University Institute of Lisbon

Department of Information Science and Technology

Supervisor Virtual Machine for VoLTE Service on a Cloud Environment

Ricardo Jorge Gomes da Silva Almeida

A Dissertation presented in partial fulfillment of the Requirements for the Degree of

Mestrado em Engenharia de Telecomunicações e Informática

Supervisor

Prof. Dr. Pedro Joaquim Amaro Sebastião, Professor Auxiliar

ISCTE-IUL

October 2019

"Fortune favours the brave."

Publius Terentius Afer

Resumo

Com o crescimento contínuo das redes de *Voice over Long Term Evolution* (*VoLTE*), juntamente com a necessidade de redução de custo de manutenção pelos Operadores Móveis, a implementação do Serviço em *Cloud* começa a ser uma aplicação comum.

Este estudo foi elaborado com o intuito de criar um método capaz de melhorar as atividades de Operações e Monitorização de um serviço de *VoLTE*, que esteja implementado numa plataforma *Cloud*. Neste estudo, encontram-se presentes conteúdos referentes aos elementos constituintes de uma rede de *VoLTE*, e é revisto em detalhe as funcionalidades do *Telephony Application Server*. Para este estudo, o *TAS* utilizado foi o *open TAS*. Neste estudo, igualmente é incluído uma explicação genérica do comportamento da *Cloud* Openstack.

O método apresentado implica criação de uma máquina virtual *Supervisor* e da sua implementação na *Cloud*. Esta máquina virtual é capaz de estabelecer ligações *SSH* com o *open TAS*, de modo a extrair o relatório de *Clear Codes*, que identifica o estado com que as chamadas foram finalizadas, para proceder a análises. A máquina virtual contém limites definidos, os quais verifica se foram excedidos. Caso este evento seja verificado, notificam o operador do sistema para um incidente.

Esta é uma proposta que apresenta as possibilidades de implementação num ambiente de *Cloud*, em melhorar e automatizar as funções de Operações e Manutenção na rede de Telecomunicações.

Palavras-chave: *VoLTE*, *Cloud*, Openstack, Máquina Virtual, *Clear Codes*, Automação.

Abstract

With the continuing growth of *Voice of Long Term Evolution* (*VoLTE*) networks, coupled with the need of Mobile Operators to reduce maintenance costs, Cloud Service deployment is becoming a common application.

This study was designed to create a method capable of improving the Operations and Monitoring activities of a *VoLTE* service that is deployed on a *Cloud* platform. In this study, we present contents referring to the constituent elements of a *VoLTE* network, and we review in detail the features of the *Telephony Application Server*. For this study, *TAS* used was the *Open TAS*. Also included in this study is a generic explanation of *Cloud* Openstack's behavior.

The presented method implies the creation of a virtual machine *Supervisor* and its deployment in *Cloud*. This virtual machine is capable of establishing *SSH* connections with *open TAS* to extract the *Clear Codes* report, which identifies the state with which calls were terminated for analysis. The virtual machine contains defined limits, which check if they have been exceeded. If this a limit is excited, the virtual machine notifies the system operator of an incident.

This study presents the possibilities of implementation in a *Cloud* environment, to improve and automate Operations and Maintenance functions in the Telecommunications network.

Keywords: *VoLTE*, *Cloud*, Openstack, Virtual Machine, *Clear Codes*, Automation.

Acknowledgements

I want to acknowledge in the first place Professor Dr Pedro Sebastião, for allowing me this opportunity to develop this concept and providing advice on the creation of this study.

I want to acknowledge Engineer Diogo Clemente, for the support provided in the design of the method and his availability. Also, I want to share my gratitude for the help of my team, in particular, Engineer Theodoros Kafadaris and Engineer Zoi Moutopoulou for providing me with the support to have availability to develop this work.

I also want to thank my family, friends and my girlfriend. They were present during the entire period of creation of this study and supported me in all kind of scenarios, in this past year.

Contents

List of Figures

Abbreviations

3G	**3**rd **G**eneration
4G	**4**th **G**eneration
AI	**A**rtificial **I**ntelligence
AMR	**A**daptive **M**ulti **R**ate
API	**A**pplication **P**rograming **I**nterface
ATCF	**A**ccess **T**ransfer **C**ontrol **F**unction
ATGW	**A**ccess **T**ransfer **G**ateway
BICC	**B**earer **I**n-dependent **C**all **C**ontrol
BGCF	**B**rakeout **G**ateway **C**ontrol **F**unction
CC	**C**lear **C**odes
CHU	**C**harging **U**nit
CMM	**C**entral **M**emory and **M**arker
CS	**C**ircuit **S**witch
DL	**D**own **L**ink
E2E	**E**nd - **to** - **E**nd
EPC	**E**volved **P**acket **C**ore
EVS	**E**nhanced **V**oice **S**ervice
GSM	**G**obal **S**ystem **M**obile Telecommunication
HD	**H**igh **D**efinition
HLR	**H**ome **L**ocal **R**egister
HSS	**H**ome **S**ubscriber **S**erver
IaaS	**I**nfrastructure **as a** **S**ervice
IBCF	**I**nterconnection **B**order **C**ontrol **F**unction
I-CSCF	**I**nterrogating **C**all **S**ession **C**ontrol **F**unction

IM-SSF	**I**P **M**ultimedia **S**ervice **S**witching **F**unction
IMPI	**I**P **M**ultimedia **P**rivate **I**dentity
IMPU	**I**P **M**ultimedia **PU**blic Identity
IMS	**I**ntegrated **M**ultimedia **S**ystem
IMS-ALG	**IMS** - **A**pplication **L**evel **G**ateway
IMS-AGW	**IMS** - **A**ccess **G**ate**w**ay
IMS-MGW	**IMS** - **M**edia **G**ate**w**ay
IN	**I**ntelligent **N**etwork
IP	**I**nternet **P**rotocol
IPDU	**IP** **D**irector **U**nit
IPSec	**I**nternet **P**rotocol **Sec**urity
IP-SM-GW	**IP** - **S**igle **M**essage - **G**ate**W**ay
ISIM	**I**P Multimedia **S**ervices **I**dentity **M**odule
GISU	**G**eneric **I**P **S**ignaling **U**nit
GPLU	**G**eneral **P**urpose **L**inux **U**nit
KPI	**K**ey **P**erformance **I**ndicator
LTE	**L**ong **T**erm **E**volution
LTE-A	**LTE** **A**dvanced
MAC	**M**edium **A**ccess **C**ontrol
MIMO	**M**ulti - **I**nput - **M**ulti -**O**utput
MGCF	**M**edia **G**ateway **C**ontrol **F**unction
MME	**M**obility **M**anagement **E**ntity
MMTel	**M**ulti **M**edia **Tel**ephony
MMTel AS	**M**ulti **M**edia **Tel**ephony **A**pplication **S**erver
MRF	**M**edia **R**esource **F**unction
MRFC	**M**edia **R**esource **F**unction **C**ontroller
MW	**M**aintenance **W**indow
NB	**N**arrow **B**and
oTAS	**o**pen **T**elephony **A**pplication **S**erver
ODFMA	**O**rthogonal **F**requency **D**ivision **M**ultiple **A**ccess
OLA	**O**peration **L**evel **A**greement

OMU	**O**peration and **M**aintenance **U**nit
OTT	**O**ver-**T**he-**T**op
PAPR	**P**eak-to-**A**verage-**P**ower **R**atio
PCC	**P**olicy and **C**harging **C**ontrol
PCRF	**P**olicy **C**harging and **R**ules **F**unction
P-CSCF	**P**roxy **C**all **S**ession **C**ontrol **F**unction
PDCP	**P**acket **D**ata **C**onvergence **P**rotocol
PDN	**P**acket **D**ata **N**etwork Gateway
PMN	**P**ublic **M**obile **N**etwork
PS	**P**acket **S**witch
PRD	**P**ermanent **R**eference **D**ocument
QoS	**Q**uality **o**f **S**ervice
TAS	**T**elephony **A**pplication **S**erver
TCP	**T**ransmission **C**ontrol **P**rotocol
T-ADS	**T**erminating - **A**ccess **D**omain **S**election
TrGW	**Tr**ansition **G**ate**w**ay
RANAP	**R**adio **A**ccess **N**etwork **A**pplication **P**art
RLC	**R**adio **L**ink **C**ontrol
RTP	**R**eal-Time**T**ransport **P**rotocol
RTCP	**R**eal-Time**T**ransport **C**ontrol **P**rotocol
RRC	**R**adio **R**esource **C**ontrol
SC-FDMA	**S**ingle **C**arrier - **F**requency **D**ivision **M**ultiple **A**ccess
SCC-AS	**S**ession **C**all **C**ontinuity - **A**pplication **S**erver
SCP	**S**ervice **C**ontrol **P**oint
SCTP	**S**tream **C**ontrol **T**ransmission **P**rotocol
S-CSCF	**S**erving **C**all **S**ession **C**ontrol **F**unction
SDP	**S**ession **D**escription **P**rotocol
SEE	**S**ervice **E**xecution **E**nvironment
S-GW	**S**erving - **G**ate**w**ay
SIP	**S**ession **I**nitiation **P**rotocol
SIP UA	**S**ession **I**nitiation **P**rotocol **U**ser **A**gent

SLA	**S**ervice **L**evel **A**greement
STU	**St**atistical **U**nit
SRVCC	**S**ingle **R**adio **V**oice **C**all **C**ontinuity
UDP	**U**ser **D**atagram **P**rotoco
UE	**U**ser **E**quipment
UICC	**U**niversal **I**ntegrated **C**ircuit **C**ard
UL	**U**p **L**ink
UMTS	**U**niversal **M**obile **T**elecommunication **S**ystem
URI	**U**niform **R**esource **I**dentifier
USIM	**U**MTS **S**ubscriber **I**dentity **M**odule
VLRU	**V**isitor **L**ocation **R**egister **U**nit
VoIP	**V**oice **o**ver **IP**
VoLTE	**V**oice **o**ver **LTE**
VoWiFi	**V**oice **o**ver **WiFi**
VM	**V**irtual **M**achine
VNF	**V**irtual **N**etwork **F**unction
VSA	**V**irtual **S**torage **A**ppliances
WB	**W**ide **B**and
WLA	**W**ork **L**evel **A**greement

Symbols

CC_{Inc}	Clear Codes of Failed Calls
$CCGroup$	Clear Code Group
$CCNameG$	Clear Code name belonging to Group
CC_{total}	Total Clear Codes of period of the Incident
CFS	Customer-Facing Service
$Growth$	Prediction of Traffic Growth
SR	Success Rate
$Refer$	*ReferentTable* function
Rf	*RefreshTable* function
T_{Inc}	Total time in minutes of the Year
T_w	Time Window threshold
W_{Inc}	Weight of Incident
$YCFS$	Yearly Average Customer-Facing Service

Chapter 1

Introduction

1.1 Context of the Study

When mobile cellular communication networks appeared, voice service was the main service available, still being a widely used service. Today, we have a new paradigm where data traffic is massively growing in mobile networks, and all new deployments are being planned with the growth of data in mind. Also, from standards point of view, a data-driven approach has been selected for future mobile networks. The IP technology and mobile services build on top of the data layer are essential pillars in this approach [1].

Following this growth in popularity of the use of data, applications capable of providing Voice over IP (VoIP) services started to appear and quickly became the main mean of communication between subscribers [2]. This Over-the-Top (OTT) threat openly caused impact in Voice Service use overall, mobile operators were forced to look for a competitive service to provide a valid option to their subscribers [3].

With the launch of Long Term Evolution (LTE), mobile operators have a VoIP service available to be used. Voice over LTE (VoLTE) provides three main advantages comparing with circuit-switched legacy and the OTT services. It provides better Quality of Service (QoS), since featuring an Integrated Multimedia System

(IMS) that provides End-to-End (E2E) QoS [4]. Provides High Definition (HD) and higher voice quality comparing both with 3G and OTT due to support modulation codecs like AMR-NB, AMR-WB and EVS [5]. Also, it is more secure than the OTT alternatives, because it does not require any connectivity with external servers of the home network, besides being able to support IP security (IPsec).

The Telephony Application Server (TAS) is an element of the IMS Core that is responsible for providing support for a minimum set of mandatory Multimedia Telephony (MMTel) services as defined by 3GPP, such as supplementary service functionality but it is also responsible for the call continuity even in inter-working scenarios between different access technologies Session Call Continuity Application Server (SCC-AS). Due to the fact of being placed in the back-end of the VoLTE service, TAS server is affected by all kinds of issues possible in a network. To prevent outage or service impact, forecast and constant monitoring are required.

1.2 Motivation

Each TAS can be provisioned to handle subscribers in the order of millions (depending on the model and vendor), and it needs to be able to serve traffic continually without any margin for an outage of service. Due to this enormous demand for readiness, it is required to be monitored, expanded and maintained carefully by the responsible support. Consequently, there is a need to develop methods and algorithms to monitor, warn and forecast possible issues with the element.

1.3 Objectives and Contributions

This study has the final goal provide insight over methods and algorithms that allow engineers to monitor and predict possible future issues in TAS on VoLTE service. These functionalities should be designed in a way to be agnostic to which environment they are implemented. However, for this research, the solution was developed in a *Cloud* environment.

The main goal is the develop a method based on Clear Codes (CC), Traffic Behaviour, Key Performance Indicators (KPI) and Alarms. Based on traffic profile analysis and status of the terminating calls, it is possible to evaluate the health of the system.

The mobile operators in which this study was performed will remain anonymous to avoid unfair competition.

1.4 Overview of the Thesis

1. First Chapter - **Introduction** - Presents the context of the study, the motivation for there choice and objectives. It also includes the organization of the document.

2. Second Chapter - **State of the Art** - It contains an overview of the concept of LTE and VoLTE. It includes a functional explanation of all the elements present in a VoLTE Network. It shows in detail the functionalities of the oTAS, the Telephony Application Server used for this study. Besides, it is also given a syntax about Key Performance Indicators, Work Level Agreements and *Cloud* infrastructure based on Openstack.

3. The Third Chapter - **Scenario Description** - It describes the "issue found" which this study intends to mitigate. Also, it shows the structure of the oTAS inside the *Cloud* and includes an explanation of the *Supervisor Virtual Machine*.

4. The Fourth Chapter - **Result Analysis** - This Chapter contains the pre-requisites determined to implement the algorithm set in chapter 3 and an analysis of the results. Also includes some observations, in the implementation, with possibility for improvement.

5. The Fifth Chapter - **Conclusion** - In this Chapter, we can find the conclusion of this study and the suggested topics for future study.

Chapter 2

State of the Art

This chapter contains an overall description of the VoLTE service in the LTE Network. It presents an overview of all the elements involved and their functionalities. It includes a description of the Clear Code Mapping and the benefits of using *Cloud* Computing in the VoLTE service.

2.1 Long Term Evolution

Long Term Evolution, also known as the Fourth Generation (4G) Technology, is an iteration of the Global System for Mobile Communication (GSM) and Universal Mobile Telecommunications System (UMTS). However, in contrast with previous Circuit-Switch (CS) model cellular systems, LTE has been designed to support only Packet-Switch (PS) services.

The Release 10, also known as LTE-Advanced (LTE-A) launched with three fundamental features, which are necessary to fulfil the throughput objectives proposed by the standards of peak data rates 100 Mbps in downlink and 50 Mbps in uplink:

- **Robust Modulation in Dense Environments** - Orthogonal Frequency Division Multiple Access (OFDMA) is used in the downlink to increase spectral and Single Carrier - Frequency Division Multiple Access (SC-FDMA) in uplink due to cost-efficient reasons. The usage of different modulated signals results in high Peak-to-Average-Power Ratio (PAPR), which results in expensive and complex UE transmitters.

- **Increased Link Capacity** - Carrier Aggregation and Multi-Input-Multi-Output (MIMO) (capable of Spatial multiplexing up to 8 layers in downlink and up to 4 layers uplink) allow for high bit rate and capacity of the network compared with previous technologies.

- **Flat IP** - Flat architecture with a backhaul based on IP and the separation of the user plane and control plane creating a scalable system (Figure 2.1).

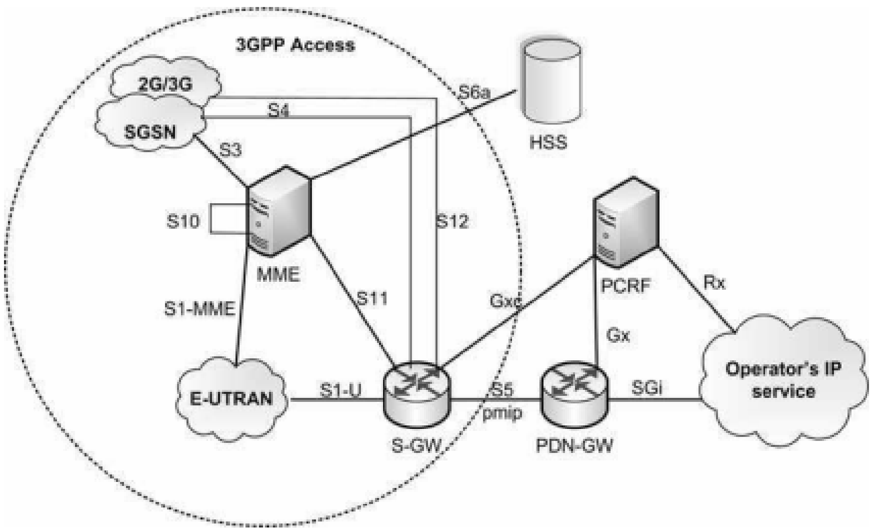

FIGURE 2.1: LTE reference model [6]

2.2 Voice over Long Term Evolution

The VoLTE solution is a standards-compliant implementation of LTE service architecture. The solution places IMS functional elements at the centre of the voice core network, managing the connectivity between subscribers and the implementation of policy control [7].

The VoLTE solution is scalable and rapidly deployed, offering rich multimedia and voice services, seamless voice continuity across access networks and the re-use of existing business and operational support components. The VoLTE solution builds upon the One Voice initiative, subsequently reformulated in GSMA Permanent Reference Document (PRD) IR.92 [8], IR.94 [9] and IR.64 [10], empowering operators to roll-out productive communications services and capabilities across both LTE and hybrid networks.

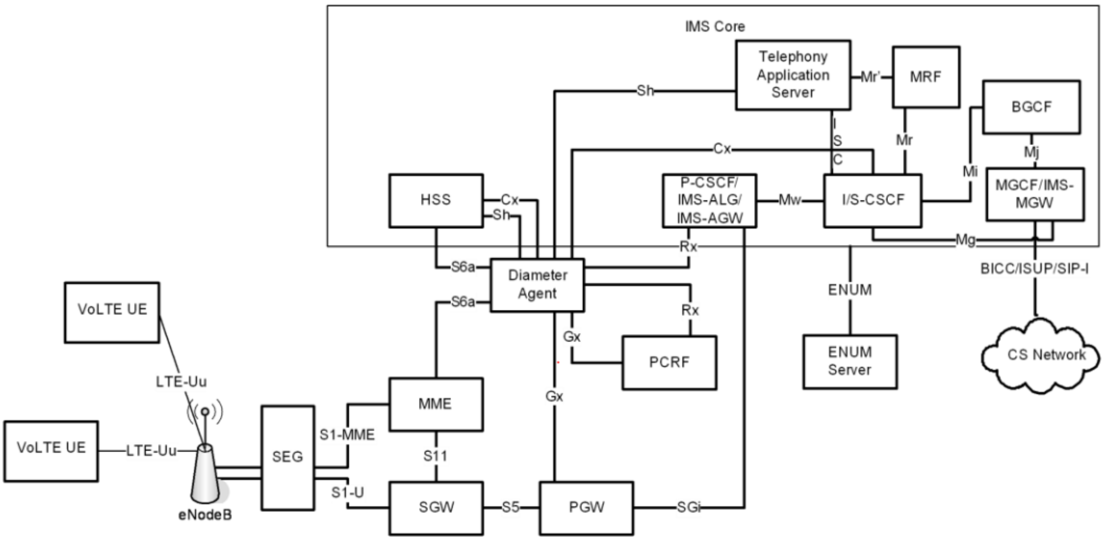

FIGURE 2.2: VoLTE Architecture [8]

The main functional nodes of the VoLTE architecture are defined by 3GPP and are described below according to the 3GPP TS 23.002 [11].

2.2.1 User Equipment

The User Equipment (UE) is used to connect to the Evolved Packet Core (EPC) via the LTE-Uu radio interface. Each UE must contain one Universal Integrated Circuit Card (UICC) and Session Initiation Protocol User Agent (SIP UA).

The most used combination of UICCs in a UE is a UMTS Subscriber Identity Module (USIM) and IP Multimedia Services Identity Module (ISIM). Focusing on

ISIM which is needed to perform registration on IMS core, it contains according to 3GPP TS 31.103 [12]:

- IP Multimedia Private Identity (IMPI) - It is a global identity allocated by the home network. IMPI contains home operators domain information.
- IP Multimedia Public Identity (IMPU) - Similar to a telephone number, it is used by any user for requesting communications to other users. It can either be a SIP URI or a tel URI [13].
- Secret Key - This long secret key is used for user authentication and SIP registration.

2.2.2 eNodeB

The eNodeB is the single node that interfaces with the UE. It is responsible for establishing the Physical (PHY), Medium Access Control (MAC), Radio Link Control (RLC), and Packet Data Convergence Protocol (PDCP) layers that include the functionality of user-plane, header-compression and encryption. It also offers Radio Resource Control (RRC) functionality corresponding to the control plane. It performs many functions including radio resource management, admission control, scheduling, enforcement of negotiated UL QoS, cell information broadcast, ciphering/deciphering of user and control plane data, and compression/decompression of DL/UL user plane packet headers.

2.2.3 Evolved Packet Core

2.2.3.1 Mobility Management Entity

The Mobility Management Entity (MME) handles the signalling and control, mobility and idle management manipulating the distribution of paging messages to the eNodeB. This facilitates optimisation of the implemented networks and allows total flexibility in expanding the capacity.

It also manages the access of the UE to the network through interaction with the Home Subscriber Server (HSS) to authenticate users. Provides the function of the control plane to enable seamless mobility between LTE and 2G/3G mobile networks and also supports lawful interception of signalling.

2.2.3.2 Serving Gateway

The Serving Gateway (S-GW) acts as the termination point between the radio access network and the Core Network. Forwards data packets to and from the eNodeB and the Packet Data Network Gateway (P-GW) and performs accounting and control of user data.

It also serves as a local mobility anchor for handovers between eNodeBs or passage between 3GPP networks and informs the user traffic in case of lawful interception.

2.2.3.3 Packet Data Network Gateway

The PDN gateway provides the connection between EPC and several external data networks. It serves as a point of entry and exit of data traffic from the UE and interface between LTE networks and packet data networks such as the Internet or fixed and mobile networks based on the Session Initiation Protocol (SIP) or Internet Protocol multimedia subsystem (IMS).

It also manages the assignment of IP addresses and supports packet filtering for each user. Also supports charging and serves as a local mobility anchor for handovers between 3GPP access systems and non-3GPP access systems.

2.2.3.4 Home Subscriber Server

The HSS is a network database that holds both static and dynamic data elements related to subscribers. The HSS provides user profile information to the MME, and IMS core during UE attach and IMS registration.

2.2.3.5 Policy Charging and Rules Function

The Policy Charging and Rules Function (PCRF) provides policy control decisions and flow-based charging controls. The PCRF determines how a service data flow shall be treated in the enforcement function and ensure that the user plane traffic mapping and treatment is following the user's profile.

2.2.4 IP Multimedia Subsystem

IMS is the control infrastructure for supporting next-generation IP Multimedia Services and consists of many separate elements which are listed below.

2.2.4.1 Proxy Call Session Control Function

The Proxy Call Session Control Function (P-CSCF) is the initial point of contact for session signalling for the IMS-enabled VoLTE UE.

The P-CSCF behaves as a SIP proxy by forwarding SIP messages between the UE and the IMS Core Network, maintains the security associations between itself and the VoLTE UE, and incorporates the Application Function aspect of Policy Charging and Control (PCC) to enable binding of the IMS session with the bearer for applying dynamic policy and receiving notifications of bearer level events. (A bearer is virtual concept used to define a connection E2E for a specific QoS. In this case, the bearer defined is the IMS bearer).

2.2.4.2 Interrogating Call Session Control Function

The Interrogating Call Session Control Function (I-CSCF) is the contact point within an operator's network for all connections destined to a user of that network.

On IMS registration, it interrogates the HSS to determine which suitable S-CSCF to route the request for registration. For mobile terminating calls, it interrogates the HSS to determine which S-CSCF the user is registered on.

2.2.4.3 Serving Call Session Control Function

The S-CSCF provides session set-up, tear-down and control including routing functions. It generates records for billing purposes for all sessions under its control and invokes Application Servers based on IFCs received from the HSS. The S-CSCF acts as SIP registrar for VoLTE UEs that the HSS and I-CSCF assign to it. It queries the HSS for the applicable subscriber profiles and handles calls involving these endpoints once they have been registered.

2.2.4.4 Media Resource Function

The MRF is a common media resource function, for use by IMS Application Servers and I/SCSCFs, to provide media plane processing independent of application types, e.g. transcoding, multiparty conferencing, network announcements, tones, etc. under the control of IMS Application Servers (VoLTE AS) as well as basic media processing functions to CSCFs. The control plane interfaces to MRFs are defined by the 3GPP references Mr, Mr' and Cr interfaces (SIP/SDP and XML encoded media service requests) while the media plane interfaces to MRFs are defined by 3GPP reference Mb for RTP/RTCP transport.

2.2.4.5 Interconnection Border Control Function/Transition Gateway

The Interconnection Border Control Function/Transition Gateway (IBCF/TrGW) is responsible for the control/media plane at the network interconnect point to other Public Mobile Networks (PMN).

2.2.4.6 IMS Application Level Gateway/IMS Access Gateway

The IMS Application Level Gateway/IMS Access Gateway (IMS-ALG/IMS-AGW) is not a stand-alone function but is located with the P-CSCF.

The IMS-ALG/IMS-AGW is responsible for the control/media plane at the access point to the IMS network.

2.2.4.7 Media Gateway Control Function/IMS Media Gateway

The Media Gateway Control Function/IMS Media Gateway (MGCF/IMS-MGW) is responsible for the control/media plane interworking at the network interconnect point to circuit-switched networks.

2.2.4.8 Breakout Gateway Control Function

The Breakout Gateway Control Function (BGCF) is responsible for determining the next hop for routing of SIP messages.

This determination is based on information received within the SIP/SDP and routing configuration data. For CS Domain terminations, the BGCF determines the network in which CS domain breakout is to occur and selects the appropriate MGCF.

2.2.4.9 Telephony Application Server

The TAS is an IMS Application Server providing support for a minimum set of mandatory Multimedia Telephony (MMTel) services as defined by 3GPP.

2.3 Open Telephony Application Server

The open Telephony Application Server (oTAS) is the Nokia solution for Voice over WiFi(VoWiFi) and VoLTE upon the One Voice initiative, empowering operators to roll-out rich communications services and capabilities across both LTE and hybrid networks.

The oTAS offers a flexible, robust, cost-optimised service delivery and market-defining enriched subscriber services in the virtualised and cloudified telecoms ecosystem.

The oTAS provides multiple standardised IMS application server roles. The oTAS software architecture also allows the single and separate invocation of the various roles. Figure 2.3 shows the oTAS functionalities and interfaces with elements within the network.

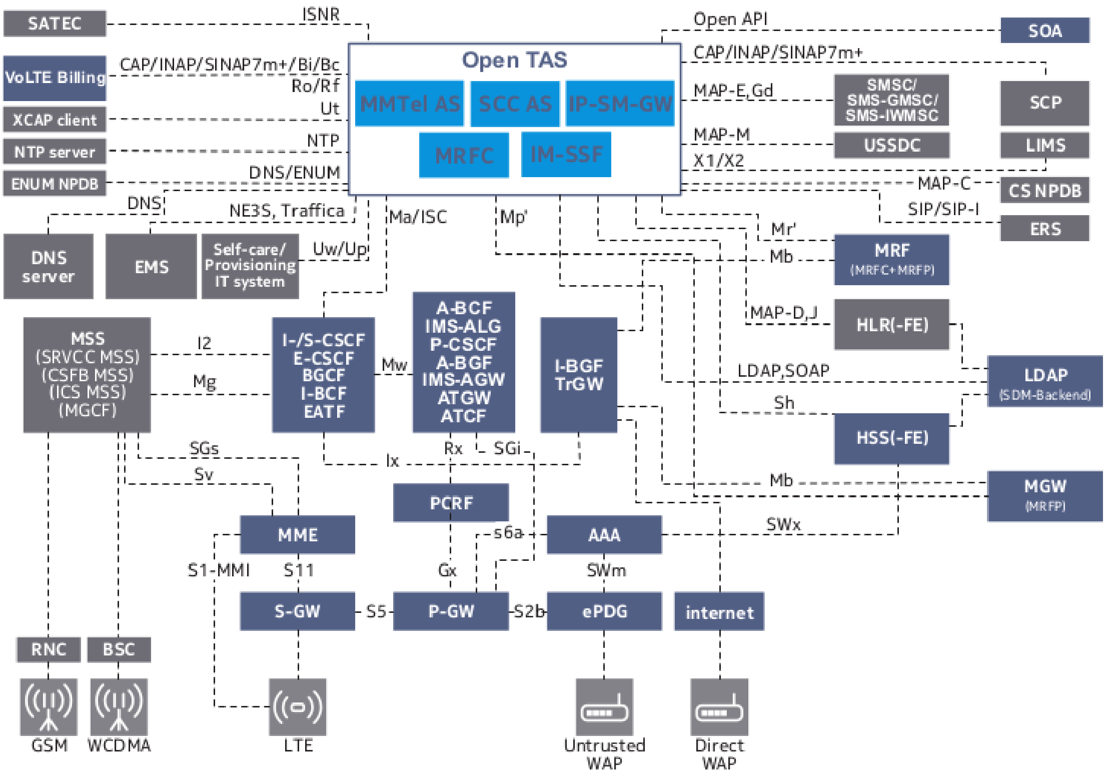

FIGURE 2.3: oTAS in VoWiFi and VoLTE Deployments[14]

2.3.1 MMTel AS

The Multimedia Telephony Application Server (MMTel AS) role manages IMS based telco sessions for voice and video, incorporating and implementing the principles spelt out by GSMA PRD IR.92 [8] and IR.94 [9].

The MMTel AS is also responsible for the management of subscriber data and subscriber services. Thus, its principal peer network elements are the respective subscriber databases. It is the intelligent and flexible MMTel capabilities that allow the oTAS to support a variety of subscriber database models.

The MMTel AS supports both Sh connectivity to any HSS and MAP connectivity to any Home Location Register (HLR). The MMTel AS interworks with the oTAS Media Resource Function Controller (MRFC) role in providing tones and announcements for audio and video calls, and, conferencing services.

The MMTel AS triggers the integrated IP Multimedia Service Switching Function (IM-SSF) role to communicate with the Service Control Point (SCP) to provide Intelligent Network (IN) services. With the IM-SSF role, over the CAP/INAP interface to the SCP, the MMTel AS also invokes the online charging service to generate session and event-based online charging data. Simultaneously, offline charging capabilities are also supported[15].

End user-controlled service management is also executed in the MMTel AS. The MMTel AS's in-built XCAP Server function allows subscriber service modification to a centralised repository from Ut/XCAP capable terminals.

The oTAS MMTel AS is involved in executing a bundle of services, including supplementary, network and regulatory services:

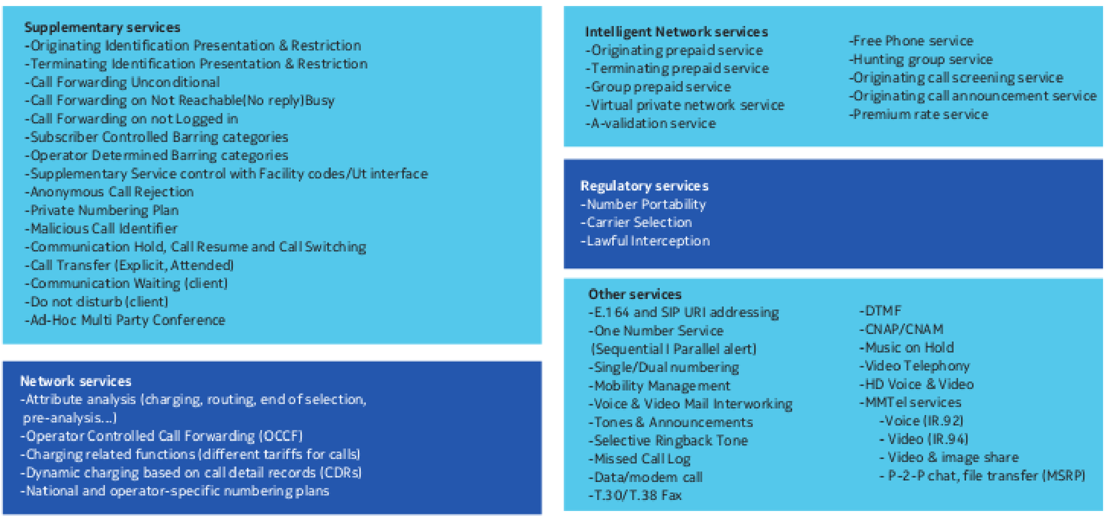

FIGURE 2.4: oTAS Supported Services[14]

2.3.2 MRFC

The oTAS MRFC manages the media components of IMS sessions, controlling the signalling mechanisms, which underlie and build up the user plane layer of a session [16].

The oTAS MRFC complements the oTAS MMTel AS role. If telephony sessions involve supplementary and IN services, or, require transcoding, the MRFC function is mandatory. It is Complying with relevant 3GPP specifications and IR.94 guidelines, the oTAS enables feature parity and service continuity across multiple access domains, by providing signalling control for bearer related media services for audio and multimedia calls and conferencing [9].

Besides assisting the MMTel AS in basic call scenarios, and calls involving base supplementary services, the oTAS MRFC role is also invoked by the IM-SSF, during call sessions involving the IN platform, whereas the MRFC provides, for example, customised and flexible tone and announcement services [8].

The MRFC complies with IR.94 guidelines by providing support, and also transparent forwarding support, for the specified feature tags, media flow direction and SDP content[9]. If required, it also provides transcoding support. The oTAS provides codec support (AMR, AMR-WB, G.711, G.723.1, G.729A/B, iLBC, GSM EFR, GSM FR, GSM HR) over the Mp' (H.248) interface [17].

2.3.3 SCC AS

To introduce VoLTE, operators have the option to implement a full-blown IMS network roll-out or choose a phased strategy to VoLTE introduction to the network. Irrespective of the approach, VoLTE network interworking with 2G/3G networks is a must. To manage continuous services across LTE and CS access domains, and centralise service execution in the IMS, 3GPP defined the Service Centralization and Continuity Application Server (SCC AS) role. The oTAS SCC AS role follows and fulfils the principles to offer IMS Centralized Services (ICS), as outlined in

GSMA PRD IR.64, and enables service continuity and service centralisation, as specified respectively in 3GPP TS 24.237 and 3GPP TS 24.292.

The oTAS SCC AS integrates the Session Continuity role to enable Single Radio Voice Call Continuity (SRVCC) in VoLTE networks. The oTAS solutions adapt to both Release 8 and Release 10 specified SRVCC solutions and network architecture. The oTAS SCC AS complies with standards in carrying out session anchoring and session transfer tasks in the home network, as well as, manages registration and session set-up task according to the particular requirements of either network models.

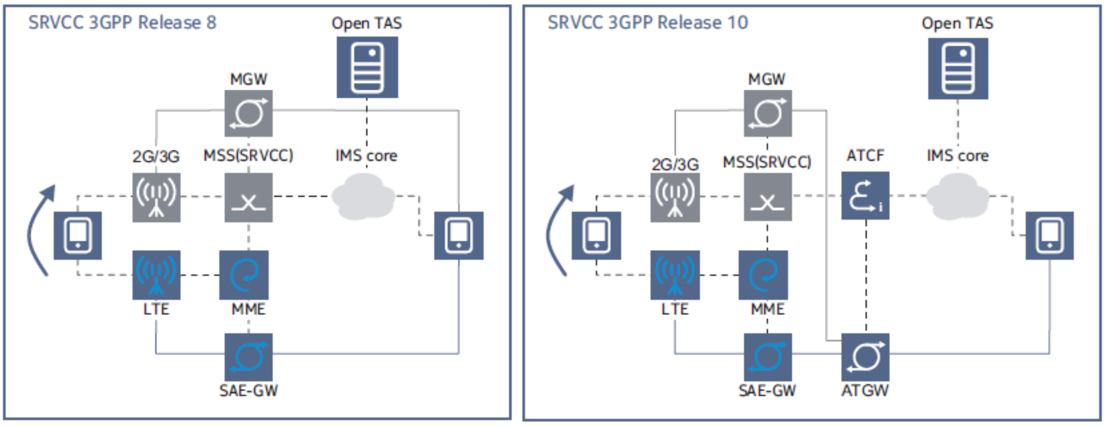

FIGURE 2.5: SRVCC Release 8 and SVCC Release 10[14]

Also, respectively, in Release 10 SRVCC architecture, with HSS support, the oTAS SCC AS supports dynamic Session Transfer Number for Single Radio (STN-SR) allocation, and, enables SRVCC based on the terminals SRVCC capability. This capability is provided with and without ATCF/ATGW deployment. The service continuity offering of the oTAS is complemented with the support of mid-call services, and, the oTAS innovation for voice quality optimisation, during SRVCC.

3GPP defines the Terminating Access Domain Selection (T-ADS) function for SCC AS. With its HSS support, complying with Release 9 specifications, the oTAS performs optimised call routing in an access conscious fashion, based on the called party T-ADS information that is, based on the called party actual radio access type and availability, as well as, based on whether IMS PS voice is supported at the called party location.

Beside the Release 9 Network-based T-ADS solution, the oTAS SCC AS also offers the Release 8 Sequential T-ADS solution for those operators who are yet to invest in HSS integration in their network, but wish to experiment with the T-ADS solution.

2.3.4 IM-SSF

The IM-SSF complements the oTAS functions with the same sophisticated service package in the IMS environment as to how Intelligent Network (IN) platforms improve CS core services. The oTAS IM-SSF role replicates and uses the CS logic, providing a solid foundation to introduce new services over proven interfaces and protocols.

The oTAS supports and implements the principles specified in 3GPP TS 23.278, and, adjusts them to the CAP interface (3GPP TS 29.078) of the CS domain. With the Configurable bearer capability for video calls functionality, the IM-SSF logic enables the simulation and the handling of the video calls as if they were voice calls, including providing the same tariff for the voice and video calls, without customisation in the IN service logic. Not strictly an IM-SSF capability but much rather to complementing the IM-SSF logic, the oTAS is capable of providing PS subscriber state information to the SCP, similarly to as it is available for CS subscribers in the CS domain. The oTAS logic is extended with the capability to respond to the SCP's MAP Provide SubscriberInfo (MAP PSI) request and provide PS subscriber state information.

2.3.5 IP-SM-GW

The IP-SM-GW role of the Open TAS delivers the SMS over IP solution, as specified by 3GPP. However, for this implementation, the service focus is on Voice Service.

2.4 Clear Codes

Clearing of a call can happen for numerous different reasons. When a call is cleared, a cause code is set to indicate why it was cleared. Since different protocols define their cause code sets resulting in different cause code sets, and also there can be internal reasons for clearing a call, not only external cause codes are defined by the protocols, but internal clear codes also.

To keep the information provided by a cause/clear code throughout the chain of network elements involved in the call, mappings should happen between the different sets of codes.

The mapping is primarily influenced by configuration settings and the activation of given features. External causes are mapped to internal causes which are then mapped to external causes again by the oTAS when the incoming and outgoing sides configuration differ, or call control modifies the internal clear code.

Signalling-specific Clear Code Reports belong to the field reporting reports which provide the operator with the following information about clear codes:

- Number of calls in the exchange;
- Clear codes of all calls;
- Clear codes of one outgoing number destination;
- Signaling-specific clear codes of calls.

The Session Initiation Protocol (SIP) in the oTAS feature divides specific, clear codes into three main parts based on the transfer protocol, so the measurement provides information on these as well:

- SIP over User Datagram Protocol (SIP/UDP);
- SIP over Transmission Control Protocol (SIP/TCP);

- SIP over Stream Control Transmission Protocol (SIP/SCTP).

The signalling-specific clear code measurement is enhanced with the possibility of setting new object types to the measurement for the SIP [18], Bearer Independent Call Control (BICC) [19], and Radio Access Network Application Part (RANAP) [20] protocol types, and it can collect data about one selected object in case of these new signalling types.

2.5 Key Performance Indicators

For performance and monitoring of a Service, it's required to design and implement metrics and counters to quantify the performance of an element or system.

KPIs are predefined statics formulas created by the operator/vendor to correlate measurements. Measurements are a function where the system collects information about the traffic and network events in the application server and then processes this information.

KPIs provides operators with real-time data on the operation, capacity, and service level of the service.

oTAS is able of providing detail measurements according to the following division:

- Overloaded parts in the system
- Faulty parts in the system
- Traffic Intensity
- Traffic Categories
- Subscriber Mobility
- Application Server Status

- Capacity

While focusing in the VoLTE service, the oTAS calculates the internal VoLTE Call Success Rate according to the CC Groups:

$$oTAS\_VoLTE\_SR = \frac{CCGroup1 + CCGroup3 + CCGroup4}{CCGroup1 + CCGroup2 + CCGroup3 + CCGroup4} \tag{2.1}$$

The Clear codes are grouped into four classes:

- Group 1 - 0000H - 00FFH: normal clearing normal clearing situations, for example, subscriber busy.
- Group 2 - 0100H - 01FFH: internal congestion clear codes set by charging and statistics and not leading to a release.
- Group 3 - 0200H - 02FFH: external congestion normal clearing situations caused by congestion outside the exchange, for example, by congestion on the radio path.
- Group 4 - 0300H - 03FFH: subscriber errors normal congestion caused by the subscriber's facility or equipment.

2.6 Contracts Agreements

Previously, it was mention that each TAS is planned and implemented to serve a few millions of subscribers. This combined with the reduced number of units in a network, usually one or two data centres (for Geo-Redundancy) creates an enormous demand for reliability and availability. Due to this, it is required to develop several processes and methodologies to ensure that the service is available on permanent bases [21].

To maintain and operate these elements, like any other element in the network, Mobile Operators require specialised teams (internals or subcontractors) to operate them.

Depending on the company policies and agreements, the process of service delivery can differ. However, for simplicity and objectivity, we will only mention the two main agreements that need to be done between the Mobile Operator and the Operational Team [22]:

2.6.1 Working Level Agreement

Working Level Agreement (WLA), also known as Operation Level Agreement (OLA), it contains a detailed description of involved parties, responsibilities and roles. It should contain instructions on how to monitor the system and access it. This document must have a classification by priority, of the different levels of severity and their escalation matrix to follow in case of an Outage (Service Unavailability). This document works as a support document for the SLA.

2.6.2 Service Level Agreement

Service Level Agreement (SLA), it is a commitment between a service provider and a client to assure the delivery of a determinate aspect of the service. It is required to define a metric to prove that the commitment is being fulfilled. For this reason, a customised KPI is created, and his formula is agreed between all parts.

In chapter 3.1.1, it is described the specific SLA limitation which also triggers the development of this study.

2.7 Virtualization

Virtualisation and *Cloud* are terms common in our days. However, it is imperative to understand the true impact of these technologies on the day-to-day business and on the evolution of technology way forward.

2.7.1 Benefits of Virtualization

Deploying a service on the *Cloud* using virtualisation bring key benefits for the provider, such as [23][24]:

- **Hardware Independence** Virtualization abstracts the hardware (HW). This abstraction simplifies life-cycle management because HW and Virtual Network Function (VNF) are decoupled and can be life-cycled independently, allowing re-use of data-centre for other network services.
- **Low entry threshold** the ability to run all VNFs for service on the same hardware and being able to chop the available hardware into smaller Virtual Machines. This means that you can install a complete small VoLTE Core Network with several VNFs on hardware, that previously would have served only a single node.
- **Enabling Automation** This means we can automate the deployment of solutions consisting of several network functions. Once a business starts to grow it can be scaled in response to the demand without having to re-implement it or to pre-order additional hardware.
- **OnDemand Scaling** Based on these mechanisms the same HW-resource can now be used to scale the capacity of any workloads. One spare server can be used to expand the capacity of any service. This implies that we completely remove several months of traffic-forecasting per network function required by the ordering and deployment cycle of special hardware.
- **Business Agility** Cost-efficient networks, in terms of investment and maintenance, allow for every business to run their network. That increases the flexibility, as businesses can evolve independently only driven by their priorities. Suddenly business development activities can introduce new features on their network and do not have to wait for the consumer services to introduce them.

2.7.2 OpenStack

OpenStack is a set of software tools for building and managing *Cloud* computing platforms for public and private clouds. Backed by some of the biggest companies in software development and hosting, as well as thousands of individual community members, many think that OpenStack is the future of cloud computing [25].

2.7.2.1 *Cloud* Computing

The *Cloud* is all about providing computing for end-users in a remote environment, where the actual software runs as a service on reliable and scalable servers rather than on each end-user computer.

OpenStack falls into the latter category and is considered Infrastructure as a Service (IaaS). Providing infrastructure means that OpenStack makes it easy for users to quickly add new instance, upon which other *Cloud* components can run.

2.7.2.2 Components of Openstack

- **Nova** is the main computing engine behind OpenStack. It is used for deploying, manage the VMs and execute computing functions.
- **Swift** is a storage system for objects and files. It uses a unique identifier to a file let OpenStack decide where to store this information instead of the traditional allocation of files in a disk drive. This process makes scaling easier.
- **Cinder** is a block storage component, which is more analogous to the traditional notion of a computer being able to access specific locations on a disk drive. The block storage system handles the creation, attaching and detaching of the block devices to servers.

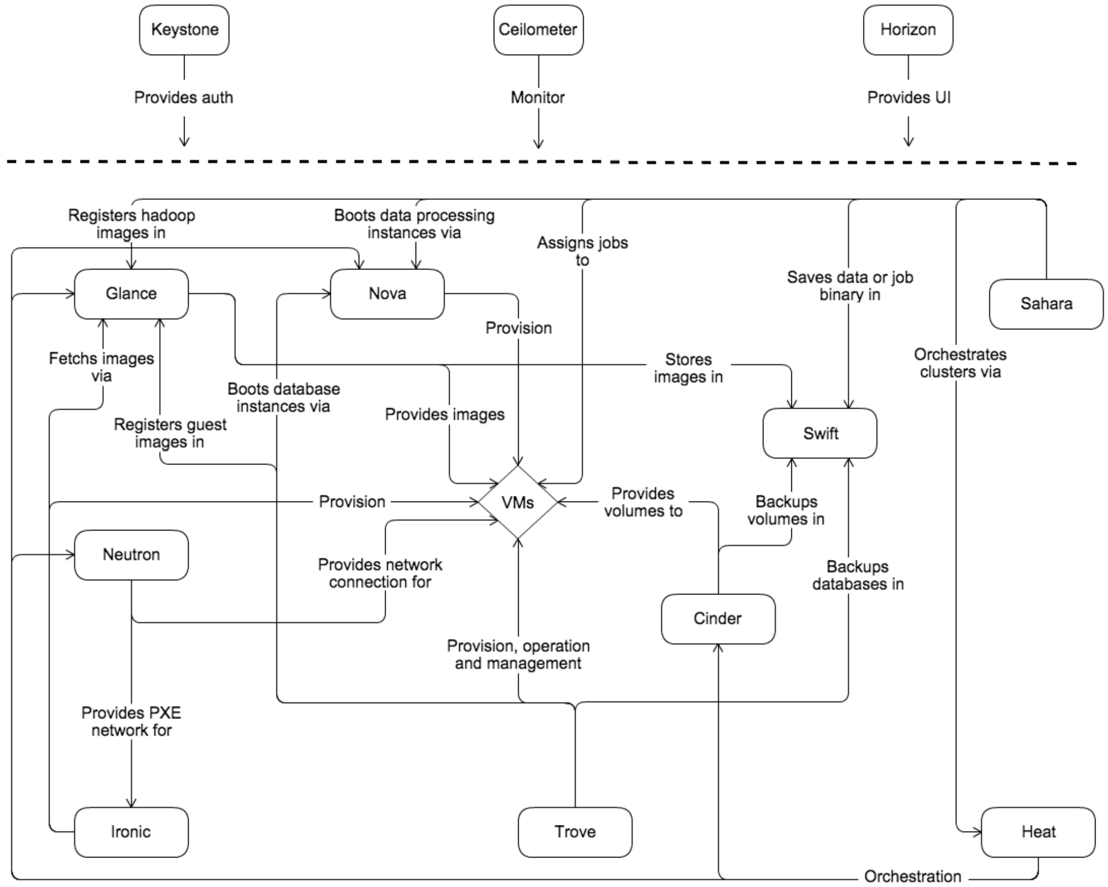

Figure 2.6: Openstack Arquitecture [26]

- **Neutron** implements the networking capability for OpenStack. It serves to guarantee that each of the components of an OpenStack deployment can communicate with one another quickly and efficiently.

- **Horizon** is the dashboard behind OpenStack. It is the only graphical interface to OpenStack. All of the components can be accessed through an application programming interface (API). Although, the dashboard groups all the OM tool in one place.

- **Keystone** provides identity services for OpenStack. It is a list of all of the users of the *Cloud*, with there services and permissions mapped.

- **Glance** provides image services to OpenStack. It is responsible for the storage of the images templates, which can be used to deploying new virtual machine instances.

- **Ceilometer** provides telemetry services. It provides a contact for billing systems, providing all the counters they need to establish billing services to individual users of the *Cloud*.

- **Heat** is the orchestration component of OpenStack, which enables developers to store the requirements of a *Cloud* application in a file that determines what resources are necessary for that application.

- **Ironic** is an OpenStack function which provisions bare metal instead of virtual machines. It may be used independently or as part of a *Cloud*.

- **Trove** is Database as a Service for OpenStack. It's designed to run entirely on OpenStack, to empower users to instantly and easily use the features of a relational database.

- **Sahara** gives users a simple way to provision clusters. After a user chooses the initial parameters required, Sahara deploys the cluster in a few minutes. Also, it provides the tools to scale an already provisioned cluster by adding or removing worker nodes on demand.

Chapter 3

Scenario Description

In this chapter, it is described the implementation environment, the issue in the study and the solution developed to solve it.

3.1 Case Study

3.1.1 SLA Limitations

For the Voice Service SLA, the KPI was created by using Clear Codes of the TAS at the end of each call. Its name is Customer-Facing Service (CFS), its formula depend on the weight of failed clear codes (a consequence of the outage) and the amount of time in which the Outage occurred:

$$CFS = 1 - \frac{W_{Inc}}{T_{year}} \quad \text{where} \quad W_{Inc} = \frac{CC_{Inc}}{CC_{total}} \tag{3.1}$$

Where:

- W_{Inc} is the ratio of outage created by the Incident;
- T_{year} is the total number of seconds in a Year.
- CC_{Inc} is the sum of all the failure *Clear Codes* created during incident;

- CC_{total} is the sum of all the *Clear Codes* that occurred during the period of the incident;

In case of an Outage, the Individual CFS is calculated for that particular incident. The YCFS is calculated by the sum of all the Individual CFSs.

$$YCFS = CFS(1) + CFS(2) + ... + CFS(n) \tag{3.2}$$

The threshold agreed on the SLA is *YCFS* equal to or bigger than 99.99%, which limits the maximum time of an outage is 52 minutes and 36 seconds.

3.1.2 Issue Found

It was found reliability in our operation section in the ability to detect an outage of service. In the past year, the interval of time between the occurrence of an incident and the detection of it was 25 minutes on average (Figure.3.1).

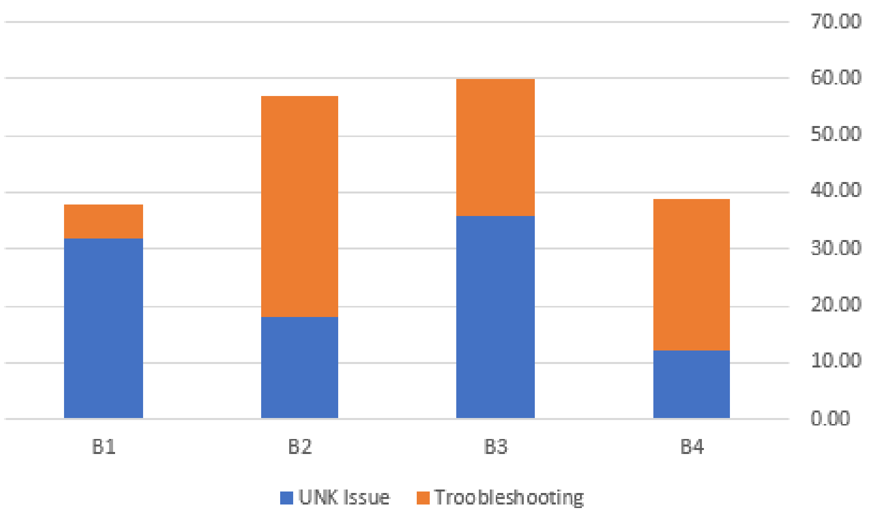

FIGURE 3.1: Real Time spent until the neutralization

The CFS is more dependent on the number of calls impacted by the incident, rather the Real-Time interval of it. Also, it is very important the reduction of the Real-Time interval.

3.2 System Architecture

In this scenario, the Telephony Application Server is deployed in a *Cloud* environment. In Figure 3.2 is described the logical architecture, the functional units and the interfaces of the oTAS in the *Cloud* environment.

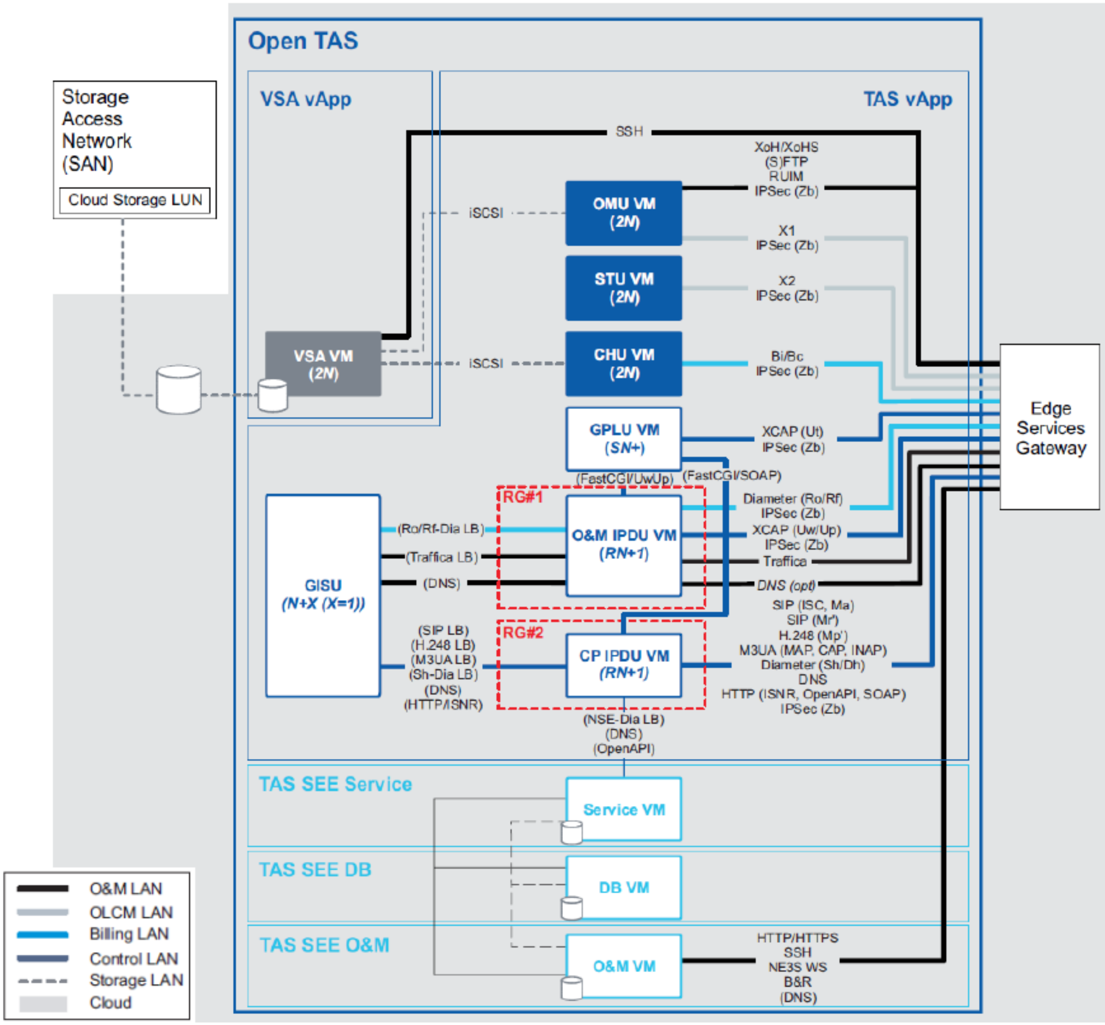

FIGURE 3.2: oTAS *Cloud* Network Architecture [27]

3.2.1 Functional Units

- **Operation and Maintenance Unit** (OMU) unit is dedicated storage devices, which serve as storage for, e.g., the entire system software of the network element, as well as, for the event buffer for intermediate storing of alarms. The OMU stores the uploaded SW packages on redundant storage devices. It incorporates the alarm system of the oTAS, which is responsible for collecting alarms from the distributed units and sending them to external OM systems. OMU is the central unit of recovery subsystem that supervises the functional unit states in the oTAS.

- **Central Memory and Marker** (CMM) serves as a centralized unit for providing such functions in routing and configuration management that could not be distributed. The CMM also stores all system configuration data and the master copies of distributed files: changes in the configuration are first written to the files in the CMM, then those are distributed, for example, to the signalling units. Also, the CMM is responsible for the central functions of signalling.

- **Statistical Unit** (STU) collects performance and measurement data from the network. It also takes care of the lawful interception reporting and IP connections related to that.

- **Charging Unit** (CHU) collects and stores charging data (CDR). CDRs are formed in the CHU memory, and based on the operator's decision they can be stored on local redundant HDD first and periodically transferred to Billing Center, or, they can be directly sent to the Billing Center when they are created. In the latter case, the local HDD can serve as fallback storage when the connection to the Billing Center fails.

- **Visitor Location Register Unit** (VLRU) is responsible for subscriber management in the oTAS includes the business logic related to the management of subscriber data, and, as such, maintains information about all subscribers currently roaming in the area of the oTAS.

- **Generic IP Signaling Unit** (GISU) is the main functional unit, which executes call and non-call related procedures in the oTAS. It is responsible for terminating the external signalling interfaces of the oTAS and processing the incoming requests.

- **IP Director Unit** (IPDU) serves as a load balancer for SIP, H.248, M3UA, Diameter and Traffica signalling in the oTAS, hiding the internal topology of the oTAS. The IPDU may also serve as an IP forwarder, for example, for DNS traffic.

- **General Purpose Linux Unit** (GPLU) hosts the XCAP server that handles XCAP requests over the Ut/Uw/Up interface. GPLUs are in load share redundancy model, so there is no additional unit. The GPLU is an optional unit, and, is only required if user-controlled supplementary service management is deployed in the network, and, the supplementary service manipulation is executed in the HSS over the Sh interface.

- **Virtual Storage Appliances** (VSA) The VSA connects the OMU pairs and CHU pairs with the shared disks of the *Cloud*.

  The Service Execution Environment (SEE) are extensions that oTAS contains to be able to inter-work with external elements as shown in Figure 3.3:

  – **Service VM** The service cluster hosts and executes the TAS SEE's service logic.

  – **Database VM** The DB cluster provides an in-memory database that stores service logic related data that is required by the service cluster to execute services.

  – **Operation and Maintenance VM** The Operation and Maintenance (OM) VM provides Fault Management and Performance Management services both locally and toward NetAct.

3.2.2 Cluster Architecture

The *Cloud* architecture selected for this deployment was the Cluster type. This type of installation presents advantages in reliability of the services *running* in the controllers and provides redundancy for the VMs (Figure 3.4) [28].

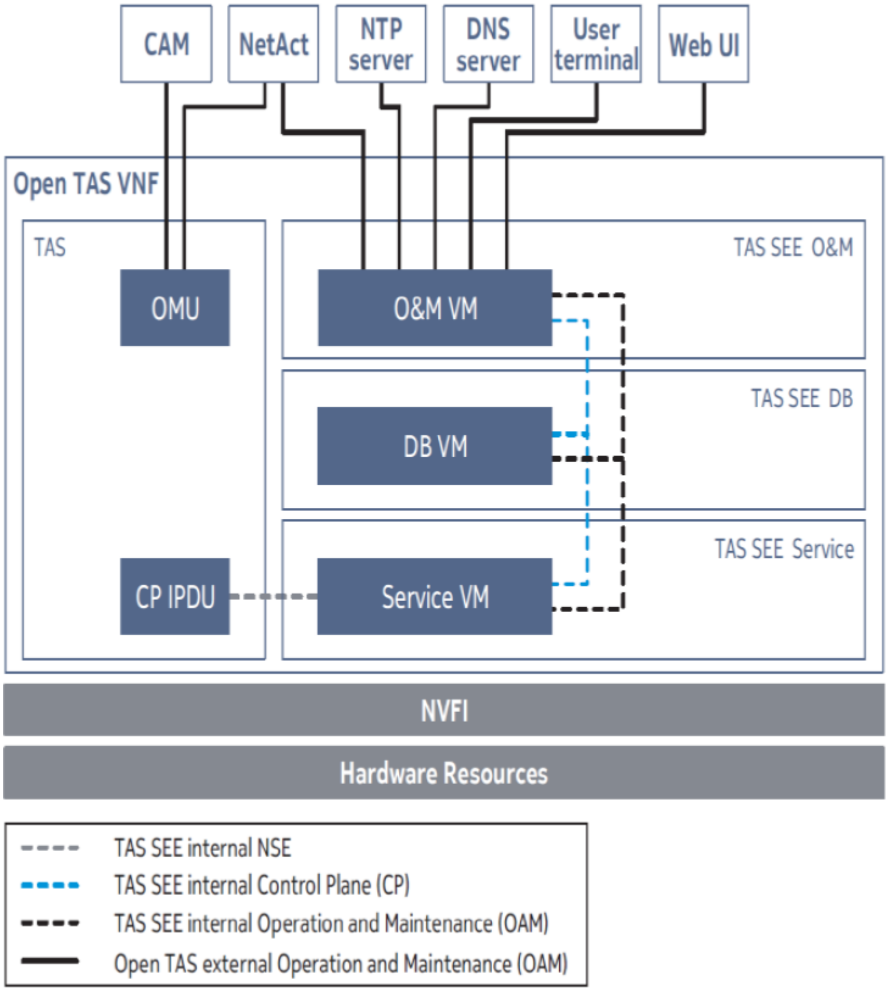

FIGURE 3.3: oTAS - Service Execution Environment [27]

3.3 Supervisor Virtual Machine

To mitigate the issue mentioned in section 3.1.2, a Supervisor VM was created and deployed in the *Cloud* environment. The VM is a Ubuntu 16.04.2 LTS image with SQL extensions.

The primary purpose of this tool is to reduce the time window for detection of the issue.

3.3.1 Extract the Clear Codes from oTAS

Within the Ubuntu VM, it was created a shell script file with the purpose of extract the Clear Code report from the oTAS and import it on the MySQL table.

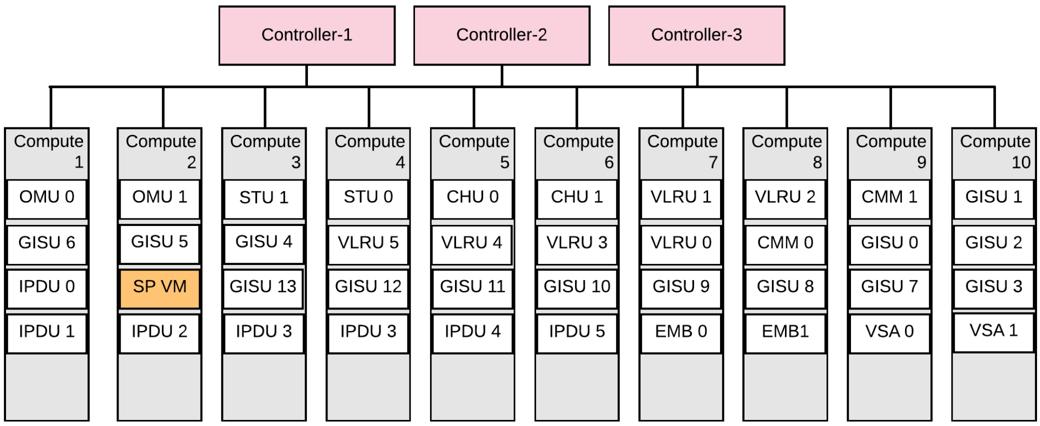

FIGURE 3.4: Cluster deployment

To be able to perform an SSH connection between the VM and the oTAS (through OMU unit), the VM is required the creation of a User Account, on the oTAS side, for this operation. For this, the user "CC_User" was created with the read rights on the oTAS.

In the Figure 3.5, it is possible to visualise the following stages:

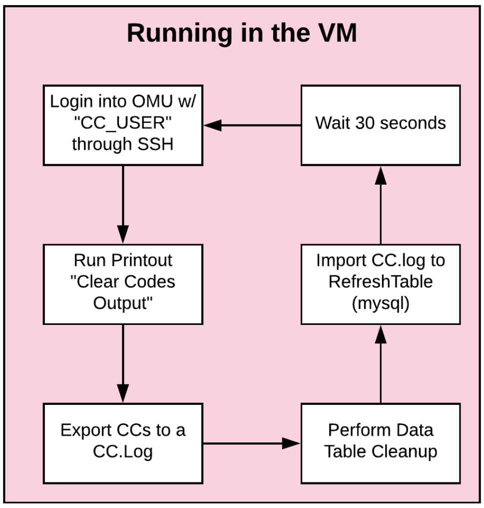

FIGURE 3.5: Algorithm to Extract CC Report from OMU

1. **Login on oTAS** - The SSH connectivity is permitted in the *Cloud* infrastructure towards the OMU unit for Operation and Maintenance purposes.

As long as the credentials of "CC_User" in use are valid and the account has access rights.

2. **Run Printout Command** - In this step, it is required to export the current CC Report generated at this point. The command and the layout of the output may vary depending on the vendor. For Nokia oTAS, this command is *ZTUT:CLR;*

3. **Export the CC report to a Log** - Save the printout of the CC report in a file.

4. **Perform Data Table Cleanup** - In this stage, it needs to organize the data to be uploaded to the table. The creation of columns and rows is performed at this step, to match the MySQL template.

5. **Import of CC Log Table to the MySQL** - The organized table is imported to the MySQL, in the table *RefreshTable*. This table is used in the next stages of the Algorithm.

6. **Repeat Timer - 30 seconds** - The Repeat timer works as a *wait timer*, to repeat the cycle. For this implementation, it is fixed at 30 seconds. However, this time can be changed or tuned.

3.3.2 Cross Check CCs in ReferenceTable

After the *RefreshTable* is been updated, it is required to perform a crosscheck with the *ReferenceTable*. This is the previous CC Table exported 30 seconds prior (in this scenario).

This process is needed because the List of CCs possible is 4096, which only a small amount is in use constantly and it is impossible to predict the appearance of a new CC. Figure 3.6 presents this algorithm.

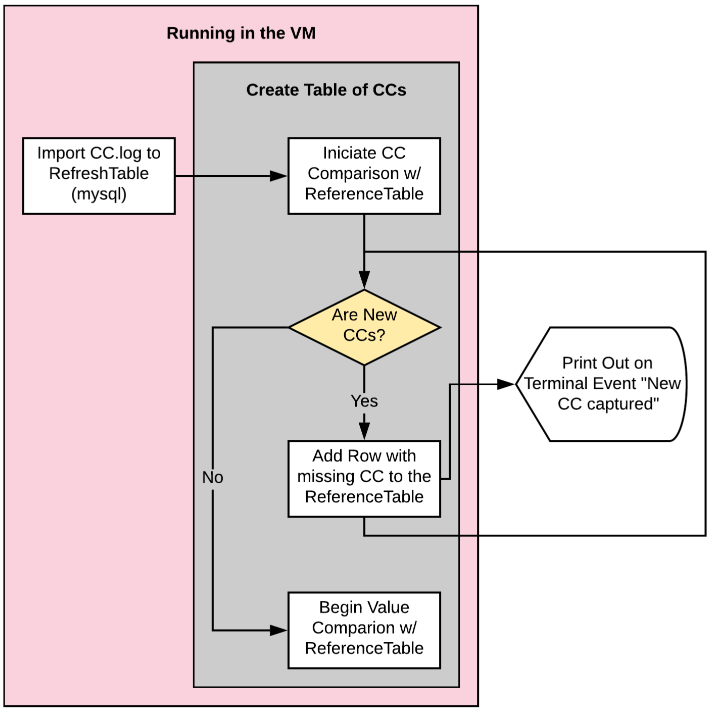

FIGURE 3.6: Function to add "New CCs to the ReferentTable"

1. **Initiate Comparison with Previous Reference** - At this moment, the goal is to find if there is a CC present on the *Refreshtable* which is not present on the *ReferenceTable*.

2. **Distinguish Existing CCs from New CCs** - Performs a Comparison between the *RefreshTable* and *ReferenceTable* in order to determine if the are "new" *CC Name* present in the *RefreshTable*, and not present in *ReferenceTable*.

3. **Addition of New CC or Bypass** - If there is determined that CC is present in the *RefreshTable* and not present in *ReferenceTable*, the row where this CC is located, in the *RefreshTable* is added to the end of the *ReferenceTable*.

4. **Supervisor Notification** - After completing the addition of the New CCs to the *ReferenceTable*, a printout notification, with the *CC Name* and timestamp, is sent to the console terminal to inform the operator.

3.3.3 Selection Cycle

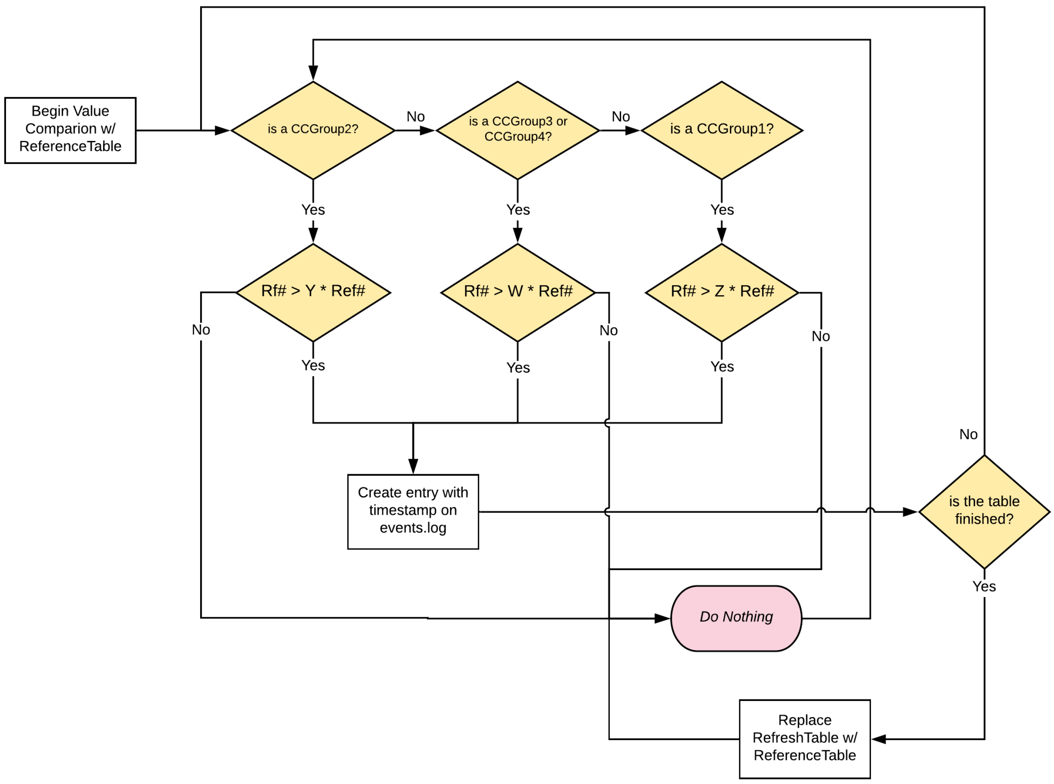

FIGURE 3.7: Selection Cycle of CC Groups

At this moment, the *RefreshTable* and the *ReferenteTable* have the same number of *CC*, making it possible to perform a comparison.

In order to be able to associate the constants and the two tables, a cycle is needed. In Figure 3.7 is detailed the steps followed:

The comparison is made by the level of priority, for this deployment, the internal failures of oTAS were of the most important ones. For this reason, they are the first ones to be checked.

1. **Search for existence of CC belonging to CCGroup2 in *RefreshTable*** - In order to begin, it is required to search on the *RefreshTable* for the existence of a *CC Name* belonging to the CCGroup2.

2. **CC belonging to CCGroup2 exists in *RefreshTable*** - If a *CC Name* belonging to the CCGroup2 in the *Refreshtable*, the number of occurrences

of that CC present in the *RefreshTable* is compared with the correspondent value in *ReferenteTable*, to determine if it is valid has an incident according the formula:

$$Rf(CCNameG2) > Y \times Refer(CCNameG2) \tag{3.3}$$

3. **Notification of Operator** - If the Incident is confirmed, a notification is written in the Operators console, with the timestamp, the *CC Name* and the number of occurrences of it.

4. **Next Groups** - This Cycle is required to continue to the other Groups (CCGroup3, CCGroup4 and CCGroup1), it is necessary to detect their existence and verify the possible occurrence of an incident. The process is the a repetition of steps **1**,**2** and **3** the CCGroup2, but the incident constant is different.

   For CCGroup3 and CCGroup4:

$$Rf(CCNameG3) > W \times Refer(CCNameG3) \tag{3.4}$$

$$Rf(CCNameG4) > W \times Refer(CCNameG4) \tag{3.5}$$

   For CCGroup1:

$$Rf(CCNameG1) > Z \times Refer(CCNameG1) \tag{3.6}$$

5. **Replace *ReferenceTable* with *RefreshTable*** - This is required for two main reasons:

   - The Goal is to find a "Trend Disruption". It is important to understand if the calls are being "cleared" at a normal rate due to traffic or due to an issue.

- On oTAS, the CC Report generated in OMU Resets after 15 minutes or before that if max memory allocated is reached due to Vendor KPI definitions. It is imperative to compare the same volumes of occurrences in *RefreshTable* and *ReferentTable*.

Chapter 4

Result Analysis

These results achieved for this study were collected by implementing the algorithm in section 3.3, to decrease the impact of the issue described in section 3.1.2 in one Mobile Operator served by oTAS.

4.1 Statistical Assumptions

This model comes with some assumptions that have to be made based on the behaviour of real traffic.

4.1.1 Historical Data

To be able to define the Thresholds for *Incident Detection*, it is essential to understand the behaviour and trend of the system.

If we check for instance the Figure 4.1, the VoLTE Call Success Rate over one month, we can visualize that, everyday issues are present. The width range of reasons (from UE radio failures, timeout delays and others) is causing the Success rate to be below 100%. The TAS is at the end of the VoLTE Call Flow being affected by E2E issues.

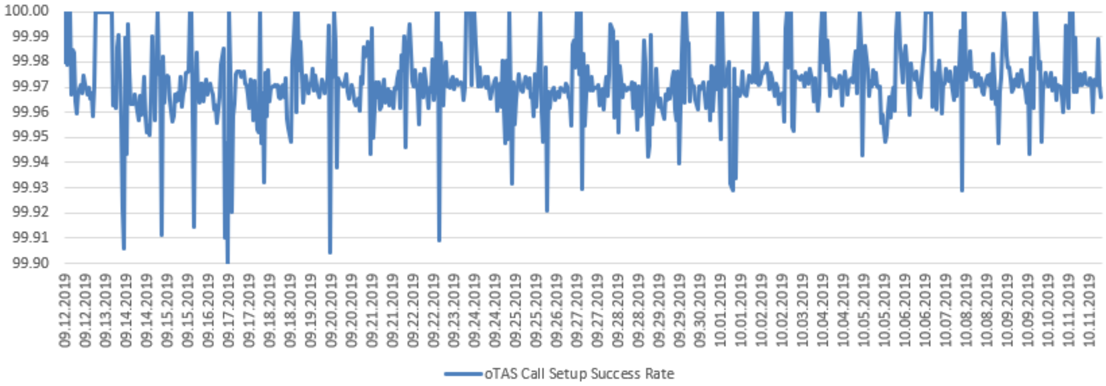

FIGURE 4.1: oTAS - VoLTE Call Success Rate

Due to the huge number of Subscribers and Calls being terminated successfully, issues are sometime hidden or impossible to detect in order to be replicated in a controlled environment. On Figure 4.2, we can visualize the volume of calls performed on the same period of the Figure 4.1. If we focus on the *CCGroup2*, *CCGroup3* and *CCGroup4*, we can view the number of permanent issues present in the network over a month, in Figure 4.3.

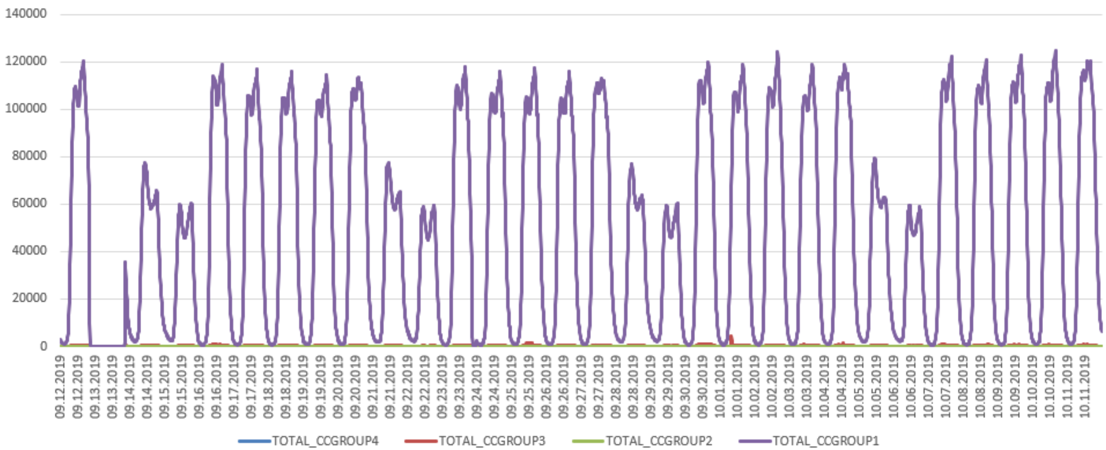

FIGURE 4.2: oTAS - Total of Occurrences of per CC Group

On Figure 4.4, we can visualized a increase of occurrences *CCGroup2* on October 1st. The Supervisor VM logged this event, and the Network Operator analysed it.

The thresholds for *Incident Detection*, for this study, were obtain with an empirical way based on the experience of the Network Operator. The values were the following:

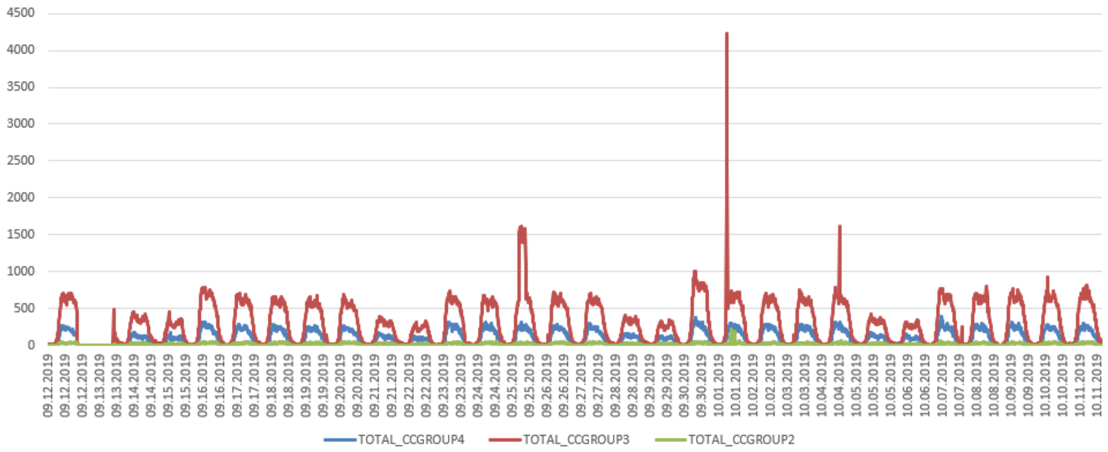

FIGURE 4.3: oTAS - Total of Occurrences Related with Fails

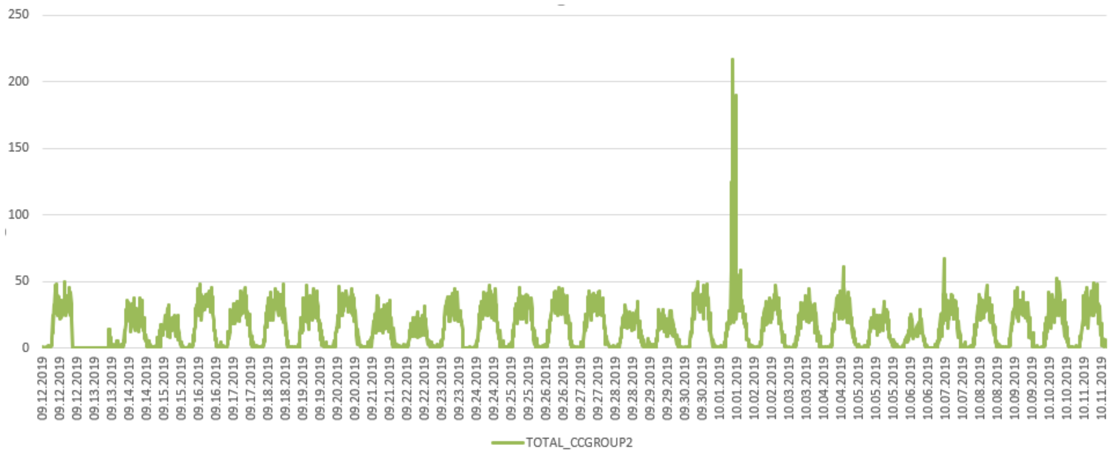

FIGURE 4.4: oTAS - Total *CCGROUP2*

- Y was the constant defined for CCGroup2, with the value *2*.
- W was the constant defined for CCGroup3 and CCGroup4, with the value *1.5*.
- Z was the constant defined for CCGroup1, with the value *1.3*.

4.1.2 Number of Subscribers

It is mandatory to assume that the number of subscribers using the system will continue to increase. The thresholds are required to are dimensioned considering with the provisioning of the subscribers.

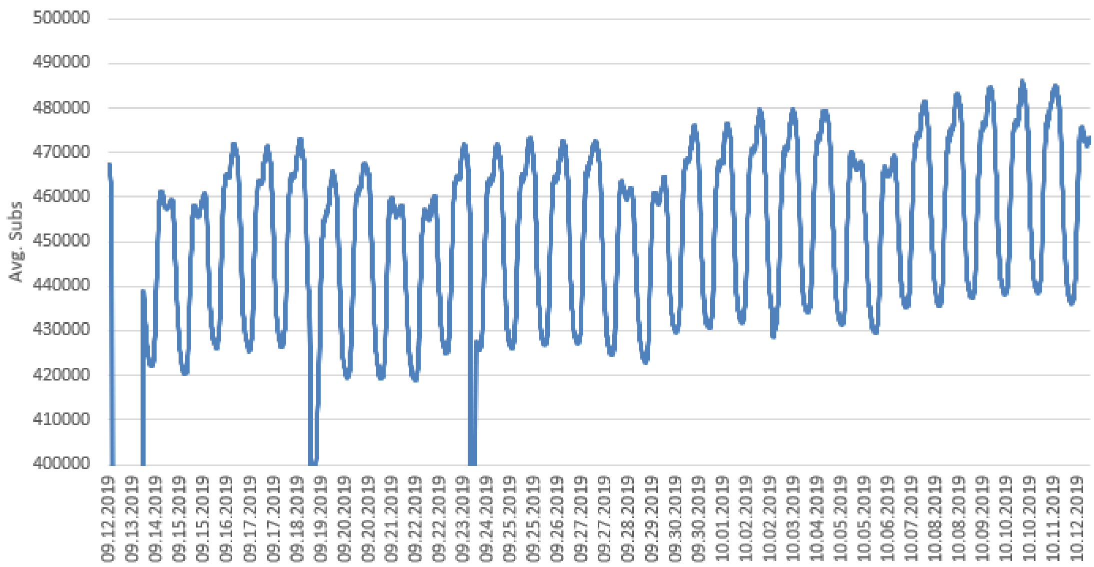

FIGURE 4.5: oTAS Number of Subscribers Growth

4.2 Achievements and Findings

In this section, this study exposed all the founding discovered by implementing this concept of a VoLTE *Cloud* environment. It contains reviews of the positive and negative aspects of this implementation.

4.2.1 Achievements

This Supervisor VM provides the advantage of continuous logging of events on the oTAS. However, it is not capable of identifying the root cause of the incident, or the impact for the end-user.

However, the primary purpose of this development was to "be able to identify incidents faster" and reduce the time in which an issue is Unknown, like mentioned on subsection 3.1.2

In Figure 4.6, it is displayed the *CFS* SLA KPI. This chart contains a contrast between before and after *Supervisor* VM implementation in *minutes of outage*. Like mentioned on section 3.1.1, this KPI depends on real-time duration of the incident, the total number of calls affected and it has a limitation of 52 *minutes of*

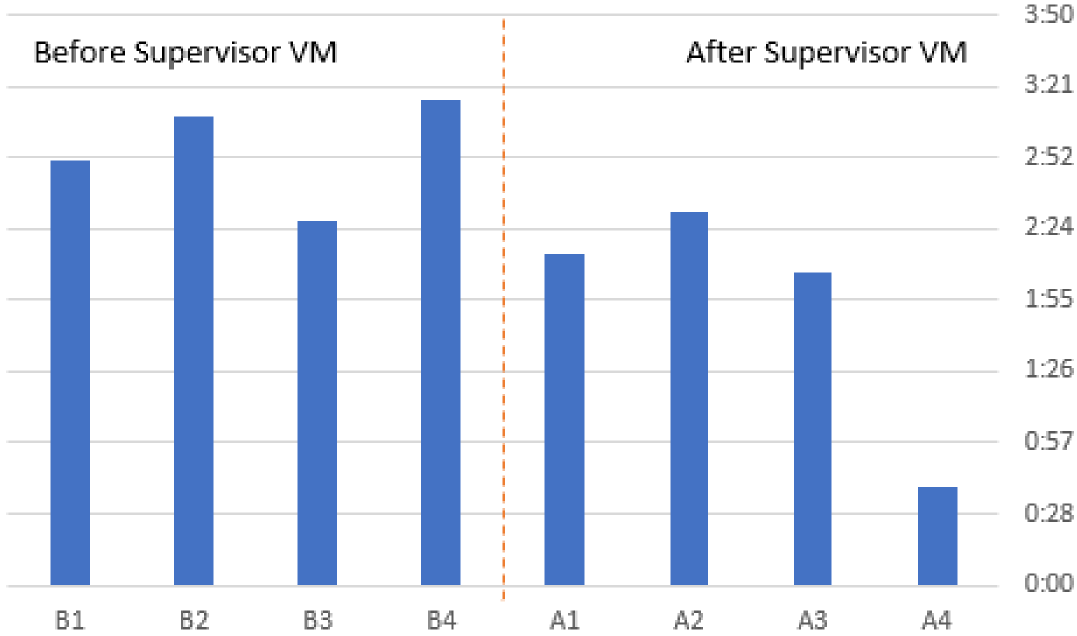

FIGURE 4.6: *CFS* - Before and After *Supervisor* VM in *minutes of outage*

outage total in a year (*YCFS*). Since the implementation of the *Supervisor* VM in this Mobile Operator, four incident occurrences occurred.

If we compare the average CFS of the incidents before (B1, B2, B3 and B4) and after (A1, A2 and A3), it is possible to visualise an improvement of CFS from an average of 2 minutes and 56 seconds to 2 minutes and 17 seconds. It is resulting in an improvement of 22.16%. The A4 was a partial outage, resulting in a significantly lower value of CFS in Figure 4.6.

If we compare the Real-Time, of the incident stages before and after, it is evident that the *Unknown issue* has been almost entirely neutralised by this application. The delays are related to human response because the logging of the machine was correct with the incident time.

Once again, if we compare the Real-Time before (B1, B2, B3 and B4) and after (A1, A2, A3, A4) it is possible to visualise an improvement in average from 49 minutes to 38 minutes. We are resulting in a combined improvement of 22.45%.

All the incidents presented in this study occurred due to different root causes. However, they were managed and solved by the same team of Network Operators.

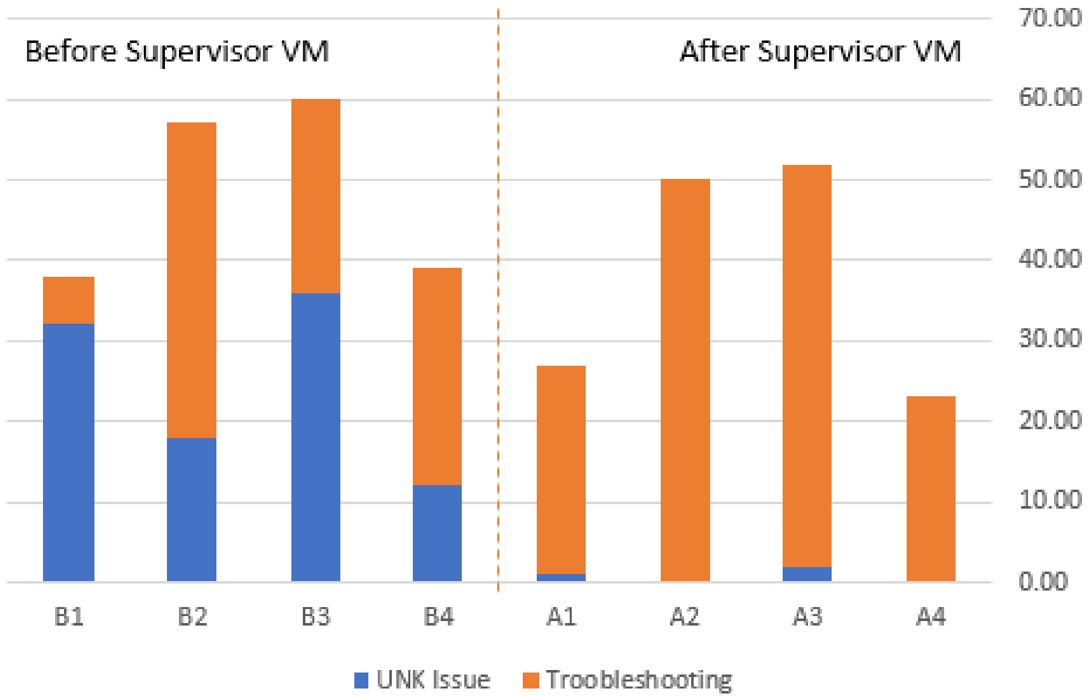

FIGURE 4.7: Real Time per Incident Stage - Before and After *Supervisor* VM in minutes

4.2.2 Findings

The main benefit of having a tool that performs autonomous monitoring and logging of equipment is the vast log events. However, a Service Operating Center (SOC) Engineer needs to be present to review the possibility of an issue continually.

This is the result of "Fixed" *Incident Detection* Thresholds, like mentioned on subsection 4.1.1.

The system is not able to adapt to the changes in traffic along the day due to the static thresholds. Also this system can not be adapted considering the increase of traffic created by the rise of Subscribers.

The *Incident* Logged referent to the Figure 4.4 was an incorrect incident. There is an increase of the *Clear Code* 603 - *NO RESPONSE FROM PARTNER PROGRAM BLOCK*, but there is no impact on the *oTAS VoLTE call Success Rate*, labeling this as a *False Incident*, on Figure 4.8.

For the future, it is imperative to find an alternative way to mitigate the high number of *false incidents* recorded like the ones in Figure 4.9 from a workload perspective for the SOC engineers but also from tool efficiency.

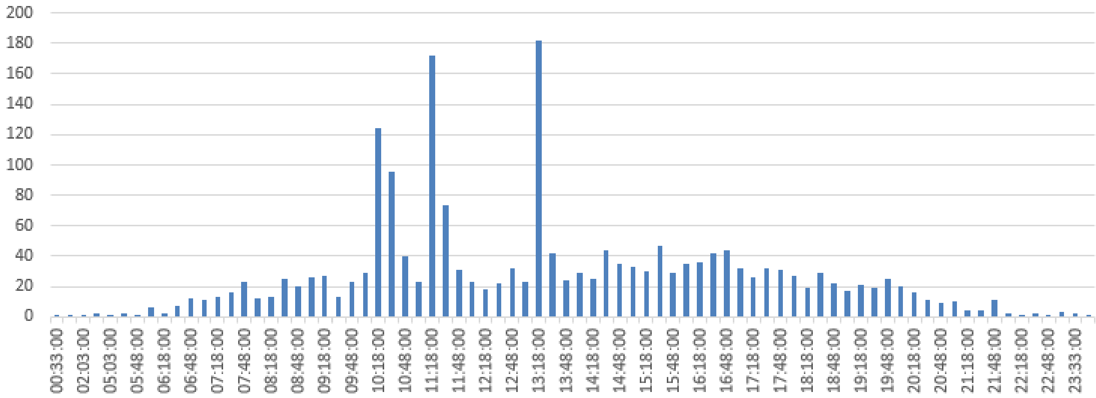

FIGURE 4.8: Clear Code 603 in a day

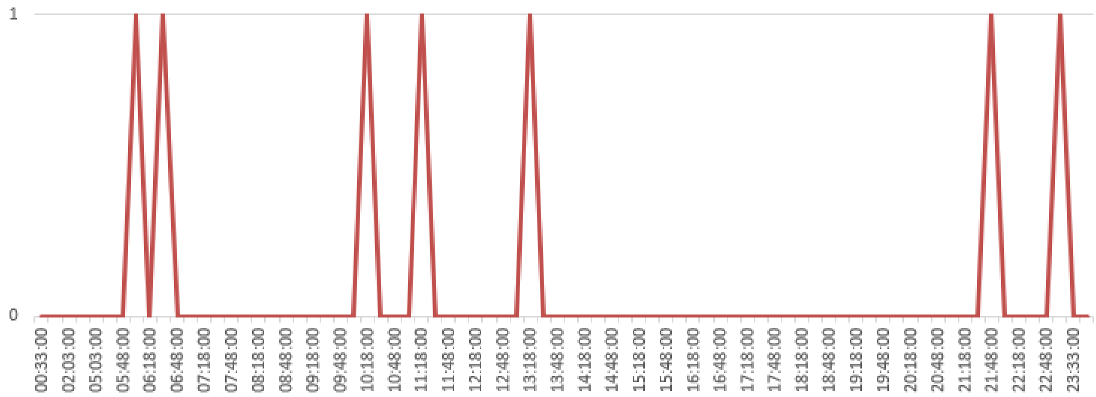

FIGURE 4.9: False Triggers of *CC* 603

It was evaluated the possibility to implement two features in the future, to mitigate these issues, which are:

- **Increase the number of *Incident Detection* Thresholds**.
- **Introduce History Prediction**.

In the next subsection, it was being described in more detail the reason why this is considered the next step to be implemented.

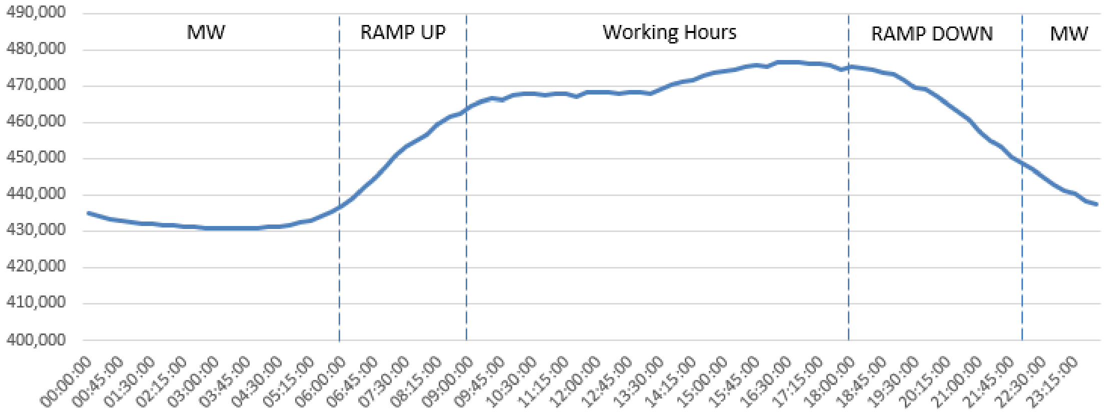

FIGURE 4.10: Amount of Subscribers during a Day

4.2.2.1 Increase the number of *Incident Detection* Thresholds

By reviewing the behaviour of the traffic on oTAS, it is noticed the presence of trends during a Week, Figure 4.10, which are the following:

- The working days have more traffic in comparison with the weekend.
- In the Mornings, the increase in traffic is faster than the Afternoon.
- There is a clear distinction between traffic before and after lunch.
- Traffic in Maintenance Window (MW) is almost zero.

Due to this, it may be considered inadequated the use of only one threshold for the same CC permanently. It becomes evident if we compare the number of successful calls in consecutive days, like in Figure 4.11.

The Goal with this action is to have different *Incident Detection* Thresholds per *CCGROUP* to the following time windows:

- **Ramp Up** (UP): From 06h00 - 09h00
- **Traffic Morning** (TM): From 09h00 - 12h00
- **Traffic Afternoon** (TA): From 12h00 - 18h00
- **Ramp Down** (TM): From 18h00 - 22h00

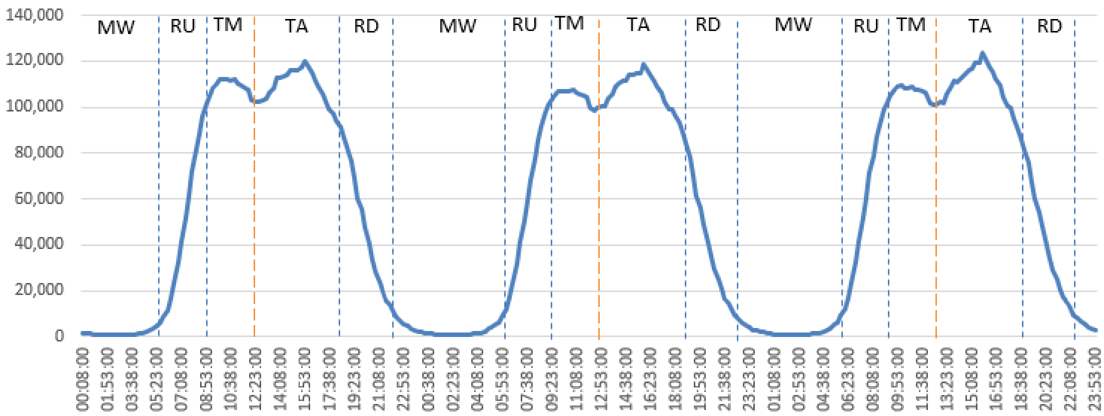

FIGURE 4.11: Amount of *CCGRP1* in 3 working days

- **Maintenance Window** (MW): From 22h00 - 06h00

With this *fine tuning*, it is expected to reduce the amount of *False Triggers* by having a unique configuration for each time of the day.

4.2.2.2 Introduce History Prediction

Like mentioned on subsection 4.1.2, the number of Subscribers is expected to increase. It is vital to allow the possibility of a threshold to adapt and predict the growth of the *CCGROUPx* based on history.

For this, we recommend using the *Growth* of the past two weeks a prediction.. For demonstration purposes, please find below the prediction of *CCGROUPx* on based on previous two weeks date.

$$Growth(w) = \frac{Total_{CC}(w-1) - Total_{CC}(w-2)}{Total_{CC}(w-2)} \tag{4.1}$$

Where $Total_{CC}$ is the sum of all the CC occurrences in that time frame.

By applying this two recommendation the formula for *Incident Detection* in the future should be:

$$(CCNameGx) > (T_W + Growth) \times Refer(CCNameGx) \tag{4.2}$$

Where T_w is the factor of the Time window at the time, mentioned on 4.2.2.1.

Chapter 5

Conclusion

5.1 Main Conclusions

The Chapter 3 describes as a method implemented to reduce the time in which an outage is undetected for the VoLTE service. The main objective of this method is to prevent to break the SLA commitment described in 3.1.1. It was implemented a Ubuntu Virtual Machine with MySQL extensions to run the process.

The Chapter 4 explains the initial assumptions required to implement the VM and analysed the results of the implementation. The outcome showed an almost complete neutralisation of the *Unknown Incident* Period causing a reduction in average of 22% in the Real-Time of the incident. Although successful, the implementation revealed some points that require improvement in the method.

Although all the incidents used for measurements had different root causes, the Operation team which resolve them is the same before and after the implementation, so their reaction time is a valid metric.

For this method to be implemented, the service needs to be deployed on a *Cloud* environment due to the need for deployment of a *Supervisor* VM.

5.2 Future Work

It is known that the need for efficiency and cost reduction is growing in the Telecommunication Market injunction with the number of subscribers increasing with a never-ending demand for service quality. These two factors are the primary motivators for a continuous trend of improvement in both techniques and standards.

The Cloudification is more present in the daily lives of the everyday life of all business, it's no surprise that the Telecommunication market will find more uses for it.

Regarding this study, it proves that it is possible for the user a VM to help monitor a system, although there are some topic which need future work:

- **Implementation of Data Analytics** to develop trend patrons and help define bottlenecks in the system. Machine learning is starting to have a crucial position in Telecommunication from a planning & dimensioning point of view.
- **Developing a Root Cause Index** would help the Operators to troubleshoot the issue faster, resulting in even better SLA KPIs. All vendor provide a health check procedure in case a specific CC is detected.
- **Implement Automatic Health Check Procedure** would also help the Operators, which has to perform this activity daily for maintenance. All kind of repetitive actions that used SSH or SFTP interfaces can be implemented in an automated way in a VM.

References

[1] M. Poikselkä, H. Holma, J. Hongisto, J. Kallio, and A. Toskala, *Voice over LTE: VoLTE*. Wiley, 2012. [Online]. Available: https://books.google.es/books?id=baCd1wIsyLsC

[2] N. Unuth, "Why whatsapp is so popular," https://www.lifewire.com/reasons-why-whatsapp-is-popular-3426372, 2018, updated: 2018-12-03.

[3] S. Esselaar and C. Stork, "Ott applications driving data revenue growth," ser. The 22nd Biennial Conference of the International Telecommunications Society: "Beyond the boundaries: Challenges for business, policy and society", June 24th - 27th, 2018, Seoul, Korea. Seoul: International Telecommunications Society (ITS), 2018. [Online]. Available: http://hdl.handle.net/10419/190337

[4] 3GPP, "Service requirements for the internet protocol (ip) multimedia core network subsystem (ims)." January 2015, v12.9.0. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/122200_122299/122228/12.09.00_60/ts_122228v120900p.pdf

[5] ——, "Codec for enhanced voice services (evs)," March 2015, version 12.1.0. [Online]. Available: https://www.etsi.org/deliver/etsi_tr/126900_126999/126952/12.01.00_60/tr_126952v120100p.pdf

[6] T. Ali-Yahiya, *Understanding LTE and Its Performance*, 1st ed. Springer Publishing Company, Incorporated, 2011. [Online]. Available: http://www.gbv.de/dms/tib-ub-hannover/623056186.pdf

[7] 3GPP, *IP Multimedia Core Network Subsystem (IMS) Multimedia Telephony Service and supplementary services*, March 2010, version 9.5.0. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/122100_122199/122173/09.05.00_60/ts_122173v090500p.pdf

[8] GSMA, *IMS Profile for Voice and SMS.* GSM Association, May 2018, version 12.0. [Online]. Available: https://www.gsma.com/newsroom/wp-content/uploads/IR.92-v12.0.pdf

[9] ——, *IMS Profile for Conversational Video Service.* GSM Association, May 2016, version 11.0. [Online]. Available: https://www.gsma.com/newsroom/wp-content/uploads/IR.94-v11.0.pdf

[10] ——, *IMS Service Centralization and Continuity Guidelines.* GSM Association, December 2016, version 14.0. [Online]. Available: https://www.gsma.com/newsroom/wp-content/uploads/IR.64-v14.0.pdf

[11] 3GPP, "Digital cellular telecommunications system (Phase 2+); UMTS; Network architecture," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) RTS/TSGS-0122228vc90, 01 2015, version 12.9.0. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/122200_122299/122228/12.09.00_60/ts_122228v120900p.pdf

[12] ——, "Characteristics of the ip multimedia services identity module (isim) application," October 2014, version 12.2.0. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/131100_131199/131103/12.02.00_60/ts_131103v120200p.pdf

[13] H. Schulzrinne, "The tel URI for Telephone Numbers," RFC 3966, Dec. 2004. [Online]. Available: https://rfc-editor.org/rfc/rfc3966.txt

[14] Nokia, "Open Telecom Application Server Cloud 16.5 MP5," Nokia, 2018. [Online]. Available: https://www.nokia.com/networks/products/telecom-application-server/#overview

[15] Yun Chao Hu and C. Granberg, "Customized Application for Mobile Enhanced Logic (CAMEL)." Institute of Electrical and Electronics Engineers (IEEE), aug 2005, pp. 1_115–1_127. [Online]. Available: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.602.7712&rep=rep1&type=pdf

[16] H. Schulzrinne and T. Taylor, "RTP Payload for DTMF Digits, Telephony Tones, and Telephony Signals," RFC 4733, Dec. 2006. [Online]. Available: https://rfc-editor.org/rfc/rfc4733.txt

[17] 3GPP, *Multimedia Resource Function Controller (MRFC) - Multimedia Resource Function Processor (MRFP) Mp interface.*, January 2015, version 12.9.0. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/123200_123299/123228/12.09.00_60/ts_123228v120900p.pdf

[18] H. Schulzrinne, E. Schooler, J. Rosenberg, and M. J. Handley, "SIP: Session Initiation Protocol," RFC 2543, Mar. 1999. [Online]. Available: https://rfc-editor.org/rfc/rfc2543.txt

[19] INTERNATIONAL TELECOMMUNICATION UNION, "Usage of cause and location in the Digital Subscriber Signalling System No. 1 and the Signalling System No. 7 ISDN User Part," *E 42671*, 2018. [Online]. Available: https://www.itu.int/rec/T-REC-Q.850

[20] 3GPP TS 23.002, "UTRAN Iu interface Radio Access Network Application Part (RANAP) signalling," *Version 3.4.0 Release 1999*, 1999. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/125400_125499/125413/03.04.00_60/ts_125413v030400p.pdf

[21] P. Wieder, J. Butler, W. Theilmann, and R. Yahyapour, *Service Level Agreements for Cloud Computing*, ser. SpringerLink : Bücher. Springer New York, 2011. [Online]. Available: https://books.google.pt/books?id=z306GUfFL5gC

[22] J. Sansbury, *Service Level Manager - Careers in IT service management.* BCS Learning Development Limited, 2017. [Online]. Available: https://books.google.pt/books?id=z306GUfFL5gC

[23] J. Gibson, R. Rondeau, D. Eveleigh, and Q. Tan, "Benefits and challenges of three cloud computing service models," in *Proceedings of the 2012 4th International Conference on Computational Aspects of Social Networks, CASoN 2012*, 2012, pp. 198–205. [Online]. Available: https://www.researchgate.net/publication/261422983_Benefits_and_challenges_of_three_cloud_computing_service_models

[24] S. Manvi and G. Shyam, "Resource management for infrastructure as a service (iaas) in cloud computing: A survey," *Journal of Network and Computer Applications*, vol. 41, 11 2013. [Online]. Available: https://www.researchgate.net/publication/259143215_Resource_management_for_Infrastructure_as_a_Service_IaaS_in_cloud_computing_A_survey

[25] redhat, "Understanding openstack," https://www.redhat.com/en/topics/openstack, 2019, updated: 2019-10-04.

[26] Openstack.org, "Conceptual architecture," https://docs.openstack.org/mitaka/install-guide-rdo/common/get_started_conceptual_architecture.html, 2019, updated: 2019-08-16.

[27] Nokia, "TAS Cloud Introduction," Nokia, 2016. [Online]. Available: https://www.nokia.com/networks/products/telecom-application-server/#overview

[28] Openstack, "Clusters," https://docs.openstack.org/senlin/latest/user/clusters.html, 2018, updated: 2018-07-05.