and gets confirmation of the subscriber's identity with information about allowed services. If the subscriber has approval of access to services, PCRF defines a PCC rule and reply with credit-control-answer, that EPS default bearer can be established. In case of default bearer to IMS, the required service is a voice service.

PCRF has a pre-defined default PCC rule to grant only access with the same QoS parameters for all subscribers in case of default bearer. This default bearer is for signalling purposes and there is no need to differentiate subscribers in control plane. After successful completion of this procedure, UE is connected to IMS via default bearer and can register to SIP server to receive or create requests for a voice call afterward.

3.3 Dedicated bearer setup and release

A dedicated EPS bearer with selected QoS for RTP media stream must be created if there is a voice call request. When the voice call is originating from the server side (IMS), IMS sends an AA-Request command to PCRF with information about required media stream, IP flow route and charging identifiers. The PCRF binds the Rx session ID with the Gx session ID, based on already saved IP address of UE.

The AA-Answer command is sent back to inform IMS that request command was acknowledged. Then the PCRF sends Re-Auth-Request command to PGW in order to install a new PCC rule and create a dedicated bearer in accordance with the rule.

The Re-Auth-Answer is also sent back to PCRF if PGW receives the request. Assuming that no other problem emerged, a dedicated EPS bearer for RTP media stream is created between UE and PGW. During the voice call, the PGW is forwarding all the IP packets of RTP media stream to IMS and back.

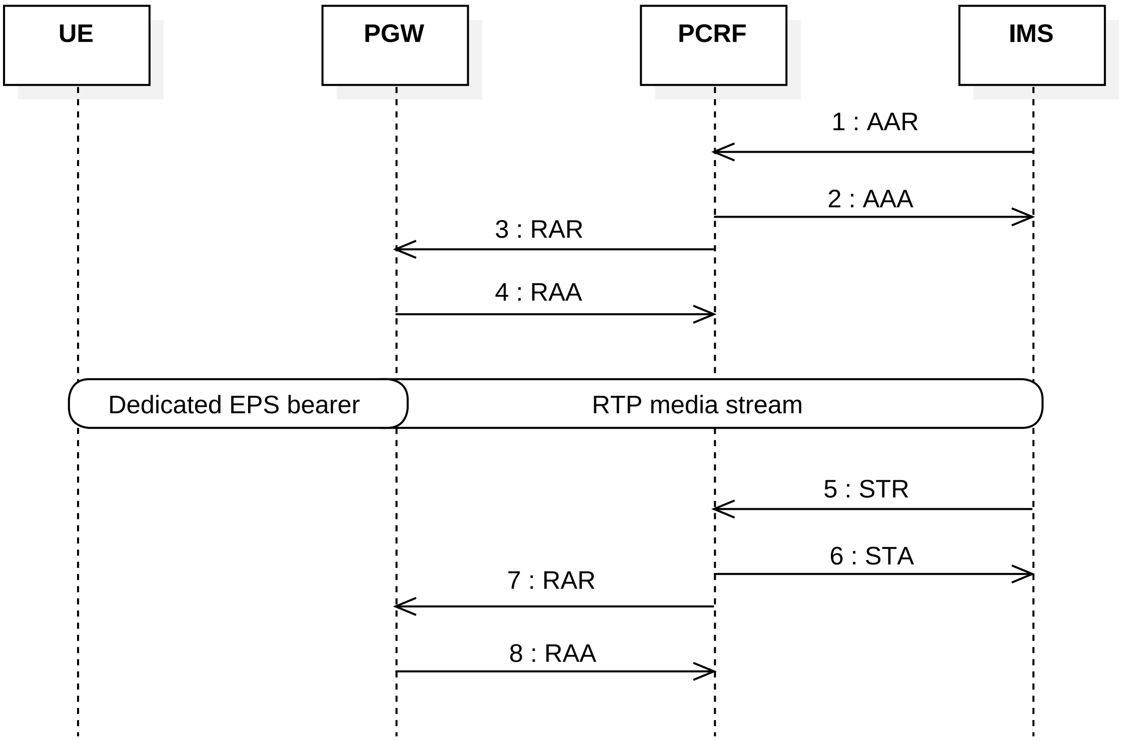


Figure 3.2: VoLTE voice call initiation and release.

When the UE or the other side ends the voice call, the IMS sends a session-termination-request command to PCRF to notify it about the end of the voice call. As usual, session-termination-answer is sent back afterward to congitm it. The PCRF sends the re-auth-request command with information to remove PCC rule and release dedicated EPS bearer. When PGW receives this command, it replies with re-auth-answer command and releases the bearer. From this moment, UE and

IMS can only use the default EPS bearer to communicate with each other, thus send SIP signalling. The entire described message flow is shown in figure 3.2.

Only necessary communication is shown in figure 3.2 and it does not reflect the reality that during the voice call, addition aa-request commands informing about changes can appear. An example might be a change in the used codec to change it to lower bitrate version.

The VoLTE voice call can be also originated from the UE instead of SIP server. In this case, the DIAMETER message flow with PCRF is the same as for server originated voice call and the only difference is in SIP signalling. The call request to SIP server is in this case being sent from UE. SIP signalling is always using the default EPS bearer which has to be already set up in order to establish VoLTE call.

3.4 Circuit Switched fallback

As mentioned earlier, the CS fallback is a situation when UE is reconnected from LTE network to the legacy CS 2G network where the voice call is not handled by the IMS. From the point of view of IMS, the voice call shall be terminated after UE leaves LTE network. Nevertheless, it would cause an undesirable interruption in the voice call, therefore there are other ways how to reconnect the voice call seamlessly from the IMS to 2G network circuit without subscriber's notice. For more details see [2].

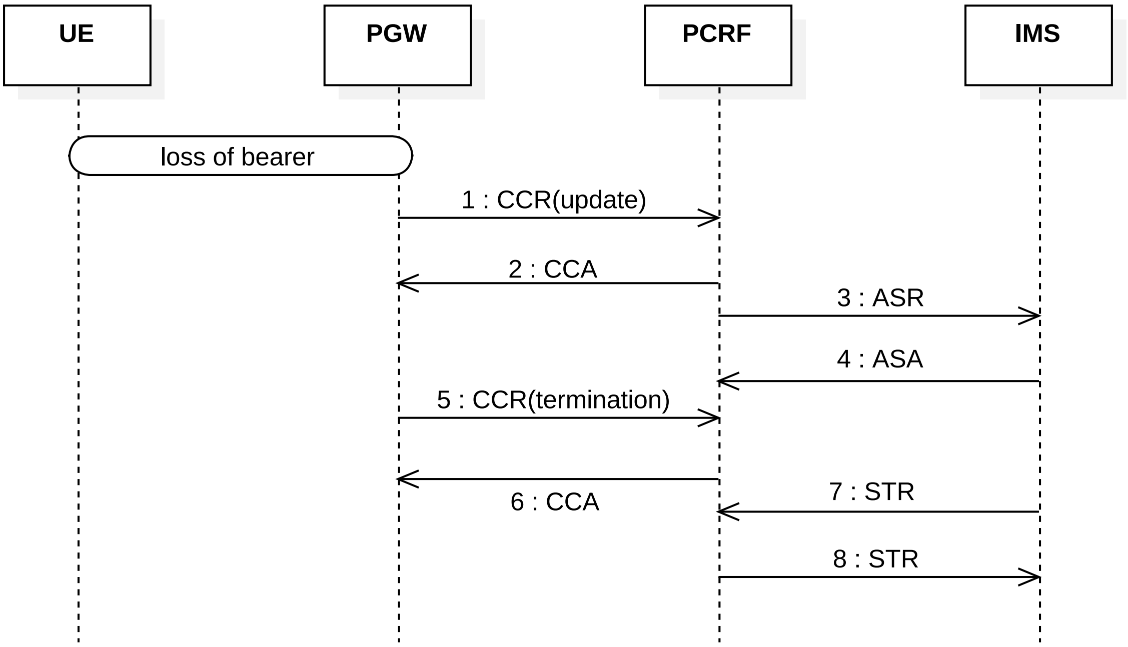


Figure 3.3: Simplified PCRF message flow during CS fallback.

When UE is disconnected from LTE network and the EPS bearer is released, the PGW informs the PCRF about the loss of bearer. However, the loss of bearer is event trigger which needs to be sent using credit-control-request type UPDATE_REQUEST command followed by another credit-control-request type TERMINATION_REQUEST to terminate the Gx session. At the same time, the PCRF sends an abort-session-request command to inform the IMS about the termination of the session. Then IMS sends a session-termination-request command to terminate the unbound Rx session. Simplified PCRF message flow during CS fallback is shown in figure 3.3.

Some messages which are being sent might seem duplicate in this procedure but the network must ensure that both Rx and Gx session is properly terminated. Otherwise, there would be an active unbounded session causing problems and possible instability.

Additionally, this procedure which is shown in figure 3.3 is also occurring when the UE with ongoing VoLTE call moves to an area without any network coverage and both the dedicated and default bearer is terminated. It is impossible to find out from the PCRF whether it resulted in CS fallback or the voice call was fully terminated due to complete loss of network coverage. Both of these events have the same procedure in PCC subsystem and are indistinguishable.

3.5 Diameter error codes

Answer of each DIAMETER message contains AVP indicating whether an error occurred. The name of this AVP is result-code. The value of the result-code AVP indicates both successful and unsuccessful request and consists of five main types of result codes, described by the first digit. Most of the DIAMETER codes used in the network can be found here [13] The result code types are as follows:

- 1xxx (Informational)
- 2xxx (Success)
- 3xxx (Protocol Errors)
- 4xxx (Transient Failures)
- 5xxx (Permanent Failure)

In case of a successful response, the result-code AVP contains value 2001 meaning DIAMETER_SUCCESS to inform that the request was successfully completed. Majority of messages in the network have this result code in the answer.

Permanent failure codes indicate application errors which are permanent and the network cannot recover from them itself. Indicating that the request failed and should not be attempted again. These error codes are especially important in the network because they indicate network problems which should be fixed in order to ensure reliable services. The list of all permanent failure codes contains a tremendous number of possible codes but only a few of them are usually used.

The number of used permanent failure codes depends on the implementation of the service and can vary based on service provider because the occurrence of every code must be specified beforehand. Some used result-code AVPs indicating permanent failures with a short description are:

- 5002 DIAMETER_ERROR_IDENTITIES_DONT_MATCH
  A message was received with a public identity and a private identity for a subscriber, and the server determines that the public identity does not correspond to the private identity.
- 5004 DIAMETER_ERROR_ROAMING_NOT_ALLOWED
  The subscriber is not allowed to roam in the visited network.
- 5012 DIAM_ERROR_SERVING_NODE_FEATURE_UNSUPPORTED
  This error is used when the HSS supports the PCSCF restoration mechanism feature, but none of the subscriber serving node(s) supports it.
- 5030 DIAMETER_USER_UNKNOWN
  The specific subscriber could not be found in the Online Charging Function (OCF).

The full list of the result-code AVP values with a short description and explanation of each can be found here [13, 14, 15]. Note that some of the result-code AVPs are still experimental and not fully specified.

CHAPTER 4

Hadoop ecosystem

This chapter describes used Big Data processing tools with explanation what is their purpose and why these tools should be used. First part is dealing with Apache Hadoop including its background, creation and main components. Then Apache Spark Framework is introduced with explanations why and how it should be used. The last part of this chapter is dealing with the problematic of web-based notebooks and how they make the workflow with Apache Spark more efficient.

4.1 What is Big Data

The increase in the amount of generated data is exponential as well as the speed at which data is generated. This data exceeding processing capabilities of conventional databases using Relational Database Management System (RDBMS) is usually associated with the term Big Data. This data is often described with three self-explanatory words: volume, velocity, and variability [16]. These words also explain the main challenges of big data. Volume for the massive size, velocity because of the speed at which the data is generated and Variability or variety mean how the data is structured because it does not matter whether it is structured, semi-structured or even unstructured. Therefore, a novel approach is required to process this data and be able to extract valuable patterns and information, which were usually hidden when using other conventional methods, such as RDBMS.

The prices of low-cost hard drives and other storage options are generally very low, therefore it is possible to store almost every data that is generated [17]. However, it poses a problem that it is cheaper to store the data than to sort, analyze and delete it. This abundance of data creates a great challenge how to read it effectively. Nevertheless, the rate at which data are being read from the hard drives have not changed much and one of the options how to overcome this limitation is using a multiple of drives in parallel, thus distributed systems.

4.2 The rise of Apache Hadoop

In the past, most of the structured data were stored and then analyzed in RDBMS. For datasets up to tens of GB, this is usually the best option. However, when datasets increase in size RDBMS are becoming very expensive and their processing power is fading slightly due to their limited parallelization. This is one of the reasons why Apache Hadoop was created. Hadoop is a framework based on the idea of distribution and parallelization of tasks. Dividing the tasks into smaller parts and distributing them among cheaper and not so powerful commodity hardware enable to achieve at least the same processing power as with RDBMS but with much lower costs. Another advantage is in its scalability. A new nodes can be added simply when more processing power is needed, instead of upgrading the whole machine with RDBMS. Nevertheless, RDBMS is still a great tool and Hadoop does not aim to replace but to complement it.

One of the main components of Hadoop is Hadoop Distributed File System (HDFS). It is a redundant file system ensuring reliability and fast reading by using redundancy. The replication factor is usually set to three, a compromise between used capacity, reliability, and speed. The second important component is a resource manager taking control of all the machines in the cluster and dividing the tasks. A resource manager in Hadoop is called YARN, and this abbreviation stands for Yet Another Resource Negotiator. The data processing models such as MapReduce, Spark and Tez. are on top of YARN [18]. This stack of the main Hadoop components is shown in figure 4.1.

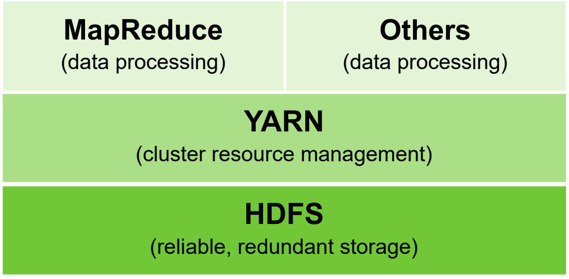


Figure 4.1: Main Hadoop components stack.

4.2.1 MapReduce

An accurate description of MapReduce paradigm from its creators [19, p. 107] is as follows:

“MapReduce is a programming model and an associated implementation for processing and generating large datasets that is amenable to a broad variety of real-world tasks.“

The main idea behind MapReduce is using divide and conquer technique, but apart from other similar algorithms one of the main advantages is the capability of automatic parallelization and distribution among nodes in a cluster. Before Hadoop, MapReduce was developed by Google to enable easier large-scale parallel computation on various data, such as web logs and crawled documents. Hadoop can be understood as an open source implementation of MapReduce framework supported by the Apache Software Foundation. Hadoop and MapReduce do not mean the same thing, Hadoop can exist alone without MapReduce as well as MapReduce. Nowadays, MapReduce is often even replaced by other data processing frameworks such as Spark or Tez and Hadoop takes care of data and resource management whereas data processing is delegated to other frameworks.

4.2.2 Hadoop ecosystem

Eventually, the term Hadoop refers to not only the base Hadoop component stack but all modules and framework that are associated and work on top of Hadoop. Therefore, it is possible to talk about the whole Hadoop ecosystem by using the term Hadoop [18]. It comprises a wide range of commercial tools and open source projects. All of these tools work on top of HDFS and YARN or alongside it. Example of the most used open source tools in Hadoop is shown in figure 4.2 below.

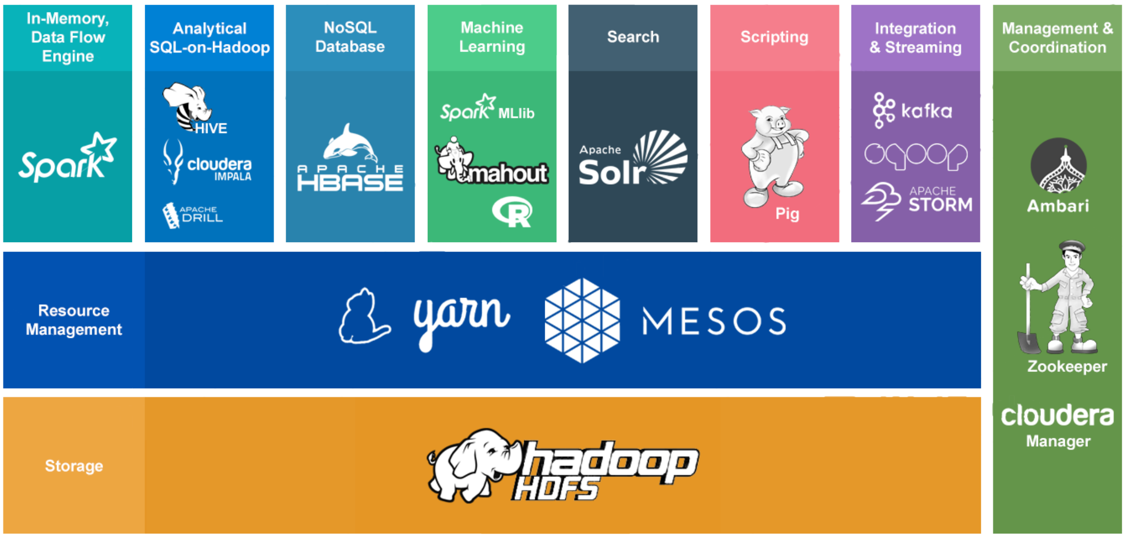


Figure 4.2: An example of Hadoop ecosystem [20].

In addition to YARN, Apache Mesos can be also used as a resource manager and job scheduler. Apache Mesos is newer and more capable sibling of YARN and is used mainly in bigger data centers, where YARN hit the limits [21]. On the left side of the figure 4.2 is Apache Spark, a powerful tool for a fast in-memory data processing which will be further explained later in this chapter. Apache Hive and Cloudera Impala are SQL query engines for data stored in HDFS. Apache HBase is a NoSQL Database intended for sparse data in HDFS. Hadoop also supports machine learning libraries, Spark MLib on top of spark and Mahout on top of MapReduce. Full-text searching and indexing can be done with help of Apache Solr, Apache Pig is a scripting language for simple and quick writing of MapReduce programs. Apache

Storm and Kafka are intended to be used in real-time data processing. Last but not least is Apache Zookeeper for keeping track of all installed Hadoop components. This task is often handled by managers of various Hadoop distributions, such as Cloudera manager in Cloudera distribution.

4.2.3 Hadoop distributions

Most of the Hadoop components are open-source software projects, therefore it is not required to pay for it. However, with the abundant number of various projects and tools, it is very difficult to set all dependencies between them correctly. Mainly, maintaining and testing their interoperability because the projects are updated and developed individually. This is the reason why Hadoop distribution came into play.

Hadoop distributions consist of multiple Hadoop projects and ensure that all the included packages and tools are working flawlessly. It is also easier to deploy because there is no need to install many standalone components and then deal with interoperability. All of it is handled by the Hadoop distribution package. There is only a need to install and deploy this one distribution package that is already tested.

Companies maintaining these Hadoop distributions are also offering support which is important and crucial especially for enterprise customers. Another advantage of ready to deploy distribution is in reliability when most of the bugs are usually removed as quickly as possible. They also offer some additional packages. However, there is one drawback of this solution and it is in its cost because it is necessary to pay for these solutions. Most of the distributions have free versions with lack of support and missing additional tools. Still, these limited versions contain most of the important packages and often are a better option to deploy than standalone packages. For example, almost all the distributions include Apache Spark framework.

Most used commercial Hadoop distributions are Cloudera, Hortonworks and Amazon Elastic [22]. Cloudera was the first to offer multiple Hadoop projects in a single package and continues to lead the market. Hortonworks is trying to offer many of its solutions as open-source and Amazon Elastic Map Reduce is a cloud-only Hadoop-as-a-service platform. The advantage of this approach is in a simple and fast scalability.

4.3 Apache Spark

Apache Spark is a framework for large-scale data processing maintained by the Apache Software Foundation. Apache Spark is also a tool for managing and coordinating the execution of tasks on data across all computers in the cluster. Spark can be either installed as standalone cluster manager or use Hadoop with YARN or Mesos, which were mentioned earlier.

Apache Spark is most often used by accessing two main APIs, Low-level API with Resilient Distributed Dataset (RDD) and Structured API with DataFrame. Spark with RDD was created in response to limitations of MapReduce, mainly by limiting the number of IO operations which are usually slow by using fast memory, typically RAM, when possible. Thanks to these changes, most of the tasks can be done more quickly than with MapReduce [23].

The RDD was the primary API when the Spark was created. RDD is immutable distributed collection of elements of processed data. It is mainly focused on unstructured data such as media streams or streams of text. DataFrame API was later introduced for the structured data where DataFrames are also an immutable collection of data. Unlike the RDD, data in DataFrames are organized into named columns which are similar to tables in conventional RDBMS. DataFrames are a higher level of data abstraction in comparison to RDDs with under-the-hood optimizations, thus it is recommended to use DataFrame when working with structured or semi-structured data [24]. Basic Apache Spark functionality is shown in figure 4.3

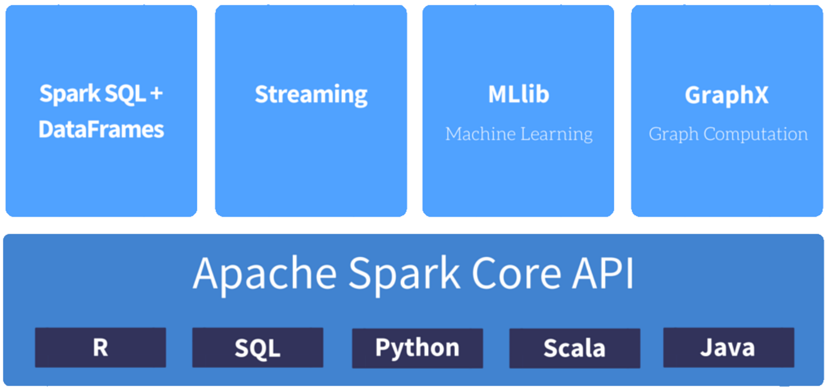


Figure 4.3: Basic Apache Spark functionalities [24].

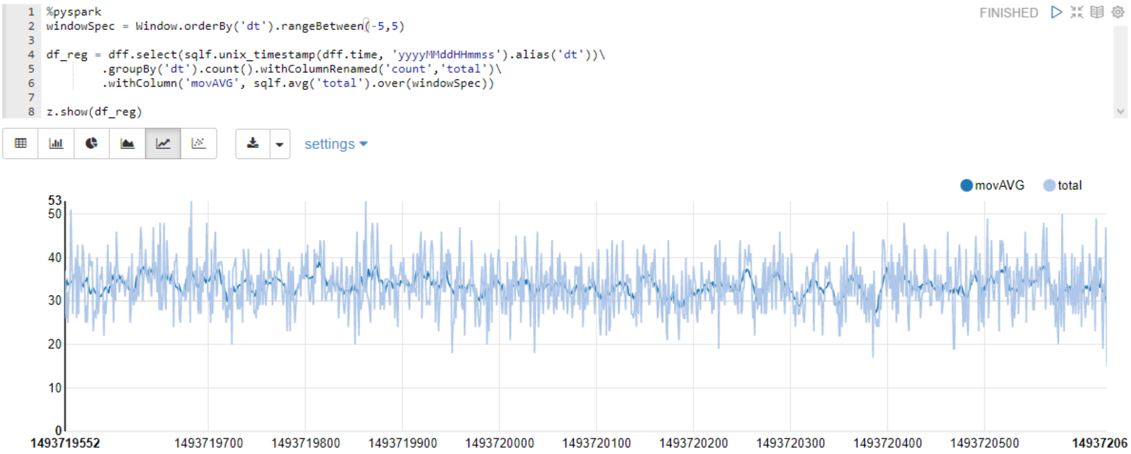
On top of structured API, additional functionalities are built into Spark. Spark streaming allows using near real-time processing. However, Spark cannot work in real-time entirely because it would be very ineffective, therefore it operates in very small micro batches. Other functionalities of Spark include MLlib built-in library for machine learning and GraphX library for graphs and graph-parallel computation. Spark can be also accompanied by other packages from the community. Sometimes these packages move into open source Spark and in next Spark release, they are an integral part of Spark. These packages mostly consist of machine learning and deep learning libraries using Spark as a core and extending its functionality [24]. A consolidated list of packages can be found here [25].

Spark's APIs support a few languages allowing to run the Spark code. The languages are Scala, Python, SQL, Java and R. Spark is primarily written in Scala, thus Scala is the most comprehensive language for the Spark. Python and Java also facilitate almost all operations in Spark whereas R and SQL languages are either limited or not available in low-level APIs. However, Structured API is available in all aforementioned languages.

4.4 Web-based notebooks

Spark code can be run by using Command Line Interface (CLI) in Linux shell. This is the easiest and most convenient way. However, when the code becomes more complex and it is required to do simple edits in the code and re-execute it again, CLI is inapplicable and the code must be saved in a separate file, which will be executed. When the code is saved in the file every time a small change is needed, it is very time demanding. Fortunately, a web-based notebooks combine advantages of both approaches, thus for shorter codes, it is a convenient way how to work with Spark.

Another advantage is a graphical interface of notebooks because it is possible to plot the graph from the output data and immediately visualize the results. The example of visualized time series is shown in figure 4.4 below. These notebooks also contain other interpreters apart from basic Spark, such as R language or SQL.



```
%pyspark
windowSpec = Window.orderBy('dt').rangeBetween(-5,5)

df_reg = dff.select(sqlf.unix_timestamp(dff.time, 'yyyyMMddHHmmss').alias('dt'))\
        .groupBy('dt').count().withColumnRenamed('count','total')\
        .withColumn('movAVG', sqlf.avg('total').over(windowSpec))

z.show(df_reg)
```

Figure 4.4: An example of Spark task in Zeppelin notebook.

There are numerous notebooks available to be used because many companies are developing them. The most used notebooks including Spark interpreter are Jupyter Notebook and Apache Zeppelin. Both are open source, thus free to use. Vendors of Hadoop distributions such as Cloudera are also developing their own web-based notebooks which are part of the distribution. Unfortunately, this additional content in Hadoop distributions is not offered for free but as a part of their price plan for enterprise customers.

CHAPTER 6

Anomaly detection

This chapter contains proposed performance indicator for anomaly detection in VoLTE with explanation what is an anomaly in VoLTE and how this performance indicator works. First, the anomalies are analyzed on a network level and then on a cell level. A few examples of cells classified as anomalous are included with an explanation why they were selected. These cells also contain a list of devices with the highest number of DIAMETER errors.

6.1 What is anomaly detection

An anomaly is defined as something that derives from standard, normal or expected values. In case of LTE network, it might be an increase or decrease of selected parameter indicating issue in the network. For example, an increase in the number of errors indicates a problem with some network subsystem. This change of state derives from expected values, and it is identified as an anomaly. The anomaly itself does not say where the problem is, it is only an indicator. An anomaly can also be a cell that derives from usual parameters in selected indicator.

Anomaly detection, also called outliers detection is identification of selected entities from the data which do not conform to expected values. In LTE network, these outliers can be cells, devices, subscribers and many more. A dataset with selected indicators is required to find outlying values indicating potential problems in the network. Finding the right indicator is the most crucial condition determining the success of finding the network problem.

There are many approaches to find the anomaly in the mobile network. The most common method is using selected Key Performance Indicators (KPIs) from the network and performing various Time-Series analyses [34, 35, 36]. However, this method cannot be utilized in this case, because the analyzed dataset contains data from only one day.

6.2 Proposed performance indicator for anomaly detection

The analyzed dataset contains data from two PCRF interfaces, Rx and Gx interface. In the first step, it is required to assign all the Rx commands and errors to cells which can be found in the Gx interface messages. The output dataset contains information about each cell and all the parameters related to this cell, the number of messages according to their type and the number of messages with errors. Based on this created dataset it is possible to calculated performance indicator for anomaly detection for each cell. This proposed performance indicator for anomaly detection PI_a can be calculated from equation 6.1 below:

$$PI_a = \frac{(n_{er}+1)}{n_{Rx}}(n_{AS}+1) \tag{6.1}$$

where n_{er} is the number of DIAMETER errors related to the selected cell, n_{Rx} is the total number of all Rx interface messages related to the cell and n_{AS} is the number of abort-session command messages related to the cell.

The PI_a given by equation 6.1 is based on the DIAMETER error ratio, which is the number of errors divided by the total number of Rx interface messages. The problem of this error ratio is in cells with high traffic because the ratio is inversely proportional to the number of messages. Then the cells with a high number of messages have low error ratio even though the total number of errors is rather high and indicating an issue with the cell. The error ratio highlights only cells with low traffic.

Another option might be using the total number of errors but it completely hides the cells with low traffic and cannot be used due to this problem. This is the reason why the error ratio in equation 6.1 is multiplied by the number of abort-session messages. This number correlates closely with the number of messages and increase the error ratio in cells with high traffic. A high number of abort-sessions is also undesirable in the network and indicates possible problems in cells.

Both the number of errors and abort-session commands are incremented by the number one because some problematic cells with low traffic have either zero errors or zero abort-session commands and the numerator is the product of these two numbers, therefore none of them must be a zero. Additionally, many errors subsequently cause abort-session command, increasing the value of PI_a even more.

The PI_a takes into account both problem of a high number of errors and a high number of aborted Rx sessions, thus VoLTE calls. The lower the PI_a the better and it converges to zero for cells without any problems. However, the zero value is unachievable and only cells with high traffic and no problems can get close to this value.

Many cells in the analyzed network have a high number of abort-session commands mainly caused by CS fallbacks. The high percentage of these messages increases the PI_a and many cells have this indicator higher than zero. It is important to mention that only cells with at least one VoLTE call or VoLTE call attempt can be analyzed otherwise there are no messages on Rx interface which can be analyzed.

The distribution of the PI_a among cells is skewed to the right as can be seen in the left graph in figure 6.1. This graph contains both number of cells in logarithmic scale and their PI_a. The PI_a distribution is very close to log-normal distribution with a few outliers. The graph on the right in figure 6.1 contains probability plot comparing the log-normal distribution of estimated parameters from the dataset with real PI_a values. It is evident that there are outliers with a high value of the indicator pointing out to possible problems with these cells.

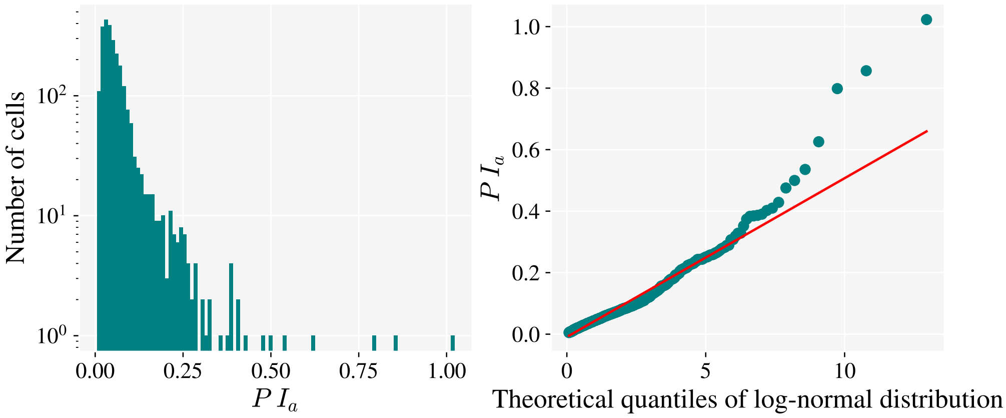


Figure 6.1: Distribution of PI_a among cells and log-normal dist. fitting.

One of the options how to select outliers is to use multipliers of standard deviation, such as three sigma rule to select outliers with values outside of this interval. However, the presence of outliers is likely to have a significant effect on the mean and the standard deviation. Additionally, this approach works best with regular normal distribution, but the performance indicator has likely log-normal distribution. Given these facts, an approach using median absolute deviation has been chosen to select the outliers [37]. The median absolute deviation is calculated given by equation 6.2 below:

$$MAD = median(\mid x_i - median(x) \mid) \tag{6.2}$$

where $x_1, x_2, x_3, \dots, x_N$ are the PI_a values for each cell and N is the number of analyzed cells in the network.

The condition for classifying a cell as an outlier using median absolute deviation of PI_a is given by equation 6.3 below:

$$\frac{\mid x_i - median(x) \mid}{MAD} > thr \tag{6.3}$$

where x_i is the PI_a value of selected cell, $median(x)$ is median of PI_a values of all the cells in the network and thr is the threshold value for selecting an outlier.

A cell is classified as an outlier when the condition in equation 6.3 is met. The threshold value *thr* is based on observation of the dataset and state of the analyzed network. Initially, a condition $thr = 15$ was chosen to select only a few of the biggest outliers. This value is the highest multiplier of median absolute deviation that falls within the range of regular PI_a.

Another reason for selecting the outlier detection using median absolute deviation instead of a simple pre-defined static value of this indicator is the possibility to adjust the border for outliers dynamically. This feature is necessary when analyzing multiple networks and each of them is in a different state. For example, when the network has a much smaller number of CS fallbacks and the distribution of PI_a values is different. The static pre-defined value would not reflect this change, unlike this method. The example of selected outliers with threshold value fifteen is in figure 6.2.

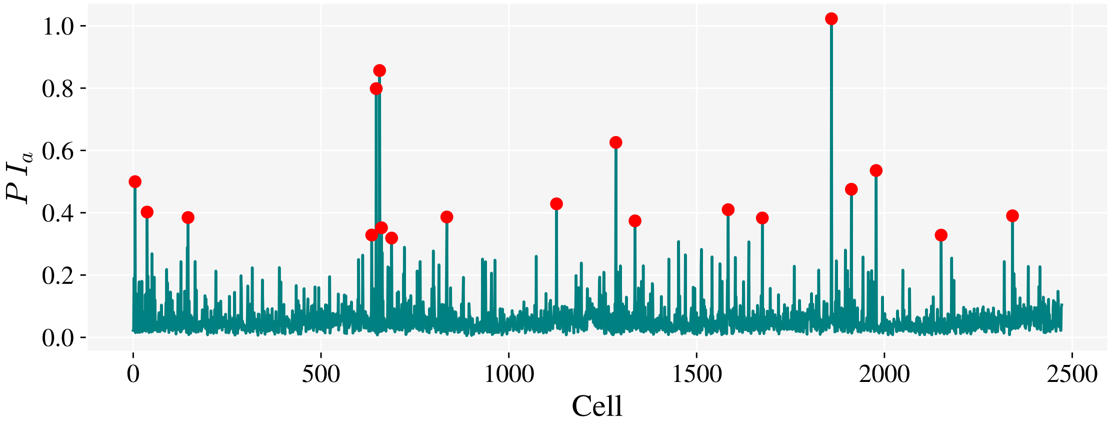


Figure 6.2: PI_a value of cells including outliers with $thr = 15$.

Almost 2500 cells were analyzed and the graph in figure 6.2 comprises the PI_a values of each cell. Selected cells with PI_a having higher than the threshold value $thr = 15$ are highlighted with red-filled circles. Only around twenty cells were classified as an outlier when threshold value $thr = 15$ was used. Their PI_a value is rather high in comparison to the rest of the cells and it is evident that these cells pose some problem and should be further analyzed.

The threshold value should be set in accordance with the required state of the network. Lower values should be selected when having higher demands on the network and vice versa. It is important to mention that each network can have a completely different distribution of PI_a based on the geographical location and landscape, and it is impossible to compare PI_a values of two different networks. The graph in figure 6.3 contains marked outliers for lower threshold value $thr = 5$. The number of the outliers is much higher than in the previous case in figure 6.2.

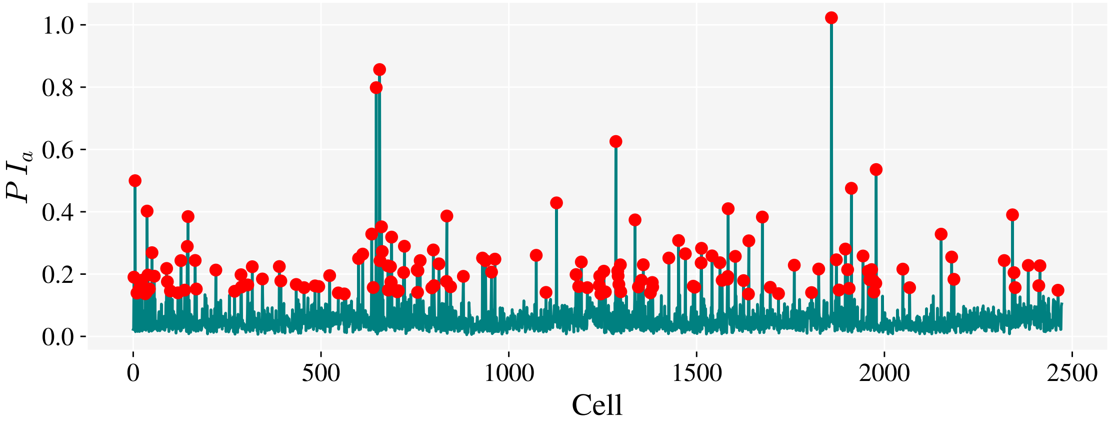


Figure 6.3: PI_a value of cells including outliers with $thr = 5$.

The PI_a values around number one or smaller do not pose a serious problem. Cells with a high percentage of DIAMETER errors and abort-session commands would have a much higher value of the PI_a. In this case, it only signalizes that either some subscribers are having a problem with the cell or the overall status of the cell is not perfect.

6.3 Performance indicator of single cells

Overall PI_a value of each cell is a perfect tool for detection of potential problems of the cells. However, it cannot be used to detect the cause of the problem, whether it is cell related or only related to a particular device. For this purpose, the PI_a must be calculated for each subscriber inside of a single cell. Then it is possible to determine that selected cell is operating normally and the problem is only caused by a particular device which does not work well with the cell. Typically, it can be a device with a different set of pre-defined timers as explained in chapter 3 dealing with DIAMETER errors.

Many subscribers have only a few VoLTE calls per day, typically one or two. Proposed PI_a reaches much higher values than zero for subscribers with a very low number of VoLTE calls. A non-problematic subscriber with very low activity can even have PI_a close to number one. It can be explained by examining the equation 6.1. A subscriber with a single VoLTE call must initiate at least two PCRF related messages, one for session establishment and another one for session termination. In case of session abortion mostly caused by CS fallback, the minimal number of messages is increased by the abort-session command to number three. In this case, the subscriber has $PI_a = 0.67$ as is shown in the example below:

$$PI_a = \frac{(n_{er}+1)}{n_{Rx}}(n_{AS}+1) = \frac{(0+1)}{3}(1+1) = \frac{2}{3} \doteq 0.67$$

In such case, the abort-session command is incremented by one and divided by three messages in total as shown in the example above. Value 0.67 is the upper limit for normally operating subscribers and a subscriber with no errors cannot have a higher value of PI_a than this. Therefore, this anomaly detection approach is using apriori information that PI_a of normally behaving subscriber should be lower than 0.67. This fact is not evident when calculating overall PI_a of the cell because there are more VoLTE calls than only one. However, this case would indicate a problem with the cell because in each cell should be more VoLTE calls than just a single call.

The most problematic cells in the network are further analyzed by computing PI_a of each subscriber inside of the cell. A device which was used during the session with this cell is assigned to each subscriber. Then the cell can be analyzed by examining the number of subscribers with PI_a value higher than 0.67. In case of a small number of subscribers in the cell with indicator value over the limit, the cell is operating normally. Only subscribers with PI_a value higher than 0.67 are experiencing a problem that should be resolved by examining used devices. An example of cells with the highest overall PI_a values are shown in the following figures, starting with the figure 6.4.

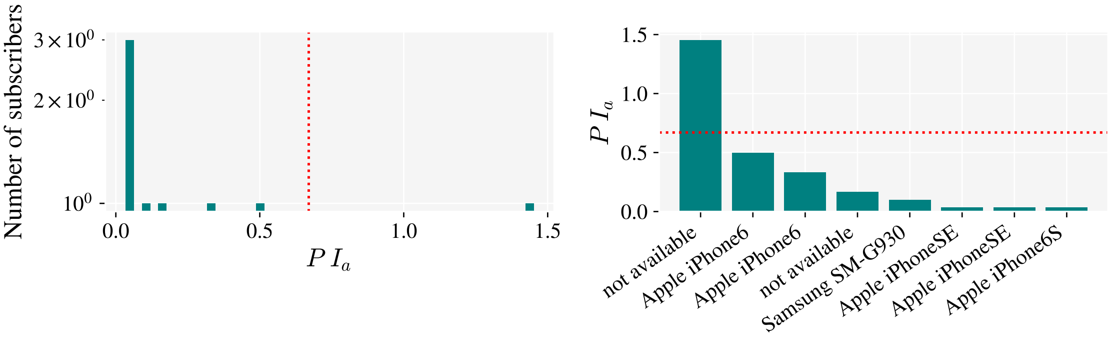


Figure 6.4: An example of problematic rural cell with low number of subscribers.

The cell which is shown in figure 6.4 is an example of a rural cell with a very low number of subscribers that have at least one VoLTE call attempt. It is impossible to detect subscribers in the cell without VoLTE call attempt because they have not made a contact with PCRF. Therefore the term subscriber will be used in the subsequent text only for the subscribers that made a contact with PCRF.

There were only eight subscribers establishing VoLTE call as can be seen from the histogram in figure 6.4. However, only one subscriber has PI_a value above the limit pointing out to a problematic device or other problem with the subscriber. The name of this device is not available because this TAC was not on the TAC list provided by the operator. Usually, newer devices are missing from the list. This cell had one of the highest PI_a value in the network caused by only a single device because of the small number of subscribers.

The cell in figure 6.5 is a regular urban cell with a high value of PI_a. This cell has a high number of subscribers, in comparison to the rural cell in figure 6.4. In this case, it is even more evident that the problem is caused by only a few subscribers because the vast majority of subscribers has PI_a lower than 0.67 as can be seen in the histogram. The border of 0.67 is marked with the red dotted line in the histogram. The same situation with unrecognized devices with a high value of PI_a repeats as in the previous figure 6.4.

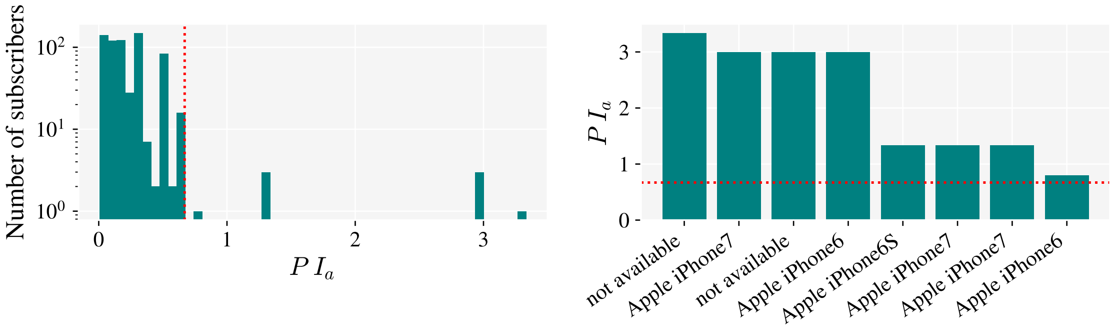


Figure 6.5: First example of problematic urban cell with high number of subscribers.

There are only Apple iPhone devices in the list with the highest PI_a values in figure 6.5 and it might seem that in this case, Apple iPhone devices are more prone to errors. It is not entirely true as is shown in the analyses in chapter 5. The reason is in much higher number of subscribers with Apple iPhone devices using VoLTE. On average, subscribers having Apple iPhone devices have the highest number of VoLTE calls, therefore almost every cell has the highest number of subscribers using Apple iPhone devices.

Another example of a problematic urban cell is in figure 6.6. Only by looking at the histogram of PI_a values of subscribers it is possible to say that the problem is related to certain subscriber's devices. In comparison to the cell in 6.5, the problematic outliers have smaller indicator value, thus a lower number of errors.

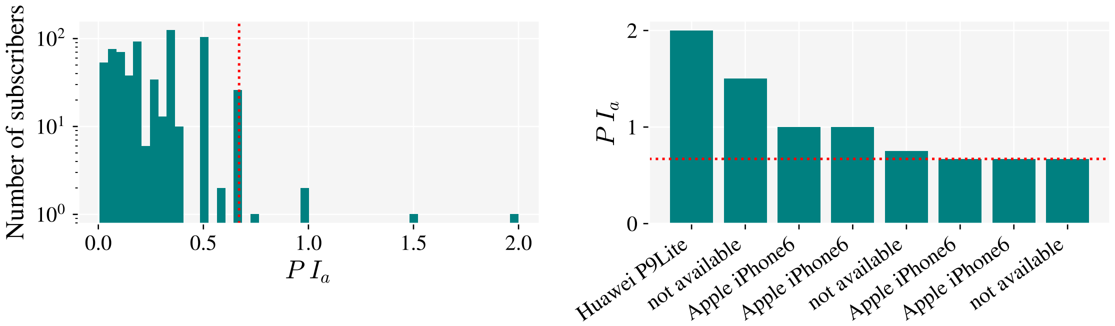


Figure 6.6: Second example of problematic urban cell with high number of subscribers.

All the previous graphs with problematics cells that were shown in figures 6.4, 6.5 and 6.6 contains cells with one of the highest overall PI_a value among all the cells. It is the reason why they were selected as an example. On the other hand, most of the cells in the network have the PI_a value closer to zero. An example of a typical urban cell with a very low overall PI_a value is shown in figure 6.7. None of the subscribers have the PI_a value higher than 0.67 indicating no major problem with the devices nor the subscribers.

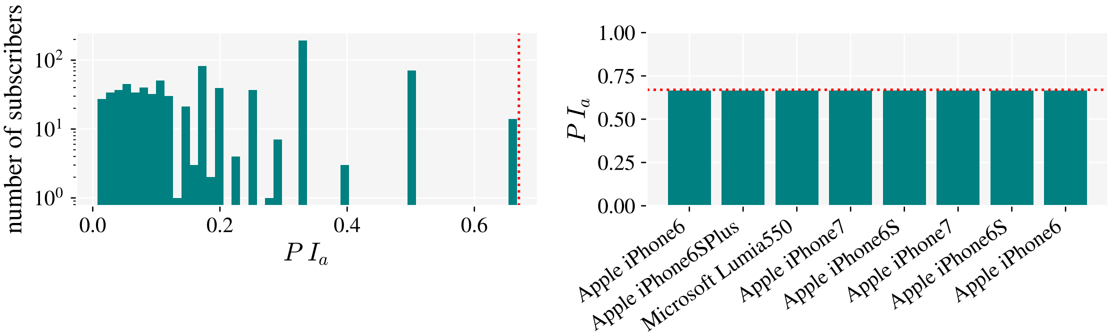


Figure 6.7: An example of urban cell with low value of PI_a.

A high number of subscribers have PI_a value either 0.67 or 0.5 in all of the examples. Majority of these subscribers have only one VoLTE call with CS fallback, respectively two of them. Although it does not demonstrate that the subscribers are experiencing errors, these high values indicate possible problems because all the VoLTE calls of these subscribers were aborted due to CS fallback or complete loss of network coverage. Basically, all the isolated peaks with higher indicator values in the histogram are mainly caused by the low VoLTE activity of selected subscribers. Most of the cells have a very similar distribution of PI_a values as the cell in figure 6.7.

Thanks to this approach using performance indicator for anomaly detection, it is possible to determine whether a cell itself has a problem or it is caused by certain individual devices. Additionally, the devices causing most problems are identified for each cell to be able to resolve this issue. The drawback of this approach is in the impossibility to analyze the cells without VoLTE traffic. However, there are many more different approaches that can be used and are easier to carry out. For example, visualization of these cells on the map as is shown in chapter 5.

CHAPTER 7

Conclusion

The aim of this diploma thesis was to develop procedures and subsequent analyses to identify problems in implemented VoLTE technology. In the first step, a thorough exploration of VoLTE and PCC implementation was conducted using provided dataset and 3GPP documentation. This data was subsequently analyzed to identify the cause of errors in VoLTE. Big Data approach was used to analyze the data, specifically Apache Hadoop with Spark was used.

This work has found out that most of the VoLTE issues in the network are related to certain cells and devices which have interoperability issues caused by a high variety of vendors. These cells and devices were detected utilizing proposed anomaly detection algorithm which is using performance indicator to find outliers in the network. This proposed performance indicator can be also used to fine tune implementation of VoLTE in the network.

Another benefit of the conducted analyses is in the possibility to localize all problematic events of VoLTE geographically and then be able to identify problematic cells. A solution was introduced using a web application with a dynamic choropleth map containing various data layers. A simple version of this web application was also created. The results of analyses can be visualized in a form of dashboard or put in a document which would be sent periodically to an email address to inform about the current state of VoLTE in the network.

Future work

With a dataset over a longer period of time, it might be possible to do time-series analyses utilizing proposed performance indicator and employ robust statistical methods, such as seasonal and trend decomposition, to find anomalies which might appear irregularly. There are many other methods how to conduct time-series analyses including various machine learning techniques. In this case, other indicators could be selected to compare how the correlation between these indicators change in time to be able to observe and classify anomalies more precisely.

Conducting the analyses is only one part of the process. The latter and often more demanding part is to implement the achieved results in the mobile network. It includes finding out the measures which should be taken after identifying the anomaly. This area is also required to be thoroughly examined.

This work illustrates that Big Data approach can be utilized even for mobile networks and data obtained from them. It is an advantageous option how to analyze problems by using data from the core network. There are also drawbacks in using this method, however, the advantages easily outbalance them.

Bibliography

[1] Ericsson. *Traffic Exploration.* Nov. 2017. URL: www.ericsson.com/TET (visited on 02/03/2018).

[2] Christopher Cox. *An introduction to LTE, LTE-advanced, SAE, VoLTE and 4G mobile communications.* Chichester, West Sussex, United Kingdon ; Hoboken, New Jersey: John Wiley & Sons, Inc, 2014. ISBN: 978-1-118-81803-9.

[3] Erik Dahlman, Stefan Parkvall, and Johan Sköld. *4G, LTE-Advanced Pro and The Road to 5G.* Boston, MA: Elsevier, 2016. ISBN: 978-0-12-804575-6.

[4] 3GPP TS 23.203. *Policy and charging control architecture, Release 14.* Oct. 2017.

[5] 3GPP TS 29.212. *Policy and Charging Control (PCC); Reference points, Release 14.* Jan. 2018.

[6] 3GPP TS 29.214. *Policy and charging control over Rx reference point, Release 14.* Jan. 2018.

[7] Miikka Poikselkä et al. *Voice over LTE (VoLTE).* English. Chichester, West Sussex, United Kingdom: John Wiley & Sons Inc., 2012. ISBN: 978-1-119-94493-5.

[8] Vaishali Paisal. "Seamless voice over LTE". en. In: IEEE, Dec. 2010, pp. 1–5. ISBN: 978-1-4244-7930-6. DOI: 10.1109/IMSAA.2010.5729423.

[9] Travis Russell. *Signaling system #7.* Sixth Edition. OCLC: ocn864700445. New York: McGraw-Hill Education, 2014. ISBN: 978-0-07-182214-5.

[10] 3GPP TS 23.228. *IP Multimedia Subsystem (IMS); Stage 2, Release 14.* Jan. 2018.

[11] Spirent. *IMS Procedures and Protocols: The LTE User Equipment Perspective.* Mar. 2014. URL: https://www.spirent.com/Assets/WP/WP_LTE_User_Equipment_Reference_Guide.

[12] G. Zorn. *Diameter Network Access Server Application, RFC 7155.* Apr. 2014. URL: https://tools.ietf.org/html/rfc7155.

[13] 3GPP TS 29.230. *Diameter applications; 3GPP specific codes and identifiers, Release 14.* Jan. 2018.

[14] 3GPP TS 32.299. *Telecommunication management; Charging management; Diameter charging applications, Release 14.* Jan. 2018.

[15] 3GPP TS 29.229. *Cx and Dx interfaces based on the Diameter protocol; Protocol details, Release 14*. July 2017.

[16] Edd Wilder-James. *What is big data?* Jan. 2012. URL: https://www.oreilly.com/ideas/what-is-big-data (visited on 03/02/2018).

[17] Jim O'Reilly. *Big Data Storage: 7 Key Factors*. Feb. 2017. URL: https://www.networkcomputing.com/storage/big-data-storage-7-key-factors/1380415337 (visited on 02/02/2018).

[18] Tom White. *Hadoop: the definitive guide*. eng. 2. ed. Yahoo! Press. OCLC: 838311718. Beijing: O'Reilly, 2011. ISBN: 978-1-4493-8973-4.

[19] Jeffrey Dean and Sanjay Ghemawat. "MapReduce: simplified data processing on large clusters". en. In: *Communications of the ACM* 51.1 (Jan. 2008), p. 107. ISSN: 00010782. DOI: 10.1145/1327452.1327492.

[20] Shailendra Chauhan. *Understanding Apache Hadoop Ecosystem and Components*. Mar. 2018. URL: https://www.dotnettricks.com/learn/hadoop/apache-hadoop-ecosystem-and-components (visited on 04/04/2018).

[21] Apache Software Foundation. *Apache Mesos*. 2018. URL: http://mesos.apache.org (visited on 04/04/2018).

[22] Bernard Marr. *Big Data?: Who Are The Best Hadoop Vendors In 2017?* Jan. 2017. URL: https://www.linkedin.com/pulse/big-data-who-best-hadoop-vendors-2017-bernard-marr (visited on 05/04/2018).

[23] Wissem Inoubli et al. "An Experimental Survey on Big Data Frameworks". en. In: *Future Generation Computer Systems* (Apr. 2018). arXiv: 1610.09962. ISSN: 0167739X. DOI: 10.1016/j.future.2018.04.032.

[24] Bill Chambers and Matei Zaharia. *Spark: the definitive guide : big data processing made simple*. English. OCLC: 982651178. 2018. ISBN: 978-1-4919-1221-8.

[25] Spark Packages. *A community index of third-party packages for Apache Spark*. 2018. URL: https://spark-packages.org (visited on 05/04/2018).

[26] Apple Inc. *iPhone 5 - Technical Specifications*. Oct. 2016. URL: https://support.apple.com/kb/sp655 (visited on 03/04/2018).

[27] Statcounter. *Mobile Operating System Market Share Worldwide*. Apr. 2018. URL: http://gs.statcounter.com/os-market-share/mobile/worldwide (visited on 05/05/2018).

[28] Yuanyuan Huang et al. "Comparison of average global exposure of population induced by a macro 3G network in different geographical areas in France and Serbia: Average EMF Exposure of Population". en. In: *Bioelectromagnetics* 37.6 (Sept. 2016), pp. 382–390. ISSN: 01978462. DOI: 10.1002/bem.21990.

[29] Geertje Goedhart et al. "Using software-modified smartphones to validate self-reported mobile phone use in young people: A pilot study: Validating Mobile Phone Use in Young People". en. In: *Bioelectromagnetics* 36.7 (Oct. 2015), pp. 538–543. ISSN: 01978462. DOI: 10.1002/bem.21931.

[30] Toon De Pessemier et al. "Analysis of the quality of experience of a commercial voice-over-IP service". en. In: *Multimedia Tools and Applications* 74.15 (Aug. 2015), pp. 5873–5895. ISSN: 1380-7501, 1573-7721. DOI: 10.1007/s11042-014-1895-4.

[31] Toon De Pessemier et al. "Quality assessment and usage behavior of a mobile voice-over-IP service". en. In: *Telecommunication Systems* 61.3 (Mar. 2016), pp. 417–432. ISSN: 1018-4864, 1572-9451. DOI: 10.1007/s11235-014-9961-9.

[32] CZSO. *Okresy - LAU 1*. Mar. 2018. URL: https://www.czso.cz/csu/rso/okresy_nuts4 (visited on 05/04/2018).

[33] Vladimir Agafonkin. *Leaflet JavaScript library*. 2017. URL: https://leafletjs.com (visited on 05/04/2018).

[34] Gabriela F. Ciocarlie et al. "Detecting anomalies in cellular networks using an ensemble method". en. In: IEEE, Oct. 2013, pp. 171–174. ISBN: 978-3-901882-53-1. DOI: 10.1109/CNSM.2013.6727831.

[35] Jun Wu et al. "CellPAD: Detecting Performance Anomalies in Cellular Networks via Regression Analysis". In: vol. 2018. Zurich, Switzerland, May 2018. ISBN: 978-3-903176-08-9.

[36] Ahmed Zoha et al. "Data-driven analytics for automated cell outage detection in Self-Organizing Networks". en. In: IEEE, Mar. 2015, pp. 203–210. ISBN: 978-1-4799-7795-6. DOI: 10.1109/DRCN.2015.7149014.

[37] Jordan Hochenbaum, Owen S. Vallis, and Arun Kejariwal. "Automatic Anomaly Detection in the Cloud Via Statistical Learning". In: *CoRR* abs/1704.07706 (2017). URL: http://arxiv.org/abs/1704.07706.