Olli Vierimaa

Cost Modeling of Cloud-Based Radio Access Network

School of Electrical Engineering

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo 2.10.2017

Thesis supervisor:

Prof. Heikki Hämmäinen

Thesis advisors:

M.Sc. (Tech.) Tiwari Bindhya

M.Sc. (Tech.) Jaspreet Walia

Aalto University
School of Electrical
Engineering

AALTO-YLIOPISTO
SÄHKÖTEKNIIKAN KORKEAKOULU

DIPLOMITYÖN
TIIVISTELMÄ

Tekijä: Olli Vierimaa

Työn nimi: Pilvipohjaisen radioliityntäverkon kustannusten mallintaminen

Päivämäärä: 2.10.2017 Kieli: Englanti Sivumäärä: 8+72

Tietoliikenne- ja tietoverkkotekniikan laitos

Professuuri: Tietoverkkotalous Koodi: S-38

Valvoja: Prof. Heikki Hämmäinen

Ohjaajat: DI Tiwari Bindhya, DI Jaspreet Walia

Mobiilidataliikenteen nopea kasvu haastaa nykyisen tavan rakentaa ja hallinnoida tämän hetkisiä radioliityntäverkkoja. Pilvipohjaista radioliityntäverkkoa tutkitaan ratkaisuksi tarjota tarvittavaa verkkokapasiteettia entistä taloudellisemmin. Tämän opinnäytetyön tarkoituksena on arvioida nykyisten ja pilvipohjaisten radioliityntäverkkoarkkitehtuurien kustannuksia riippuen verkon rakenteesta ja liikenteestä. Tämä toteutetaan luomalla kustannusmalli, joka perustuu asiantuntijoiden haastatteluihin. Mallin avulla on mahdollista vertailla annetun verkon eri arkkitehtuurien kokonaiskustannuksia ja selvittää taloudellisin radioliityntäverkkoarkkitehtuuri verkolle.
Mallinnuksen tulokset osoittavat, että pilvipohjaisen radioliityntäverkon taloudelliset hyödyt ovat riippuvaisia dataliikenteen kasvusta verkossa. Vähäisellä dataliikenteen kasvulla pilvipohjaisen radioliityntäverkon kustannusedut ovat kyseenalaiset, mutta suurella dataliikenteen kasvulla saadaan selviä säästöjä verrattuna nykyisiin arkkitehtuureihin.

Avainsanat: Cloud RAN, C-RAN, kustannus mallintaminen, kustannus analyysi, liiketoiminta, TCO, pilvi

AALTO UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING

ABSTRACT OF THE
MASTER'S THESIS

Author: Olli Vierimaa

Title: Cost Modeling of Cloud-Based Radio Access Network

Date: 2.10.2017 Language: English Number of pages: 8+72

Department of Communications and Networking

Professorship: Communications Ecosystems Code: S-38

Supervisor: Prof. Heikki Hämmäinen

Advisors: M.Sc. (Tech.) Tiwari Bindhya, M.Sc. (Tech.) Jaspreet Walia

The rapid growth of mobile data traffic is challenging the current way of building and operating the current radio access network. Cloud-based radio access network is researched as a solution to provide the required capacity for rapidly growing traffic demand in more economical manner.
Scope of this thesis is to evaluate the costs of different existing and future radio access network architectures depending on the given network and traffic scenario. This is done by creating a cost model based on expert interviews to solve the most economical solution for the given network in terms of total cost of ownership.
The results show that the cloud-based radio access network's cost benefits are dependent on the expected traffic growth. In the low traffic growth scenario, the cost benefits of cloud-based radio access network are questionable, but in the high traffic growth scenario clear cost benefits are achieved.

Keywords: Cloud RAN, C-RAN, cost modeling, cost analysis, business, TCO, Cloud

Preface

This thesis was conducted at Nokia Oyj in Espoo Finland. First of all I'm grateful that I was given this interesting topic to work with in Nokia. I would want to thank all my colleagues at Nokia and especially Sumit Krishna who was able to guide me with my work. I would also want to thank my thesis advisor Tiwari Bindhya for his guidance, support and all the insightful discussions we had.

Additionally I want to thank my professor Heikki Hämmäinen for supervising my thesis, giving valuable feedback and advising me even before my thesis, and Jaspreet Walia for being my second thesis advisor and giving feedback on my work.

Finally I'm grateful for the continuous support I have gotten from my family and friends.

Otaniemi, 23.8.2017

Olli M. Vierimaa

Contents

Abbreviations

API	Application Programming Interface
APRU	Average Revenue Per User
ARQ	Automatic Repeat Request
BBM	Proprietary Baseband Module Unit
BBU	Baseband Unit
BS	Base station
BSC	Base Station Controller
CAGR	Cumulative Annual Growth Rate
CAPEX	Capital Expenditures
CoMP	Co-operative Multi-Point processing technology
CoMP JT	Co-operative Multi-Point processing technology Joint Transmission
COTS	Commercial off-the-shelf
C-plane	Control Plane
CPRI	Common Public Radio Interface
C-RAN	Centralized Radio Access Network
CUE	Connected User Equipment
D-RAN	Distributed Radio Access Network
E2E	End-to-End
eNB	Evolved NodeB
EPC	Evolved Packet Core
E-UTRAN	Evolved Universal Terrestrial Radio Access Network
EPS	Evolved Packet System
GSM	Global System for Mobile Communications
HARQ	Hybrid Automatic Repeat Request
HSDPA	High Speed Downlink Packet Access
HSS	Home Subscriber Service
HSUPA	High Speed Uplink Packet Access
HTTP	Hyper Text Transfer Protocol
HW	Hardware
ICI	Inter-Cell Interference
ICIC	Inter-Cell Interference Coordination
IoT	Internet of Things
JT	Joint Transmission
LTE	Long Term Evolution
LTE-A	Long Term Evolution Advanced
M2M	Machine-to-Machine
MAC	Medium Access Channel
MEC	Mobile Edge Computing
MIMO	Multiple Input Multiple Output
MME	Mobility Management Entity
MNO	Mobile Network Operator

Abbreviations

NEP	Network Equipment Provider
NFV	Network Function Virtualization
NRT	Non-Real-Time
OAM	Operation and Management
OBSAI	Open Base Station Architecture Initiative
OPEX	Operational Expenses
OTT	Over-The-Top
PCRF	Policy and Charging Rules Function
PDCP	Packet Data Convergence Protocol
PDN	Packet Data Network
PDN-GW	Packet Data Network Gateway
PHY	Physical Layer
QoS	Quality-of-Service
RAN	Radio Access Network
RAT	Radio Access Technology
RF	Radio Frequency
RLC	Radio Link Control
RNC	Radio Network Controller
RRC	Radio Resource Control
RRH	Remote Radio Head
RT	Real-Time
SAE	System Architecture Evolution
SDN	Software Defined Networking
S-GW	Serving Gateway
SW	Software
TCO	Total Cost of Ownership
TE	Terminal Equipment
UE	User Equipment
UMTS	Universal Mobile Telecommunications Systems
VoIP	Voice over Internet Protocol
VM	Virtual Machine
VNF	Virtualized Network Function
WCDMA	Wideband Code Division Multiple Access

1 Introduction

1.1 Motivation

No more than 20 years ago, people did not see owning a mobile phone or being available around the clock as a necessity. Now being without your phone can leave you feeling naked, as the mobile phone has almost become a physical extension of its users. Mobile connectivity has become essential for many people. Features like mobile voice, messaging, applications and video are becoming crucial part of business and consumer user's lives.

We have come a long way from 1G networks offering basic mobile voice and messaging services to 4G networks, where we have all, what the Internet can offer in our palms. The world is more connected than ever before. With the forthcoming and rising popularity of technologies such as Internet of Things, Machine-to-Machine (M2M) communication and digitalization of products and services a clear trend can be seen that the world is moving towards providing ubiquitous connectivity anywhere and anytime. Moving towards 5G we are on the verge of fully experiencing self-driving cars, wearable devices, smart cities, augmented reality and virtual reality. Evolution is happening from connecting people to connecting the world.

Global mobile traffic grew 63 % only in 2016 from the previous year, and projections are that it will increase seven-fold between 2016 and 2021. Respectively, nearly half a billion mobile devices and connections were added globally in 2016 . In 2021 there will be 1.5 mobile devices per capita, a total of 11.6 billion exceeding the projected world's population at that time (7.8 billion). In figure 1 we can see the significant rise in mobile traffic and especially the rise in popularity of 4G mobile broadband, which will be accountable for three quarters of global mobile traffic in 2021.[1]

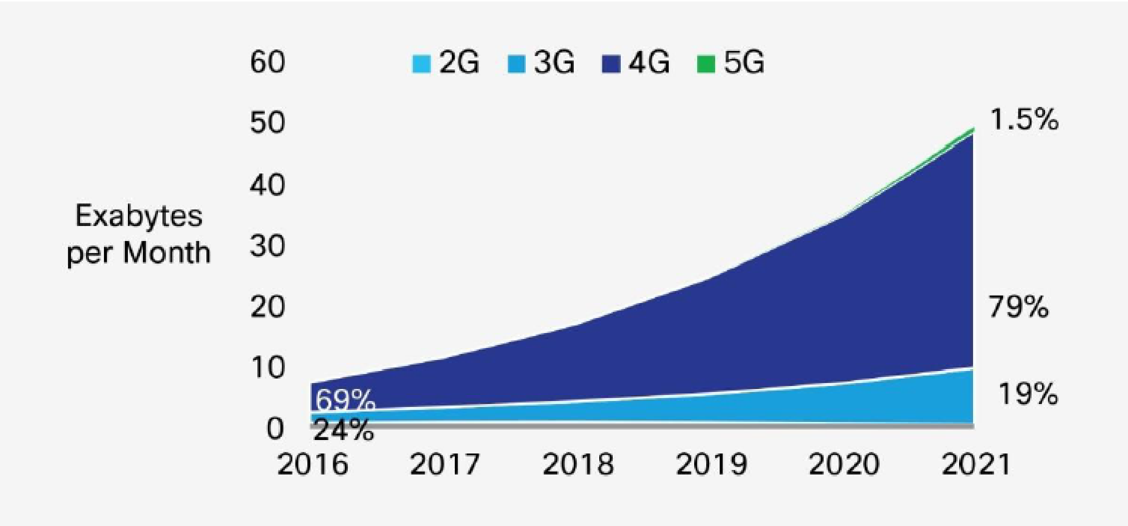

Figure 1: Mobile network data usage forecast and popularity of 4G. [1]

It is clear that both, the mobile traffic and the number of mobile devices are growing dramatically. Thus, the need for investing into the mobile network architecture to improve the capacity is imminent. At the same time, people are using more data in the network and expecting prices to stay the same. In figure 2, we can see that globally total mobile revenue growth is slowing down and Average Revenue Per User (ARPU) has been decreasing for multiple years and the same trend will likely to continue in the future [2]. For Mobile Network Operators (MNOs) this means that meeting the network capacity demands must be achieved with much lower costs. This has made MNOs to become more aware of their network's Total Cost of Ownership (TCO).

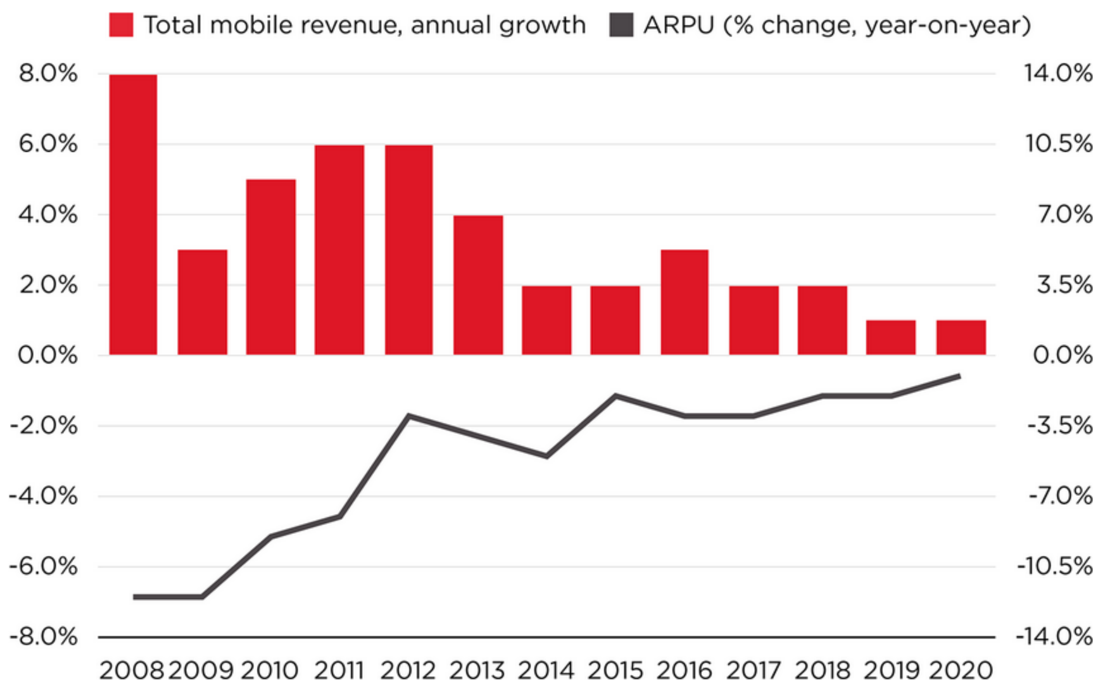

Figure 2: Mobile revenue growth and ARPU in 2008-2020. [3]

Radio Access Network (RAN) provides the connectivity between the users and the core network and it is the MNO's main asset for providing features such as high data rates, Quality-of-Service (QoS) and network availability to its users. Currently, 80 % of the MNO's Capital Expenditures (CAPEX) is spent on the RAN to meet the rising demands of the growing number of users. Operating and maintaining the RAN is also expensive as RAN's Operating Expenses (OPEX) account for 60 % of the TCO. For MNO's, RAN is the most expensive part of the network, which is why it is very attractive to find a solution to reduce its TCO.[4]

IT world has been using virtualization successfully for its benefits such as reduced costs, flexibility and scalability in data centers. Leveraging cloud computing and virtualization in the Radio Access Network could be the solution for MNOs to meet the demand for much needed capacity with lower costs. Cloud-based Radio Access

Network (Cloud RAN) utilizes cloud computing in the Radio Access Network by moving and centralizing the baseband processing into the cloud. Cloud RAN can be a promising solution for MNOs to reduce investments in base station equipment, site rentals and power consumption. Therefore, effectively reducing TCO of their network to become more profitable and competitive.

1.2 Research Problem

The main research problem of this thesis is: "What is the most economical Radio Access Network architecture for the given use case and existing cellular network deployment?"

The objective of the research is to create a cost model that can be used to compare different existing and future RAN architectures in terms of costs in the given use case scenario. This can give a better understanding of the feasibility of the future RAN architectures in terms of costs.

1.3 Research Scope and Methods

The scope of this thesis is to perform cost evaluation of the different Radio Access Network architectures (Distributed-RAN, Centralized-RAN, Cloud Distributed-RAN and Cloud Centralized-RAN) to solve what is the most economical network architecture deployment for the given use case scenario. The scope consists of wide area network analysis and does not include local area solutions such as small cells. Motivation for the thesis is to research the feasibility of the Cloud RAN architecture as the future RAN architecture in terms of costs and get a better understanding of Cloud RAN's value proposition. Cost modeling of Cloud RAN is focused on analyzing the changes in the baseband processing in the architectures.

This thesis is based on qualitative analysis and the methods used in this thesis consists of literature analysis, expert interviews and internal company document analysis. Tools used in this thesis were Microsoft Excel for creating the Cloud RAN TCO model and www.draw.io for creating figures.

1.4 Structure of the Thesis

Chapter 2, Background, provides the introduction to mobile networks. Consisting of the history and technologies in mobile networks and providing the background for understanding the Cloud RAN technology. Chapter 3, Feasibility of Cloud RAN explains the motivation for researching the business benefits of Cloud RAN, challenges in the current RAN, Cloud RAN's business benefits and present the motivation for Cloud RAN's cost modeling. Chapter 4, Cloud RAN TCO Model explains the created Cloud RAN cost model in more detail. Chapter 5, Analysis consists of analyzing the results of using the Cloud RAN TCO model in the example case. Chapter 6, Conclusions consists of discussion about the the results, possible exploitation of the results and future work for researching the cost benefits of Cloud RAN.

2 Background

2.1 Mobile Networks

Mobile networks revolutionized the way of communicating between people by enabling users on the move to connect with each other in large areas. Mobile networks provide coverage to users by dividing areas into cells. Each cell in the network consists of cell sites with base stations, which usually contain three antennas that are pointed to different directions providing coverage in form of a cell thus creating a cellular network.

Mobile network users are referred to as User Equipment (UE), which includes mobile phones, tablets and data cards used by computers. UE consist of two components, which are the User Service Identity Module or SIM-card and the Terminal Equipment (TE). Base stations in 2G and 3G networks are connected to a centralized intelligent controller also known as Base Station Controller (BSC) or Radio Network Controller (RNC) as seen in figure 3. In 4G, the RNC functionality was integrated into the LTE base stations that are also known as evolved NodeBs (eNBs). RNCs and eNBs then provide the connectivity to the core network and from there enabling users to communicate with each other through external networks. [5]

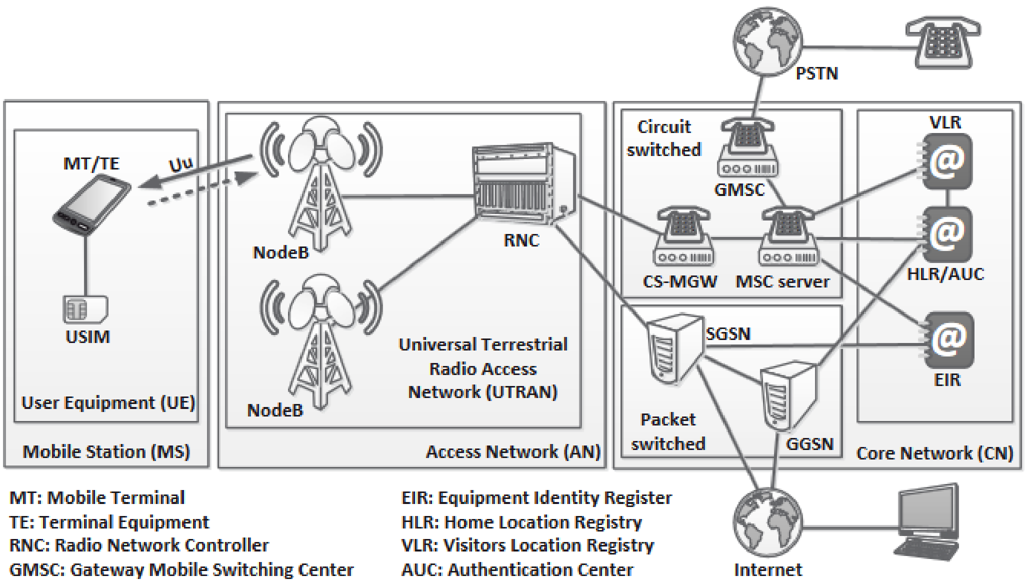

Figure 3: 3G UMTS network architecture.[5]

Mobile networks have evolved a lot in relatively short time as seen in figure 4. First generation of mobile networks (1G) was developed in early 1980's and it only provided basic analog voice feature for its users. 1G network was the foundation for the mobile networks. At that time each country had its own 1G system, which limited the usage and prevented economies of scale. [6]

The second generation networks (2G) (figure 4) were released in 1991 under the name of Global System for Mobile Communications (GSM) and they replaced the 1G systems in most countries. The first GSM call was made in 1991 in Finland by Prime Minister Harri Holkeri in Radiolinja's network [7]. 2G included digital voice and simple data services for subscribers and it made mobile phones truly global. During the 2G era, mobile phone prices went down and the mobile phone penetration increased from 2 % to 76 % between the years 1993 and 2002. [6] It can be said that 2G brought mobile phones for the masses.

The third generation (3G) using the Universal Mobile Telecommunication Systems (UMTS) standard was introduced in 1998. 3G was based on the 2G system and shared the similar network architecture (figure 3). 3G introduced the Wideband Code Division Multiple Access (WCDMA), which uses larger bandwidth than was available in 2G. Compared to previous generations, 3G improved the data rates and spectral efficiency significantly. 3G brought mobile access to audio content, video streaming and video calls with its higher data rates of 2 Mb/s. New 3G releases such as High Speed Downlink Packet Access (HSDPA) and High Speed Uplink Packet Access (HSUPA) improved the 3G even further and some operators marketed it even as 4G. 3G introduced the mobile broadband to users with its high data rates and mobile access to wide variety of content over the Internet. 3G started the evolution towards smart devices and Internet style mobile applications. [5]

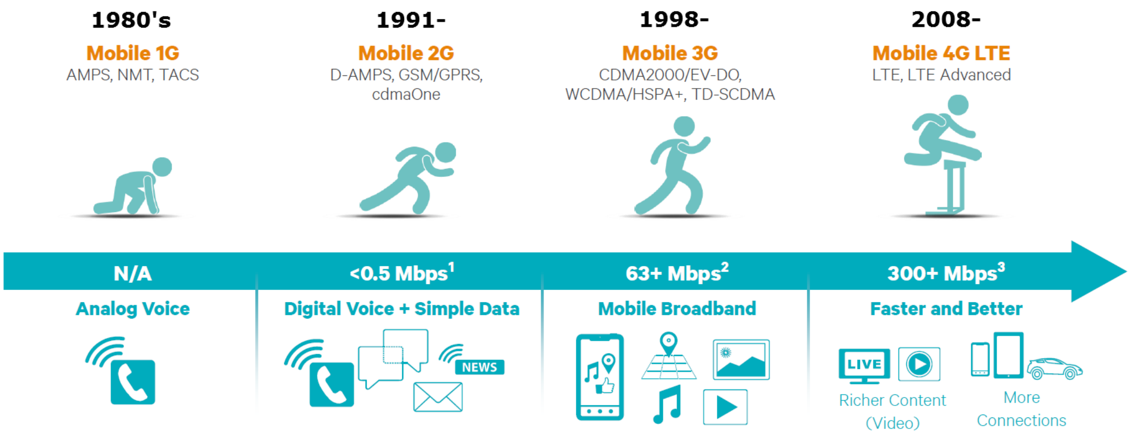

Figure 4: Mobile network evolution from 1G to 4G. [8]

The current mobile network technology Long Term Evolution (LTE) or sometimes referred to as the fourth generation (4G) was finalized in 2008 and publicly released in 2009. LTE released in 2009 did not actually full-fill the International Mobile Telecommunications-Advanced (IMT-A) specifications for 4G and it is therefore more accurately described as 3.9G. Later on LTE Advanced (LTE-A) was released, which full-filled the 4G standards and could be therefore described as true 4G. [9] Where the previous generations were still supporting circuit-switched networks 4G was designed to support only packet switching with its all-IP network. One of the major changes in 4G compared to previous generations was the simplified and flat RAN. The RAN in 4G only consists of eNBs and there is no centralized intelligent controller such as the RNC in previous generations. The reason for distributing the intelligence among the base stations was to reduce system vulnerability caused by RNC failures, network complexity, connection set up time and the required handover time. Motivations for 4G were to ensure the continuity of competitiveness of the 3G system for the future, offer higher data rates and quality of service to meet the user demands, optimizing the packet switched system, reducing costs and network complexity. [5]

2.2 LTE Network Architecture

LTE system (figure 5) was designed to be completely packet switched all-IP based network, providing mobile IP connectivity between the UE and Packed Data Network (PDN). LTE system architecture can be divided into two parts, which are the Evolved Universal Terrestrial Radio Access Network (E-UTRAN) and System Architecture Evolution (SAE). The E-UTRAN evolved from 3G UMTS RAN and it is the RAN technology of LTE providing the network access for UE in the network. SAE supports the evolution of the packet core network, which is also known as the Evolved Packet Core (EPC). Together E-UTRAN and SAE form the Evolved Packet System (EPS). EPS uses EPS bearers to route IP traffic between PDN gateway (PDN-GW) and UE. Bearer is defined as IP packet flow with predefined QoS characteristics. E-UTRAN and EPC are both responsible for setting up and releasing bearers depending on the application's QoS requirements. LTE allows multiple bearers to be established for users, which means that the user can have a voice call using the Voice over Internet Protocol (VoIP) and browse the Internet at the same time with Hyper Text Transfer Protocol (HTTP).[5]

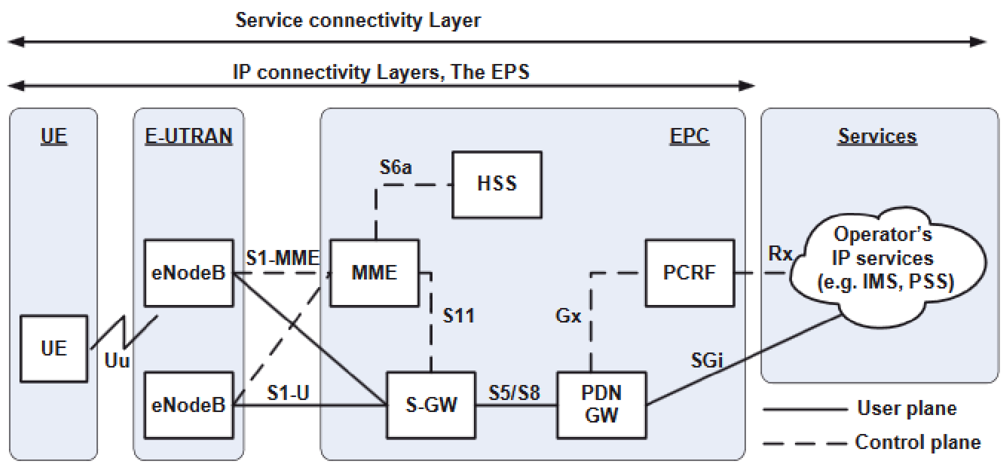

Figure 5: LTE system architecture.[5]

LTE EPC is also known as the core network and its main function is to provide necessary functionalities for the users and create their required bearers. EPC (figure 5) consists of three main entities, which are Mobility Management Entity (MME), PDN-GW and Serving gateway (S-GW). Other entities in EPC are the Home Subscriber Service (HSS) and Policy and Charging Rules Function (PCRF). MME provides the control functions and signaling for the EPC and it is only involved in the C-plane. MME functions include mobility management, authentication, security, roaming, bearer establishment and handovers. PDN-GW is the user connectivity point for the user traffic and it is responsible for assigning IP addresses for the UE and classifying user traffic into different QoS based bearers. S-GW acts as the main gateway, where all the user traffic goes through, for the user traffic between the eNB and PDN-GW. S-GW also performs other functions such as routing, forwarding and gathering charging information. [10]

LTE E-UTRAN consists of interconnected eNBs, which are using the X2 interface to communicate with each other and S1 interface to connect to the EPC. LTE eNBs acts as a bridge between the UE and EPC, providing the necessary radio protocols for UE to be able to send and receive data securely with the PDN-GW. E-UTRAN radio protocols consist of User plane (U-plane) and Control plane (C-plane). U-plane protocols are used to transfer the actual user data over the LTE network, whereas the C-plane protocols are used to control and establish user connections and bearers within the E-UTRAN. The main LTE radio protocols are:

– **Radio Resource Control (RRC)** performs handover functions such as handover decisions and transferring UE from serving eNB to target eNB. RRC is also responsible for controlling the periodicity of the Channel Quality Indicator (CQI) and maintenance and set-up of radio bearers.[5]

– **Packet Data Convergence Protocol (PDCP)** is responsible for IP header compression, which is important to reduce the overall overhead and thus improving efficiency over the radio interface. PDCP also performs ciphering and integrity protection functions.[5]

– **Radio Link Control (RLC)** handles segmentation and integration of PDCP packets, retransmissions and guarantees in-sequence packet delivery to higher layers. RLC also performs error corrections using the Automatic Repeat Request (ARQ) methods.[5]

– **Medium Access Channel (MAC)** is responsible for scheduling air interface resources in downlink and uplink and satisfying user QoS over the air interface. MAC also handles the Hybrid Automatic Repeat Request (HARQ).[5]

– **Physical Layer (PHY)** handles radio related issues such as modulation/demodulation, coding/decoding and multi-antenna mapping.[5]

The existing RAN architectures are based on proprietary baseband processing hardware. These include the Distributed Radio Access Network (D-RAN) (figure 6), which is the current architecture of choice and Centralized Radio Access Network (C-RAN) (figure 7), which has been used mainly in highly populated events such as concerts and recently by MNOs in Asia.

The base stations in the existing RAN architectures consist of proprietary Baseband Units (BBUs) that are responsible for the baseband processing and Remote Radio Heads (RRHs), which contain the base station's radio functionalities. BBUs and RRHs are interconnected with optical fiber using Common Public Radio Protocol (CPRI) and the base stations are connected to the core network with Ethernet backhaul. [11]

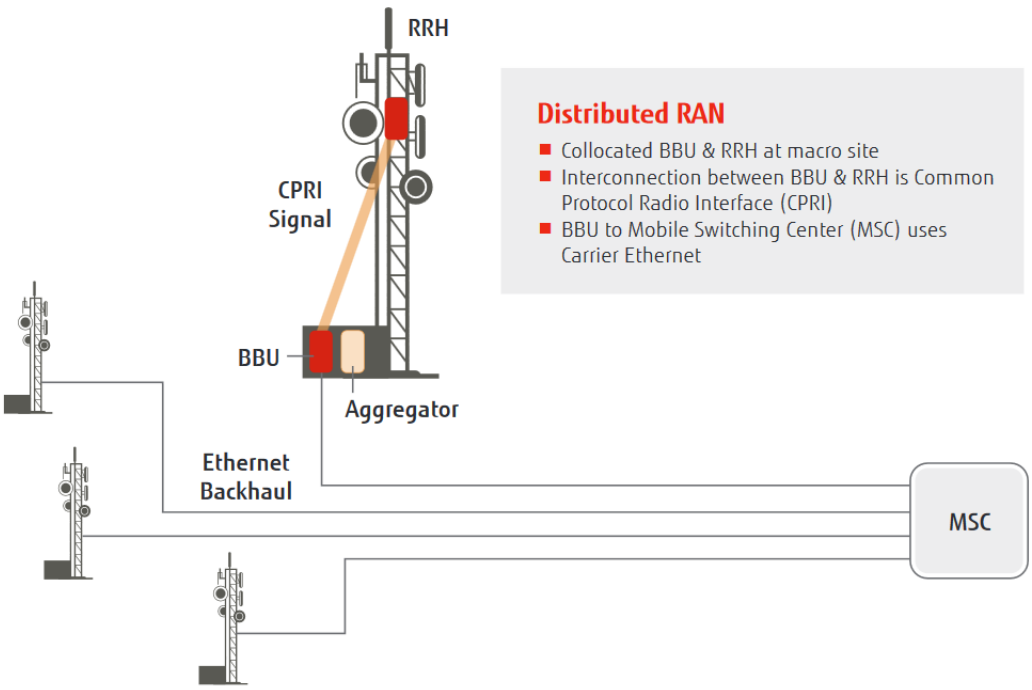

Figure 6: D-RAN architecture.[11]

D-RAN (figure 6) offers a simple and flat architecture, where the base stations can handle all the RAN functionalities and there are no other entities such as the RNC in 3G between the base stations and the core network. Due to D-RAN Ethernet backhaul's relaxed latency and range limitations base stations can be deployed easily even in rural areas with long distances. The problem with the current D-RAN architecture is the costs. Building, upgrading and operating base stations in D-RAN architecture is expensive due to network architecture's distributed nature. Challenges of the current D-RAN architecture are explained in more detail in chapter 3.1.

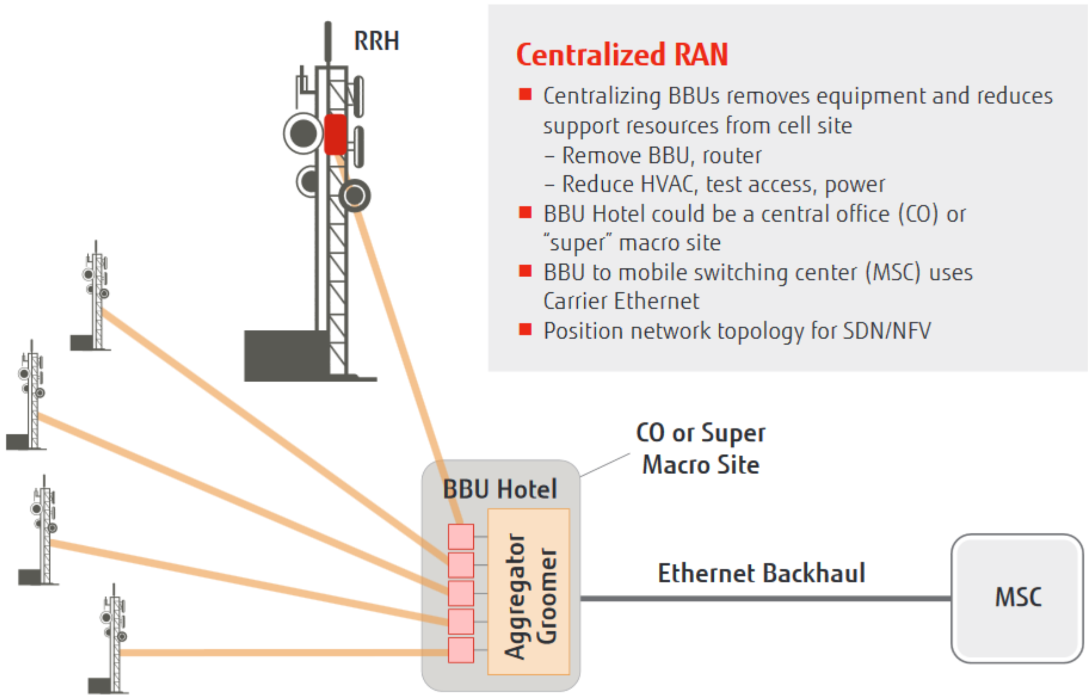

Figure 7: C-RAN architecture.[11]

C-RAN (figure 7) evolves the D-RAN architecture further by physically separating the BBUs from base stations and pooling the BBU functions to centralized location such as "BBU hotel" or sometimes referred to as Point-of-Presence (PoP). In BBU hotels the BBU pool can operate multiple RRHs [12]. C-RAN adds a new element, the BBU hotel, in the overall RAN architecture, but simplifies the base stations themselves so that they only consist of antennas and the RRH. Fronthaul network between the RRH and BBU requires high speed and low latency optical link [13]. These requirements introduce range limitations to base station and BBU hotel deployment. C-RAN is best suited for densely populated areas, where the optical fiber network is already in place and the best pooling gains can be achieved with centralization. [13]

C-RAN architecture aims to solve the challenges that MNO's are facing by increasing bandwidth capacity economically. Centralization of BBUs reduces required hardware, site acquisition costs, allows more efficient maintenance and increases spectral efficiency. Thus, improving network performance and reducing RAN's TCO for MNOs.[14] Current C-RAN deployments are based on proprietary hardware, which does not allow scalability, flexible and adaptive software deployment and leaves the potential of cloud computing unused [13].

2.3 5G

The fifth-generation networks (5G) will be the next evolution for mobile networks towards truly connected world. 5G is aiming to provide ubiquitous connectivity for any kind of device and application that can benefit from being connected. 5G network is not based on single Radio Access Technology (RAT), but rather a collection of different access and connectivity technologies used together to meet the demands of future mobile networks.

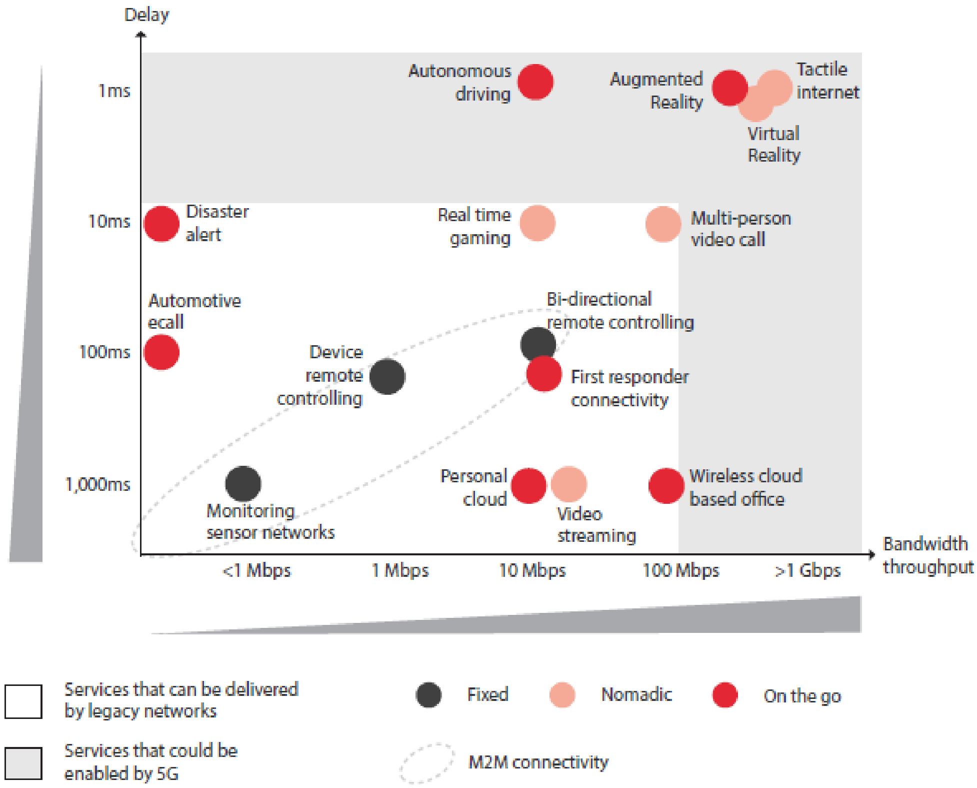

Figure 8: Bandwidth and latency requirements for different use cases in mobile networks.[15]

Standardization and specification of 5G is still in progress as IMT-2020 is being currently worked on [16]. Most defining specifications for 5G are supporting 10-100 times more connections, providing high data rates of over 1 Gb/s and sub-1ms latency. Reaching these specifications will enable revolutionary new use scenarios for mobile networks (figure 8) such as autonomous driving, augmented and virtual reality and tactile internet.[15]

5G's capabilities can be realized with three different use cases, which are the evolutionary wide area 5G, revolutionary wide area 5G and revolutionary local area 5G. Evolutionary wide area 5G aims to enhance the current 4G in every way by improving the quality with higher bandwidth, reliability and availability and lower latency. Revolutionary wide area 5G aims to provide massive M2M communications that supports massive number of low cost devices with long battery lives sending small volumes of data. Revolutionary local area 5G aims to create the foundation for critical M2M communication applications such as traffic safety, control of critical infrastructure and wireless connectivity for industrial processes. This is made possible by ultra-reliable, very high availability and ultra-low latency.[17]

5G will face massive increase in data traffic, which will challenge the current network architecture and especially the RAN [15]. This exponential increase in data traffic will occur as more devices will be able to access the Internet, devices become more powerful, bandwidth hungry services will become more common and more devices are integrated into everyday life by different industries.[13] One vision for the 5G network architecture is End-to-End (E2E) cloud network, where Cloud RAN will be essential part of the network to solve the challenges caused by exponential increase in data traffic economically.

2.4 Cloud Based Radio Access Network

Mobile operators are facing a difficult situation meeting the demands of increasing mobile traffic, limited spectrum availability and mass adaptation of mobile broadband, which are challenging the ways of building the traditional Radio Access Network [18]. As LTE's spectrum efficiency is approaching Shannon limit, the most obvious way of increasing network capacity is either by adding more cells or implementing techniques such as Multiple Input Multiple Output (MIMO) or Massive MIMO. However, increasing cell densification has led to rising costs and use of MIMO has resulted in increasing inter-cell interference and high costs.[14] [19] Thus it is clear that new solution is needed to tackle the challenges of traditional RAN with low costs.

According to China Mobile Research Institute [4] the future RAN solution should have:

1. Lower costs,
2. High spectral and energy efficiency,
3. Open platform for supporting multiple standards and enable smooth evolution,
4. Platform for additional revenue generating services.

Cloud RAN sometimes referred to as C-RAN was first introduced by IBM under the name of Wireless Network Cloud [IBM] and later on researched further by the China Mobile Research Institute [4]. It is now on the sights of multiple companies such as Nokia [20], Huawei [21], Intel [22], ZTE [23], Ericsson [18] and Texas Instruments [24]. Cloud RAN is expected to be the solution to address the challenges faced by the MNOs and full-fill the goals of the future RAN.

The innovation behind Cloud RAN is centralizing baseband processing of multiple base stations into virtualized baseband unit pools (BBU pools) to achieve pooling gains for the baseband processing. Virtualized BBU pools can respond to non-uniform traffic more efficiently by dynamically allocating resources on demand

[25]. New BBUs can also be added, upgraded and removed easily, which improves network scalability and maintenance. Cloud RAN architecture requires fewer BBUs than the traditional RAN architecture, which means that lower energy consumption and potential cost reductions in CAPEX and OPEX can be achieved. Virtualized BBU pools can also be shared with operators, which makes it possible to offer Cloud RAN as a service.[19] Cloud RAN's co-operative radio provides higher spectral efficiency and since Cloud RAN is based on open platform, it opens the possibilities for new revenue sources with technologies such as Mobile Edge Computing (MEC). It is a network architecture to provide low bit-cost, high spectral efficiency and low energy consumption by utilizing cloud computing in the RAN.[4]

Cloud RAN technology is aiming to support the existing Radio Access Technologies from 2G to 4G and provide the foundation for the future 5G mobile networks [4][18]. This thesis focuses on Cloud RAN deployment in the LTE network. Cloud RAN will be usable in most typical RAN deployment scenarios such as macro and pico cells as well as indoor coverage. Additionally, other RAN deployments can work as complementary with Cloud RAN [4]. Competing technologies such as Small cells and Massive MIMO can also benefit the Cloud RAN deployment [19]. For example, Massive MIMO can potentially improve secrecy and energy efficiency of Cloud RAN [26].

2.4.1 Base Station Evolution Towards Cloud

Base stations consist of different functions that can be divided into baseband processing functionalities and radio functionalities. Figure 9 shows the different LTE base station functions that exist between the BBU and RRH. Traditionally base stations have been consisting of the baseband processing, but the Cloud RAN architecture aims to centralize baseband processing to locations such as BBU hotels and data centers. Centralized baseband processing allows dynamic resource allocation and better hardware utilization. RAN architectures and the base stations can consist of the following parts:

- **Remote Radio Head (RRH)** is located at the cell site and it provides the wireless signal coverage for the cell site area. RRH consist of antenna module, transmitter and receiver functions, power amplifiers, digital to analog conversion, analog to digital conversion and filtering (figure 9) [27]. RRHs connect UE to the mobile network with high rata rate radio frequency (RF) signals in the downlink and forward the baseband signal to the BBU for processing. [19][28]

- **BBUs and BBU Pool** include functions from layer 1 to layer 3, which are responsible for the baseband processing in the network (figure 9). BBU pool consists of co-located and co-operating virtualized BBUs serving a cluster of RRHs.[29] Depending on the used Cloud RAN architecture's level of centralization and functional split (chapter 2.4.2), BBUs can be located at cell sites, BBU hotels or data centers.

– **Fronthaul network** provides the connectivity between the BBU and RRH most commonly using optical fiber with Common Public Radio Interface (CPRI) [30] or Open Base Station Architecture Initiative (OBSAI)[31] .[19].

– **Midhaul network** provides connectivity between the BBUs and BBU pools using IP or MPLS interface. Midhaul network is required in Cloud RAN architectures where the baseband processing is split.

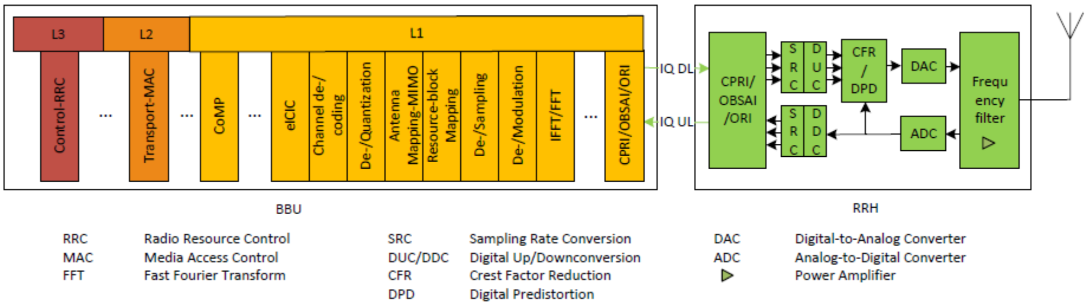

Figure 9: Base station functionalities divided between the BBU and RRH.[32]

From 1G network deployments to future radio network solutions of Cloud RAN, base stations have gone through three different architectures.[19]

Traditional base station, (figure 10a) used in 1G and 2G mobile networks, consist of base station equipment cabinet and the cell tower with the antenna module on top. Hardware equipment such as the power unit, baseband processing modules and radio functionalities are located inside the equipment cabinet. The equipment cabinet is in the close proximity of the cell tower and it is connected with coaxial cable to the antenna.[19]

Distributed base station (figure 10b) was introduced in the deployment of 3G networks and currently most of the base stations operate on this architecture. In this architecture, base station is separated into to BBU and RRH. This allows the BBU to be placed in more suitable locations, which can provide cost savings on rental and maintenance costs. RRHs are separate entities with their own power and cooling systems, which allows them to be placed on rooftops and up on poles to reduce costs for cooling and air-conditioning.[19] Separation of BBU and RRH was the first step towards BBU hotelling and centralization.

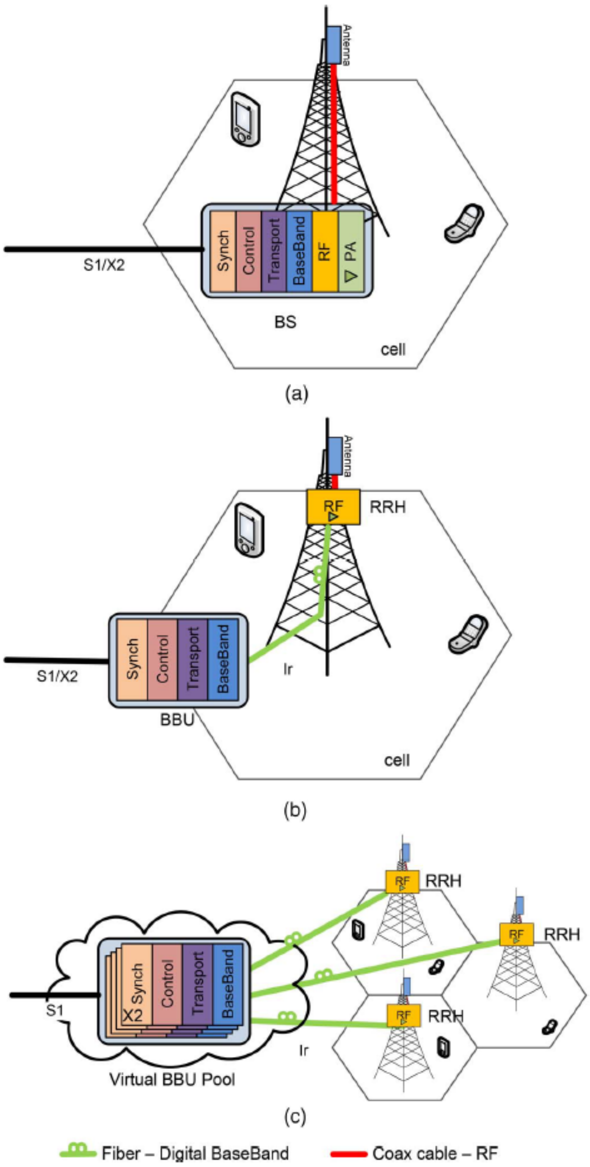

Figure 10: Base station architecture evolution. (a) Traditional base station in 1G and 2G. (b) Distributed base station in 3G and 4G. (c) Cloud RAN base station.[15]

Cloud base stations in Cloud RAN (figure 10c) take the next step in the separation of BBU and RRH components by centralizing BBUs to form a virtualized BBU pool. RRHs are co-located with the antennas at the cell sites and BBUs are centralized to BBU hotels or data centers. Virtualized BBU pools allow utilizing base stations more efficiently by allocating BBU resources on demand depending on the traffic load.[28] Fronthaul network connects RRHs to the BBU pool and the backhaul network connects the BBU pool to the mobile core network. Fronthaul connection between RRH and BBU pool is done with high bandwidth, low latency optical transport links to enable large data flows with low latencies between the entities.[19]

2.4.2 Cloud RAN Functional Splits

Cloud RAN was initially envisioned to have fully centralized baseband processing, where most of the base station's functions would be executed at centralized location and only the RF functionalities would be executed at the distributed cell sites. Fully centralized baseband processing can offer the best multiplexing gains, but it comes with strict bandwidth and latency requirements for the fronthaul network, which can offset the benefits. Fully centralized baseband processing relies on optical fiber for the fronthaul network to support high bandwidth and low latency requirements between the RRH and BBU. In fully centralized solution, transporting raw signal data on the CPRI/OBSAI interface between the RRH and BBU causes a heavy burden on the fronthaul in terms of required bandwidth and latency requirements. Therefore limiting the distance between the RRH and BBU. This has led into exploration of different functional splits for Cloud RAN. [33]

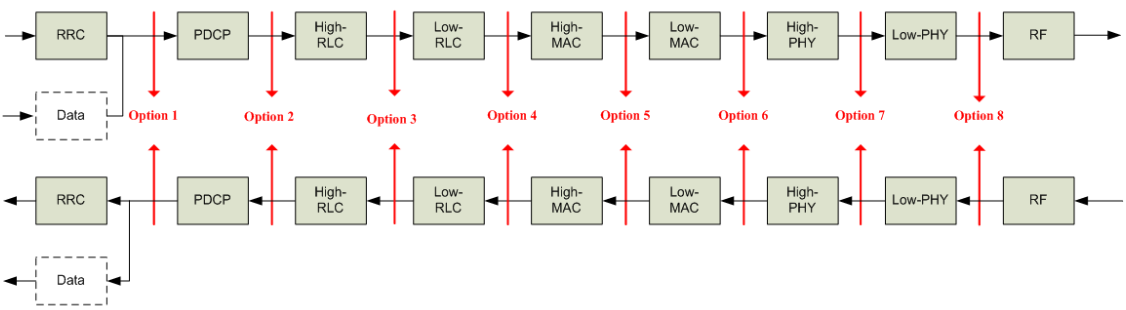

Figure 11: Different functional split options realized by the 3GPP. [34]

3GPP has studied the different possible functional splits consisting of options 1 to 8 for Cloud RAN, where the fully centralized solution is realized as option 8 (figure 11). Currently the most attractive functional splits are the options 2 and 3, which divide the functionalities between Non-Real-Time (NRT) and Real-Time (RT) functions. In the NRT-RT split, the NRT functions are executed at the centralized location and RT functions executed at the distributed sites.[34] Functions on the higher layers have more relaxed requirements for bandwidth and latency, which makes them more suitable to be centralized. On the other the hand, the lower layer functions require high bandwidth and very low latency and are therefore more attractive option to be deployed near the cell sites and RRH.[33]

Compromising functional splits such as the NRT-RT split have much more relaxed requirements compared to fully centralized solution, while still offering multiplexing gains.[33] It is also important to consider criteria such as backhaul and hardware capabilities, traffic demand and energy efficiency, when choosing the most optimal functional split for Cloud RAN [13]. Fully centralizing the baseband processing still remains as the end goal in Cloud RAN for achieving the best multiplexing gains.

2.4.3 Virtualization in Cloud RAN

Virtualization is one of the main concepts behind Cloud RAN together with centralizing the baseband processing. In Cloud RAN, virtualization is used to create the virtualized BBU pools that are operated on multiple Commercial off-the-shelf (COTS) servers.[29] Virtualization technology separates resources into virtual entities from the underlying physical hardware. This enables the dynamic allocation of resources such as memory, processing power and storage to be used by different applications. In virtual environment applications and functions operate on top of the Virtual Machines (VMs). VMs can share resources effectively to achieve better utilization, scalability and efficiency. VMs are managed and created by hypervisor, which is the virtualization layer on top of the hardware. [35] Adding or removing VMs depending on the traffic for example can improve the hardware utilization even further. In Cloud RAN VMs are responsible for handling the different BBU functionalities inside the virtualized BBU pool.

Network Function Virtualization

Network Function Virtualization (NFV) is a new network architecture concept that allows decoupling the network functions from the network hardware. The purpose of using NFV is to virtualize different network functions from proprietary hardware to Virtual Network Functions (VNFs) on COTS hardware.[36] In Cloud RAN (figure 12), NFV is used to virtualize network functions from the proprietary baseband hadware to virtualized BBU pools or often also called VNFs on the cloud servers. [29]. VNFs usually consists of different VMs, which have their own dedicated functions. VNFs are also independent of the underlying hardware and can therefore be easily created, moved, copied and deleted [36].

NFV is based on COTS hardware and, therefore, there is no vendor lock-in. The underlying hardware can be chosen from a variety of different traditional manufacturers. Use of hardware that supports open standards, platforms and services can enable faster innovation and reduce life cycles through software updates rather than hardware updates. Another benefit of having a virtualized server platform close to the mobile edge is the ability to host services such as content caching, which the end-users in mobile networks can experience as improved performance due to lower latency and faster access to content.[37] Available un-used server capacity can also be sold to parties that can benefit from computing on the mobile edge.

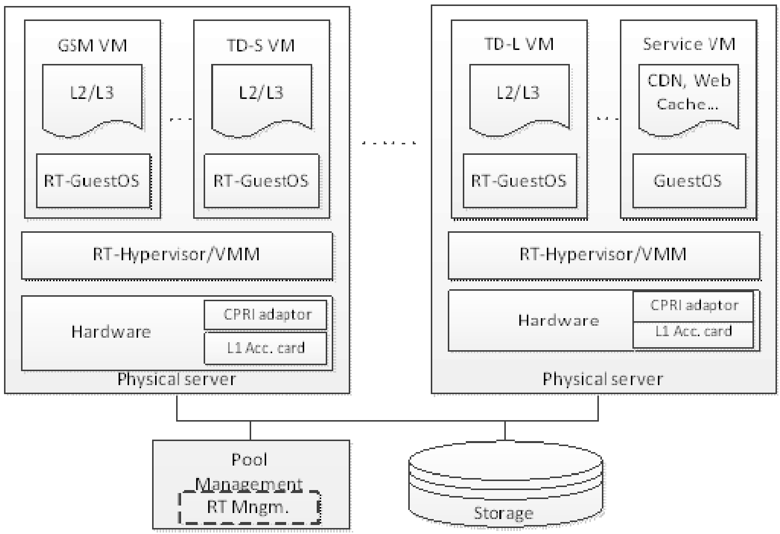

Figure 12: Example of virtualization in the Radio Access Network.[29]

Software Defined Networking

Software Defined Networking (SDN) is another important concept in virtualization of baseband resources in Cloud RAN. SDN decouples network control plane and forwarding plane, which enables programmability of the network control function and allows the underlying physical infrastructure to be abstracted from applications and services.[38]

In SDN architecture, software based centralized SDN controller is responsible for the network intelligence and maintains the overall view of the network. SDN controller simplifies the network structure and operations by managing the network from centralized location instead of having to deal with management of multiple vendor specific devices and protocols separately.[38] Another key feature of SDN is open interfaces between devices and controllers, which allows more flexibility for the operator to manage the network. Where NFV replaces the proprietary hardware with software running on top of COTS hardware, SDN replaces the standardized networking protocols with centralized control [36].

SDN architecture in Cloud RAN provides the ability to enable fast deployment of new software applications and dynamically adapt to the changing traffic patterns in the network. OpenFlow developed by Open Networking Foundation is common protocol used between network devices and controllers in SDN southbound interface.[39]

NFV and SDN in Cloud RAN

Cloud RAN can be seen as an application of NFV and SDN, where NFV enables the easy scalability and creation of new virtual BBU instances as needed and SDN acts as the connection between the virtual BBU instances and the radios that need them. Application of both NFV and SDN makes it possible to run different radio standards on the same BBU pool hardware platform and integrate their processing resources into the whole BBU pool. NFV and SDN are not dependent on each other and can be applied to the network separately, but they complement each other and integrating both technologies can provide the scalability and flexibility required from the future mobile networks.[38]

2.4.4 Cloud RAN Architectures

Two different Cloud RAN architectures featured in this thesis are Cloud Distributed-RAN (Cloud D-RAN) and Cloud Centralized-RAN (Cloud C-RAN). Both of these architectures are based on the NRT-RT functional split, where the RT functions are operated on the proprietary hardware and NRT functions on the COTS cloud servers. Therefore the BBUs in Cloud RAN can be thought of consisting of both, the proprietary baseband hardware and COTS cloud servers.

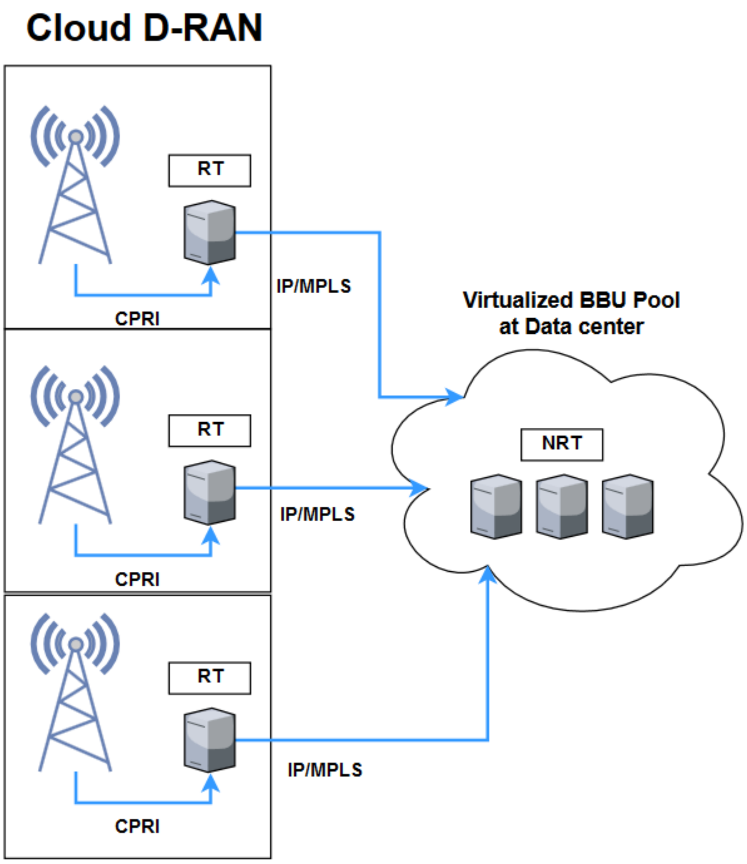

Figure 13: Cloud D-RAN architecture.

In Cloud D-RAN (figure 13) the proprietary baseband hardware that is responsible for the RT functions is located near the RRHs. Therefore, cell sites remain similar to current D-RAN architecture. COTS cloud servers responsible for NRT function processing are centralized to data centers. Midhaul network in Cloud D-RAN uses Ethernet links with IP/MPLS interface to connect the BBUs located at cell sites and the virtualized BBU pools. Performing part of the baseband processing at cell sites removes the need to use CPRI/OBSAI interface in large fronthaul network that requires high bandwidth and low latency to connect RRHs and BBUs. This means that there are much more relaxed requirements for connecting the cell sites and data centers [33]. Even though part of the baseband processing is located at cell sites, Cloud D-RAN benefits from centralization and virtualization [40].

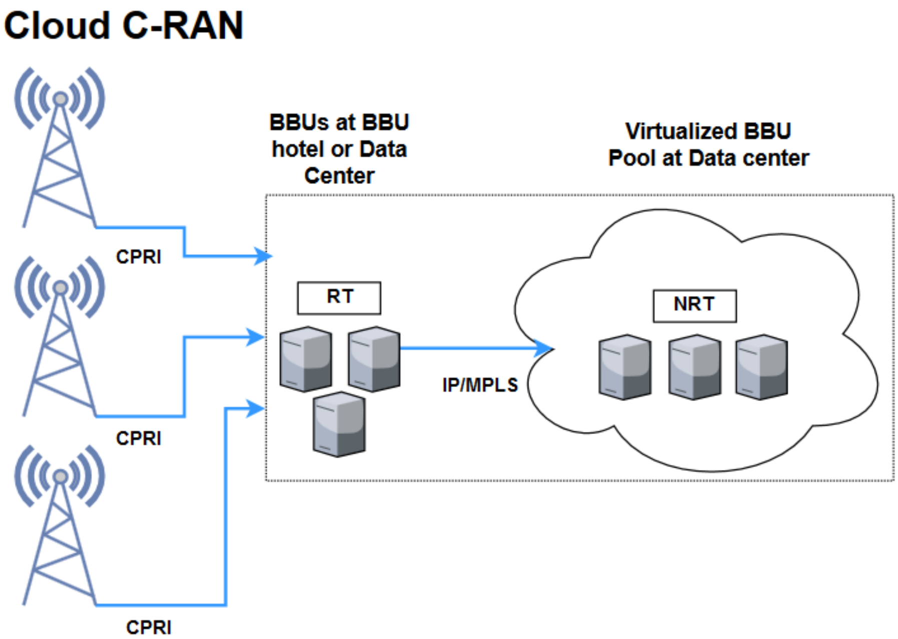

Figure 14: Cloud C-RAN architecture.

In Cloud C-RAN (figure 14) cell sites are consisting only of RRH and the baseband processing is moved to BBU hotels or data centers depending on the distance between the RRH and BBU. The maximum feasible distance between the RRH and BBU is dependent on the maximum allowed latency between the two, which is determined by the hardware's processing time and light's speed [41]. For longer distances BBU hotels are required to host the proprietary hardware that cannot be centralized to data centers. Optical fiber is required in the Cloud C-RAN fronthaul network to connect RRHs and BBUs. This means that only areas with existing fiber access may be chosen for Cloud C-RAN deployment or investments are required to construct the optical fiber network [13]. On the other hand, centralizing the proprietary baseband hardware means that the hardware can be used more efficiently, upgrading and managing the hardware is easier and less space is required at cell sites than in the distributed architecture.

2.4.5 Mobile Edge Computing with Cloud RAN

Mobile Edge Computing (MEC) aims to position computing and storage resources on the RAN edge to improve content and application delivery for 4G and 5G network users. [42] MEC environment is characterized by very low latency, high bandwidth and real-time access to radio network information, which can enable deployment of new applications and services such as IoT, video analytics and connected cars. Deploying distributed data centers capable of content caching and processing is key for achieving low latency in the mobile networks, which is also one of the defining characteristics of 5G.[43]

MNOs implementing MEC can create a new ecosystem and re-position themselves in the value chain by opening their networks for authorized third parties to develop new applications and services for the network users.[43] For example, in large events such as formula races MEC can be used to enhance the experience by allowing users attending the event to gain additional real-time information and video streaming directly to their mobile devices. Users can watch real-time video streams on the cars they want to follow from different angles, see and hear the expressions of drivers when they are getting passed by and follow lap times and pit stop times in real-time. These real-time applications have the potential to enhance experiencing events to completely new level and add more value to the overall event experience. [44] MEC's features can also help to differentiate the MNO's solutions from Over-The-Top (OTT) players and offer competitive edge over them. MNOs can launch their own OTT platforms with MEC to create new services with better user experience and features than the OTT players can offer.[43]

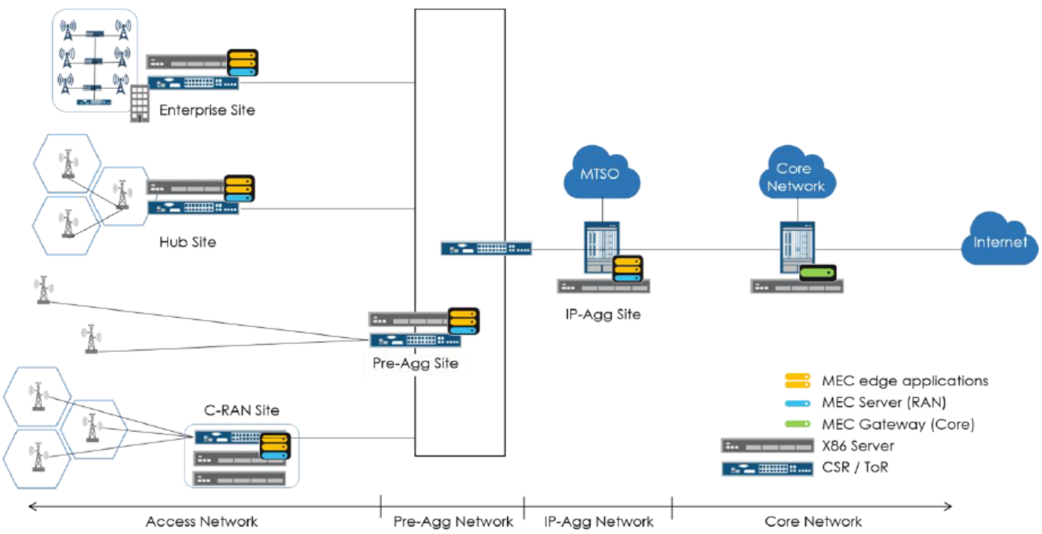

Figure 15: MEC deployment in Cloud RAN architecture.[42]

Cloud RAN offers a great opportunity for MNO's to implement MEC, which might not be an obvious connection at first as Cloud RAN is aiming to centralize baseband processing, while MEC requires distributed architecture to work efficiently.

Nevertheless, there is synergy between them as Cloud RAN operates on COTS servers. These servers can also be used to host MEC applications that are running on the centralized edge instead of the core network. Additionally, it is also unlikely that MNO would centralize all of their baseband processing to a single data center, because in case of failure in the data center the whole network would go down. Multiple data centers or other aggregation points are required to guarantee high availability for the mobile network. Therefore Cloud RAN COTS servers would be distributed in the network, which offers ideal opportunity to deploy MEC.

Cloud RAN can enable large scale outdoor MEC deployment (figure 15), which is not only limited to large events. Including MEC applications into Cloud RAN COTS servers that are located in multiple data centers around a city can offer great opportunities for new applications and services. Real-time video analytics in cities can be used for analyzing traffic, local businesses can begin using augmented reality applications and content caching to offer better Quality-of-Experience for their customers and active device location tracking can be used for mobile advertising [43]. MEC can bring additional revenue sources for MNOs and Cloud RAN can make it possible by providing the infrastructure with COTS servers on the mobile edge.

2.4.6 Advantages of Cloud RAN

Cost savings in CAPEX and OPEX: Centralizing and virtualizing the baseband processing and moving site support equipment to BBU hotels and data centers reduces the amount of equipment needed at the cell sites, thus reducing CAPEX. Use of cloud computing and centralization also makes upgrading the network capacity easier and cheaper. Hardware upgrades can be made by adding more servers to centralized locations, instead of installing new hardware to every cell site separately. Cell sites in centralized architecture have much simpler functionality and require less power when the baseband processing is centralized, but the number of cell sites will remain the same. RRHs can be attached to places like roof tops and poles, where there is no need for cooling and they can operate with minimum site support and management. Therefore the network construction speed is quicker and cost savings can be achieved with reduced power consumption, less equipment and better resource utilization. Centralization will also reduce site rental and operation and management costs of the network.[4]

Capacity and spectral efficiency improvements: Base stations in Cloud RAN's BBU pool can work together and easily share signaling data, traffic data and Channel State Information (CSI) of active UE's in the network due to the nature of centralized baseband processing. Cloud RAN also enables easy implementation of joint processing and scheduling, which can help to mitigate inter-cell interference and improve spectral efficiency. For example, Co-operative Multi-Point processing technology (CoMP in LTE Advanced) can be implemented in Cloud-RAN infrastructure with ease.[4]

Adaptability to non-uniform traffic: Traditional base stations are dimensioned for peak capacity to be able to provide network coverage to users even if a large event like a music festival is being held in the area. This means that a lot of processing power is going to waste during the time when the base station is not running at full capacity. With centralization and virtualization of baseband processing in Cloud RAN, BBU pools can handle traffic from multiple cell sites and the dimensioning can be done for the average capacity of all connected base stations. In Cloud RAN the required baseband processing capacity in BBU pools is expected to be much less than the sum of capacity requirements of individual base stations.[19] This is due to pooling gain, as in BBU pools the peak capacity requirement does not have to be taken account in every single base station and base stations can be dimensioned for the average capacity. Thus, the resource utilization is much better than in the current RAN architecture.

Energy efficient: In centralized architecture baseband processing hardware, air conditioning and other site equipment are centralized and better utilized, thus power consumption can be expected to decrease. Cloud RAN's virtualized BBU pools also enable better resource utilization during low traffic hours, when it is possible to scale down the number of virtualized BBUs to reduce power consumption. Co-operative radio technology decreases the distance between the UE and RRHs due to reduced interference among RRHs and thus allowing higher cell density. Reduced distance between UE and RRH means more energy efficient signal transmission, prolonged UE battery life and overall decreased power consumption in the RAN.[4]

Smart Internet traffic offload and services on the edge: Centralization of baseband functionalities provides a point for traffic offload and content management to reduce the traffic on the operator's core network that is caused by growing Internet traffic from smartphones and other devices. Benefits of this are reduced back-haul and core network traffic and costs, reduced latency for the users, which means better Quality-of-Experience for the users.[45]

2.4.7 Challenges of Cloud RAN

RRH and BBU pool connection with optical fiber: In fully centralized Cloud RAN architecture fronthaul must carry a significant amount of baseband sampling data in real-time for several kilometers over an optical link between the RRH and BBU pool. Centralized BBU pool should be able to support 10-1000 cell sites, which is why such a huge amount of data is required to be carried in the fronthaul network CPRI interface. Additionally, the fronthaul network must support strict requirements for transport latency, latency jitter and it must be cost efficient.[4] Building a large optical fiber fronthaul network to support Cloud RAN can be expensive, which can make Cloud RAN unattractive option compared to other RAN architectures in terms of costs.

Data compression techniques in the fiber and introduction of new transport nodes for fronthaul transmission can be used to reduce the fiber consumption. Thus, reducing the heavy burden of the fronthaul.[29] Moving towards more distributed functional split can also be considered as an option to reduce burden on the fronthaul [33].

BBU pool interconnection and clustering: Centralized architecture aggregates large number of BBUs to same physical location, which requires special attention to guarantee network security and reliability. High system reliability is important in centralized architecture to recover from unit failures and errors and to allow flexible resource sharing between BBUs. This can be achieved with high bandwidth, low latency, cost efficient switch network with flexible topology that interconnects BBUs inside the BBU pool. Switch network allowing flexible routing of digital baseband processing signals between any RRH and BBUs can prevent the failure of the entire system in case of single BBU failure happens.[4]

Optimal cell clustering in BBU pools is required for achieving multiplexing gains and preventing the BBU pool and transport network from overloading. The BBU pool should consists of cells from different traffic areas such as office, residential and commercial areas to optimize the number of active BBUs and RRHs in the BBU pool.[19] Optimal cell placement in BBU pools should maximize the BBU pool resource utilization, by combining cells so the total sum of the traffic does not differ much at any time.

Advanced co-operation with transmission and reception: Current LTE networks suffer from much more severe interference problems than 2G or 3G networks, due to large number of deployed small cells to achieve higher data rates. Important aspects of Cloud RAN are also to improve spectral and energy efficiency and reduce Inter Cell Interference (ICI), which can be done with collaborative radio and joint signal processing techniques. CoMP Joint Transmission (JT) algorithms are viable in Cloud RAN and can be used to improve the system performance in the previously mentioned aspects.[29]

CoMP JT algorithms require end-user data, uplink and downlink information to be shared between Cloud RAN base stations. Information such as end-user data packages, UE channel information and cloud base station's scheduling information that might be shared can require real-time processing. This means that Cloud RAN base station interfaces should be designed to support high bandwidth and low latency to achieve the real-time processing requirements with low backhaul transmission delay and overview.[4] Combining CoMP JT with cell clustering algorithm will reduce complexity of scheduling, which means that well designed scheduler in Cloud RAN can impact the spectral efficiency [19].

Base station virtualization: One of the biggest challenges of base station virtualization is meeting the strict demands of mobile signal real-time processing constraints in the virtual environment. Implementing dedicated hardware accelerators can help achieving the requirements. Successful base station virtualization is a lot harder in mobile networks than in standard IT data center setting due to extremely strict requirements for real-time processing in wireless communication. For example, Time-Division Duplexing Long Term Evolution (TDD-LTE) system requires that an ACK/NACK must be sent back to UE or base station under 3ms after the frame is received. Standard IT data centers cannot meet such requirements so the base station virtualization needs special optimization and design to meet the requirements of the wireless mobile communication.[29]

3 Feasibility of Cloud RAN

Mobile networks have evolved a lot in the last 10-15 years. Features such as mobile gaming, live video streaming and mobile payments were limited or not even possible some time ago, but now they are considered as core features in our mobile phones. During 3G and 4G the mobile networks evolved a lot in terms of bandwidth and latency and at the same time the capabilities of mobile phones have improved a lot as well. Better mobile networks and the rising adoption rate of smart phones has led into increasing adoption of advanced multimedia mobile applications, which then has contributed to increased mobile data traffic [1].

Future of communications will be very different from what it is today. Forthcoming of IoT, M2M communications and 5G will change the way we experience communications and it will most likely mark the next evolution in the mobile communications. These technologies will enable us to connect everything from our shoes to cars, and even cities to the Internet. For this connected world to become reality, the mobile networks and especially the RAN needs to be able to provide better bandwidth, latency and connect the massive number of devices economically.

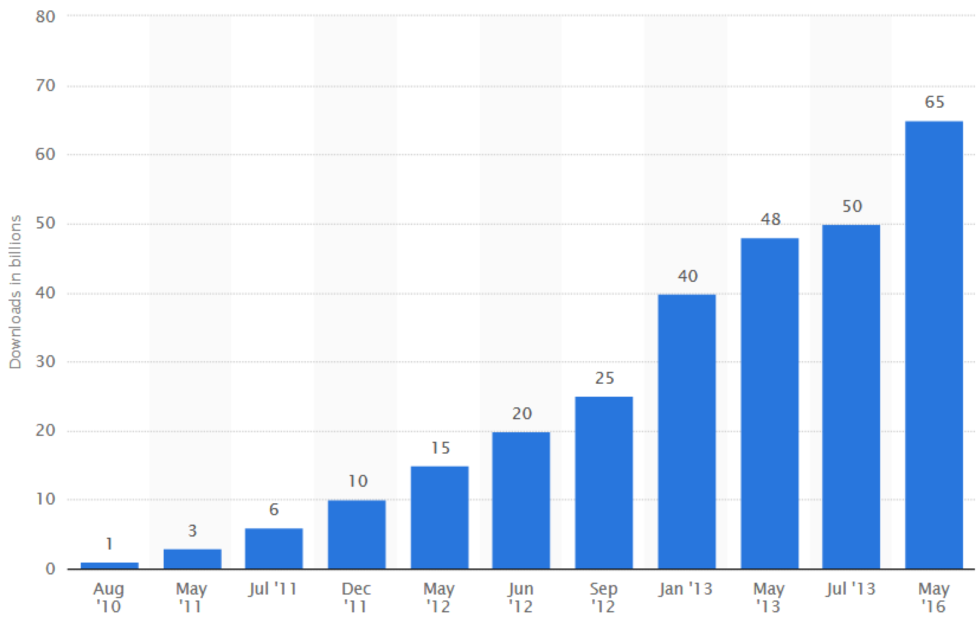

Figure 16: Number of mobile application downloads in Google's Play store during 2010 and 2016.[46]

Mobile networks are also driving innovation and growth across industries. Mobile applications have become part of our everyday lives as the number of downloaded applications in Google's Play store has increased from 1 billion to 65 billion during 2010 and 2016 (figure 16) [46]. OTT applications such as the messaging application WhatsApp has reached massive popularity among users with its 3,6 billion users at the end of 2016. OTT applications have been cannibalizing on MNO's core services, which has led MNOs into increasing the collaboration with technology start-ups to develop more open models of innovation. According to GSMA, MNOs tripled their investments into technology start-up companies between 2014 and 2015 [3].

Digital platforms are driving businesses to reinvent their business models to provide new and innovative digital products and services. Broad range of customer focused sectors are innovating how business is done through digital platforms and looking forward to adding intelligence into their products. Industrial companies are also moving towards digital platforms with cloud computing by putting automation, real-time and big data analytics into cloud environments. These applications combined with new scalable connectivity technologies such as Low Power Wide Area (LPWA) solution will be essential for unlocking the potential of Industrial Internet of Things.[3]

RAN is essential for providing wireless Internet connectivity for devices and sensors in the network, enabling new applications and services in different industries and bringing new business opportunities with technologies such as IoT and MEC. Cloud RAN is ranked among the top five technology disruptors and it is often brought up as the solution for the rapidly increasing traffic volumes and number of connections [12]. Cloud RAN's cloud-based architecture enables radio networks to be opened up for new services and applications through open Application Programming Interfaces (APIs) and access channels. This allows open innovation and collaboration with different businesses on the mobile edge, which will help MNOs to capture more value, differentiate and achieve a competitive edge over their competition. Cloud RAN can be the solution for MNOs to meet the demands of increasing number of connected devices economically, providing new business opportunities through IoT and MEC and accelerating innovation and growth across industries by providing open platforms on the mobile edge.

This chapter evaluates Cloud RAN's feasibility by explaining the challenges of today's RAN architecture, presenting the possible business benefits of Cloud RAN and featuring the motivation behind the cost modeling of Cloud RAN.

3.1 Challenges of today's RAN

Challenges of today's RAN architecture are mostly related to growth and expansion of the network due to increasing number of connected devices and mobile network traffic. Existing wireless network infrastructures were made for handling voice traffic and the change to data oriented traffic has overwhelmed the networks especially in densely populated areas.[45]

High costs: MNOs must significantly increase their mobile network capacity in order to satisfy the growing mobile data traffic demands of the users. Building and operating a dense network infrastructure is expensive with the current RAN architecture [47]. At the same time, the MNOs are experiencing high saturation levels, rapid technology changes and declining ARPUs in their highly competitive marketplace [3]. This presents a difficult challenge to provide the required network capacity, while still maintaining reasonable profitability. The current RAN architecture will become challenging to keep competitive in the future if the traffic is expected to grow rapidly. On top of that the current RAN is not prepared for 5G deployment [4]. Future RAN must be able to satisfy the demands of growing data traffic with lower costs and be ready for 5G.

High energy consumption: One of the easiest ways to increase network capacity to meet the traffic demands of growing network for MNOs have been increasing the number of base stations. For example, China Mobile doubled the number of its base stations in 5-year period to provide better network capacity and coverage. However, doubling the number of base stations also doubles their energy consumption and carbon dioxide emissions in the network. MNOs must plan for more energy efficient network architectures to reduce the energy consumption and carbon dioxide emissions.[4]

Base station over-dimensioning: The nature of wireless networks allows subscribers to move freely in the network, which creates changing traffic loads on certain areas. For example, during business hours central office areas experience significant growth in traffic, when large numbers of subscribers move from home to work and when the subscribers commute back to home the traffic moves with them to the residential areas. This creates poor base station utilization rate as the base stations have to be dimensioned for the peak hours and are therefore often over-dimensioned for the capacity that is needed most of the time. Outside the busy hour period, resources are not utilized in the best possible way in the current RAN architecture. Sharing base station resources between different areas is a good way to utilize these resources more efficiently and reduce the required base station over-dimensioning.[45]

Closed solutions: Current RAN is based on proprietary hardware and software, which makes the network difficult to program, lacks the agility to dynamically meet the requirements of new applications and building a cost effective multi-vendor infrastructure is difficult due to interoperability issues. Proprietary hardware and software also limits open innovation for new services and applications due to closed access points and platforms.[47]

Spectrum availability and Interference: Acquiring new available spectrums for the radio network is difficult due to their scarcity and regulatory limits. LTE is designed with the frequency reuse factor of one to improve spectral efficiency in the network, which differs from previous generations that had the frequency reuse factor more than one. Since the cells in LTE operate on the same frequency band there is no avoiding the inter-cell interference between the neighbor cells, which leads into reduced throughput performance. In LTE it is also common to have coverage overlapping with neighboring cells, which makes reducing co-channel interference important especially in large scale networks. Interference coordination technologies such as Inter-Cell Interference Coordination (ICIC) and CoMP have been developed to solve these problems, but the benefits gained from them are limited in the current distributed RAN architecture. On the other hand, centralized Cloud RAN architecture can implement these technologies more successfully and reduce the inter-cell interference, increase the networks throughput performance and achieve better spectral efficiency.[4]

3.2 Cloud RAN's business benefits

Cloud RAN aims to tackle the challenges of the current distributed RAN architecture and its business benefits can be recognized on the following areas.

Cost effective: One of the corner stones of Cloud RAN is the ability to scale resources efficiently to meet the traffic and service growth demand, which removes the need for over-dimensioning cell sites and the network for the peak traffic demand [20]. New mobile applications based on augmented or virtual reality such as Pokémon GO can cause an unexpectedly heavy burden on the RAN due to application's heavy signaling and uneven traffic demand. Cloud RAN provides high scalability and dynamic resource management in the network for heavy signaling increase and uneven traffic demand.[48]

Cloud RAN's more effective scaling for traffic growth and large-scale pooling gains can bring savings to both CAPEX and OPEX. [20] Improving economics of RAN operation is among the most important drivers for MNOs together with system efficiency, ability to improve scalability and resource utilization [42]. Cloud RAN aims to meet the demands of rapidly increasing mobile data traffic with low costs.

Efficiency: Combining the use of licensed and unlicensed spectrums in Cloud RAN makes it possible for MNOs to maximize the use of their radio technology assets through multi-connectivity technologies such as inter-site carrier aggregation, LTE Dual Connectivity, LTE and Wi-Fi link aggregation (LWA), Licensed Assisted Access (LAA) and the upcoming 5G-LTE Wi-Fi from the cloud. Cloud RAN can also enhance peak data rates and improve spectral efficiency by delivering and coordinating key features from the cloud.[20]

Centralizing and virtualizing the baseband processing in Cloud RAN improves the utilization of baseband processing hardware with more effective use of resources. Therefore, Cloud RAN also has the potential to reduce the energy consumption in RAN and provide an eco-friendlier infrastructure.

Open Innovation: Cloud RAN's cloud-based architecture is flexible and scalable and it has the agility to meet the demands of new applications and services. The use of COTS cloud servers offers the ability to build cost effective multi-vendor solutions with better compatibility. Cloud RAN supports open innovation on the mobile edge with easy MEC deployment on the same cloud servers.

Open APIs and access channels can be used to develop new business opportunities, applications, services and plug-ins that can be smoothly implemented into RAN [20]. MEC enables content caching and extremely low latency capability for services and applications that can provide better end-user experience than the OTT players [49]. Ability to provide applications and services on the mobile edge and the support for massive IoT has the potential to pique interest of many different businesses from consumer and industrial sectors. Collaborating with these businesses can help MNOs to excel their influence in their business ecosystem, differentiate from the competition and allow them to capture more value.

5G and IoT readiness: 5G is estimated to be worth 220 billion euros for the MNOs, and respectively the Internet of Things is estimated to be 484-billion-euro business opportunity [49] [50]. Cloud RAN will bring the foundation for 5G with improved spectral efficiency, peak data rates and lower latency. Existing LTE-A, transport and cloud infrastructure can be re-used for fast deployment of new 5G access. Cloud RAN's multi-connectivity layer can be used to host 5G, LTE and Wi-Fi Cloud based multi-layer access network to save costs on individual RAN deployments. Multi-layer access network in 5G will enable new Internet of Things services that requires high reliability and low latency.[20] Cloud RAN's ability to support rapidly increasing number of devices and improved features such as lower latency will be essential for unlocking the true potential of IoT.

3.3 Cost modeling of Cloud RAN

Motivation for cost modeling of Cloud RAN is to analyze the business benefits on the cost effectiveness and efficiency of Cloud RAN, presented in the previous section. The goal is to recognize the cost positions of the different RAN architectures depending on the given use case scenario. Analyzing the TCO of the existing D-RAN and C-RAN architectures with the future's Cloud D-RAN and Cloud C-RAN architectures is essential in terms of Cloud RAN's feasibility. Cloud RAN is expected to provide better scalability for the growing mobile traffic and reduce costs in the RAN. If Cloud RAN cannot reduce the RAN's TCO, MNOs might be reluctant to invest into it.

Cost modeling of Cloud RAN deployment and its potential cost benefits have been covered slightly in the recently published literature. "Cost modeling for SDN/NFV based 5G Networks" [51] presents a cost model comparing the network costs of traditional RAN architecture and SDN/NFV based Cloud RAN architecture in 5G. "Analysis of CAPEX and OPEX benefits of wireless access virtualization" [52] analyzes the cost benefits of virtualized RAN to the traditional RAN. Both of these models use macro level approach for the modeling that does not consider the network traffic or the used the hardware, which has significant impact on the benefits of Cloud RAN deployment and the TCO. Thus, these results can give a general idea about the possible benefits as the analysis is done only on the macro level.

Cost modeling of Cloud RAN in this thesis aims to provide more detailed view on the cost positions of the different architectures in the given scenario. The model's network dimensioning is based on real baseband processing hardware specifications and existing and upcoming RAN solutions to give a more realistic view on the feasibility of Cloud RAN in terms of costs. Cost modeling of Cloud RAN is explained in more detail in the chapter 4, chapter 5 consists of the analysis of modeling an example case and chapter 6 consists of conclusions from the results.

4 Cloud RAN TCO Model

Cloud RAN TCO model evaluates the cost position of D-RAN, C-RAN, Cloud D-RAN and Cloud C-RAN architectures to solve the most economical RAN deployment option for the given subscriber traffic and network configuration. The model is based on network dimensioning, mathematical modeling and cost calculations. Network dimensioning is used to calculate the number of base station equipment, sites and other network elements as well as their configuration. Mathematical modeling is used to estimate required site visits for baseband capacity upgrades and required space for the network equipment to calculate site rental costs. The cost calculations part calculates the CAPEX and OPEX from the results of the network dimensioning.

The following assumptions are made in the model:

- Functional split used in the modeling of Cloud RAN is the NRT-RT split.
- D-RAN architecture is assumed to be the current architecture that most of the MNOs are using in their Radio Access Network.
- In large networks, it might not be possible to completely centralize the network in C-RAN and in Cloud C-RAN deployments, because some cell sites might be located too far to be centralized. Therefore in the model, C-RAN and Cloud C-RAN networks have been assumed to be consisting of D-RAN or Cloud D-RAN as well.
- CAPEX calculations do not include RF hardware as it has been assumed to remain the same in all the architectures.
- The model does not include the construction cost of new cell sites or the costs for building or renting a fiber network.
- In practice cell sites can have different cell configurations, but in the model all cells are remapped into equivalent identical cells thus creating a homogeneous area.

The model follows the flow chart in figure 17 and it can be divided into three parts, which are the model input data, network dimensioning and the cost calculations. First part of the model is the model input data, which consist of required user input data and secondary data for the network dimensioning and cost calculations. This includes the cell site data, network configuration data, hardware and software data (HW & SW data) and financial data. Second part of the model is the network dimensioning, which consists of baseband module (BBM) dimensioning and cloud server dimensioning to calculate the required network hardware and software licenses. The third part is the cost calculations, where the CAPEX and OPEX are calculated for the architectures. Output of the model is TCO analysis of the evaluated RAN architectures for the given time period of 1-5 years.

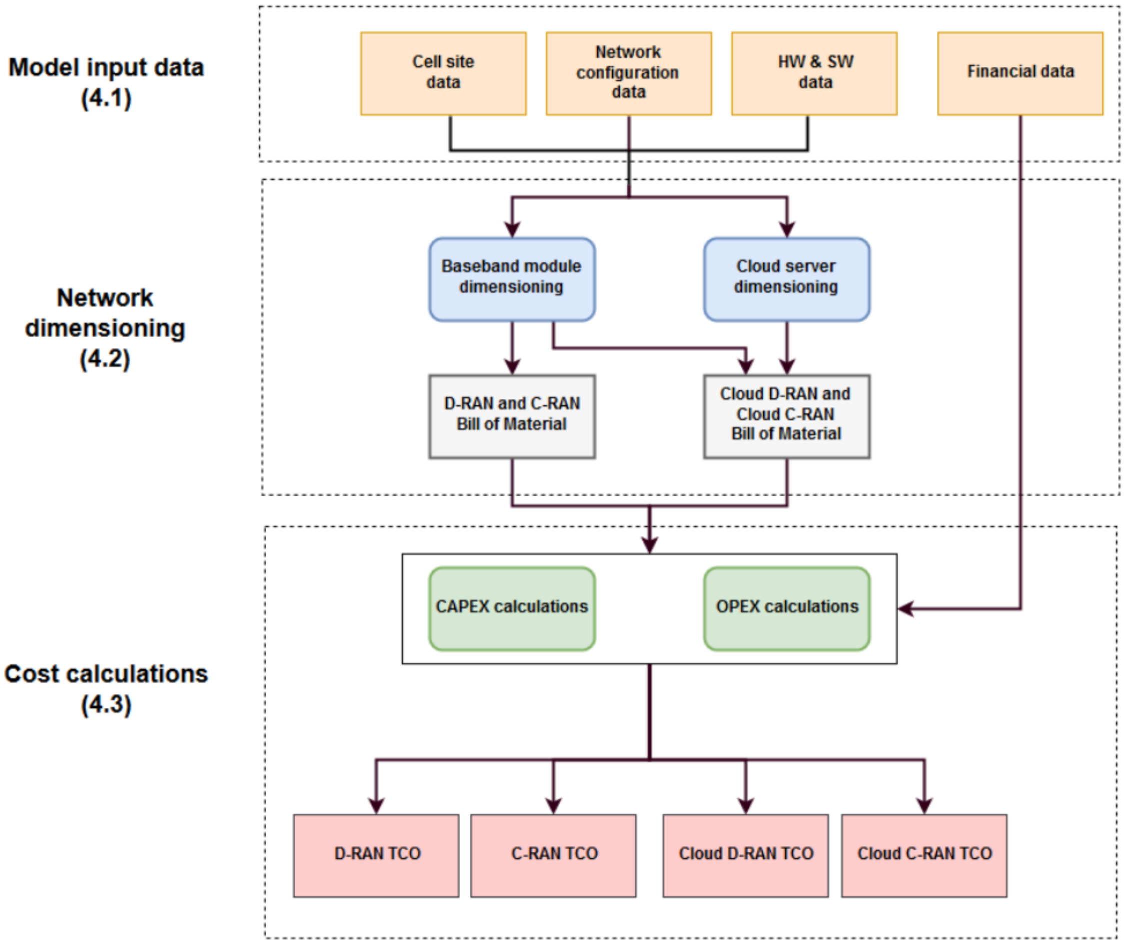

Figure 17: Cloud RAN TCO model structure.

This chapter will describe the created Cloud RAN TCO model, by explaining first the different input data sets and then proceeding to explain the network dimensioning and the cost calculations for calculating the TCO figures for the evaluated architectures.

4.1 Model input data

Four different data sets are required as inputs for the Cloud RAN TCO model calculations. The data sets are cell site data, network configuration data, HW & SW data and financial data.

4.1.1 Cell site data

This model uses some input data from the cell site configuration and traffic. The data is collected from the user (MNO) and entered as an input to the model. Cell site data (table 1) specifies parameters such as the cell site cell configuration, data throughput and C-plane traffic at cell sites.

Table 1: Cell site data

Name	Input parameters	Info
Cell configuration	Number of – 5/10 MHz 2x2 MIMO – 5/10 MHz 4x4 MIMO – 15/20 MHz 2x2 MIMO – 15/20 MHz 4x4 MIMO cells located at the cell site.	Used to calculate total cells in the network and peak downlink and uplink throughput per site. Required for all the cell limitation and cell connectivity calculations in the network dimensioning.
Data throughput	– Average busy hour throughput per site – Busy hour share of the traffic per day on site – Share of downlink and uplink traffic throughput	Used to calculate data throughput volumes for Cloud RAN and daily data volume per site. Provides the data for throughput requirement calculations in the cloud server dimensioning.
C-plane traffic	– Maximum Connected User Equipment (CUE) per site – Average CUE per site – C-plane load per CUE	CUE per site indicates the number of connected devices per site. C-plane load per CUE states the signalling frequency of these devices. CUE per site is also used to calculate the number of RRC connected users in the network. These parameters are used for the network dimensioning requirement calculations.

4.1.2 Network configuration data

Network configuration data (table 2) specifies the overall network structure for the different architectures. Network configuration data quantifies the overall need for the networking hardware and software licenses by the different RAN architectures. This data is collected from the user (MNO) and entered as user input to the model.

Table 2: Network configuration data

Name	Input parameters	Info
Network configuration	– Total number of cell sites – Number of D-RAN and C-RAN type configured sites in centralized architectures	Key parameters in the network dimensioning for calculating the overall network hardware and software licenses required by the different architectures.
Centralization	– Number of BBU hotels and data centers – Number of cell sites served by each BBU hotel or data center	Required for dimensioning the needed network hardware in centralized architectures.

4.1.3 HW and SW data

HW and SW data (table 3) consist of secondary data about the used network hardware and software, specifications, performance, capabilities and capacity limits.

Table 3: HW and SW data

Name	Input parameters	Info
BBM data	– Cell connectivity and C-plane capacity limits for the BBMs – BBM maximum and minimum configurations per cabinet – BBM unit energy consumption.	Used for BBM dimensioning and calculating energy consumption in the network
Cloud server data	– Number of CPUs per server – CPUs required by the different VMs – VM and VNF capacity limits for dimensioning requirements – Estimated pooling gains for Cloud RAN	Used for the VM, VNF and cloud server dimensioning .
Software licenses	– Active RF, baseband and Cloud software licenses	Used for calculating the costs for the required network software licenses.

4.1.4 Financial data

Financial data (table 4) includes prices for the used hardware and software licenses to calculate the CAPEX, costs for operating the network to calculate the OPEX. Data is based on secondary internal and external data.

Table 4: Financial data

Name	Input parameters	Info
CAPEX items	– Network hardware and software license prices	Required to calculate the network CAPEX.
OPEX items	– Price of energy – Rental costs for cell sites, BBU hotels and data centers – Cell site visit costs for hardware upgrades and installations – Network OAM costs per cell site	Combined with network configuration and network dimensioning bill of material to calculate the network OPEX.

4.2 Network dimensioning

The evaluated network architectures can be divided into two categories, which are the existing bare-metal based architectures D-RAN and C-RAN and the future Cloud RAN based architectures Cloud D-RAN and Cloud C-RAN. Network dimensioning for the bare-metal architectures consists only of the BBM dimensioning, as it is completely based on the proprietary hardware. Network dimensioning for Cloud RAN architectures requires BBM dimensioning and the cloud server dimensioning as both proprietary and COTS hardware is used. The required RF, baseband and cloud software licenses are calculated using the subscriber traffic and network configuration data.

Output of the network dimensioning is bill of material for each year consisting of all the required baseband processing hardware and software licenses in bare-metal and Cloud RAN architectures.

4.2.1 Baseband module dimensioning

First part of the network dimensioning is calculating the required BBM hardware in the different architectures. In the distributed architectures, the BBM hardware is located at cell sites and thus the dimensioning is done on a per cell site basis. In the centralized architectures, the hardware is located at Points-of-Presences (PoPs), which includes BBU hotels and data centers, and the dimensioning is done for these PoPs instead of cell sites. BBM hardware is assumed to be consisting of three different units, which are the capacity plug-in unit (CAP), the common plug-in unit (COM) and the cabinet unit (CBN). The BBM units are described in more detail in the table 5.

Table 5: BBM unit function description

BBM unit	Function
Capacity unit	Provides the baseband processing capacity. CAP consists of two cell sets that can be used individually for the baseband processing. The number of cells a single CAP unit can support is dependant on the cell's bandwidth and MIMO setting. Higher bandwidth and MIMO require more processing power, thus more CAP units will be required. In C-RAN architecture, the CAP units have less available processing capacity for cells due to C-RAN requiring more processing power to turn interference from neighbor cells into useful traffic.
Common unit	Responsible for processing the C-plane traffic and linking multiple BBMs together. COM includes transport and centralized control functions for the supported radio access technologies as well as antenna data routing.
Cabinet unit	Cabinet for hosting the CAP and COM units. CBN connects the underlying CAP and COM units to each other and provides backplane for the internal communication, as well as air conditioning for the underlying units.

BBM dimensioning follows the chart in figure 18 and it is done for the CAP, COM and CBN units separately. BBM dimensioning uses data from the HW & SW, cell site and network configuration data sets for the calculations. HW & SW data provides the data about BBM hardware capacity for cell connectivity and the C-plane traffic, cell site data provides the cell site's cell configuration and C-plane traffic and network configuration data provides the information about the network structure (number of cell sites, BBU hotels, etc.) to do the dimensioning for the whole network.

BBM dimensioning considers the cell configuration and the C-plane traffic on sites in the calculations, which form the two dimensioning requirements for the BBM units that are explained in tables 6 and 7. Additional limitations exist from the RF ports, but they are not considered in the model. For Cloud RAN, the C-plane traffic requirement can be ignored in the BBM dimensioning, because the C-plane traffic is mostly handled by the cloud servers and it is therefore not a limiting factor in the Cloud RAN BBM dimensioning. Also, due to pooling gains in the cloud, average traffic per cell site can be used in the calculations instead of peak traffic per cell site.

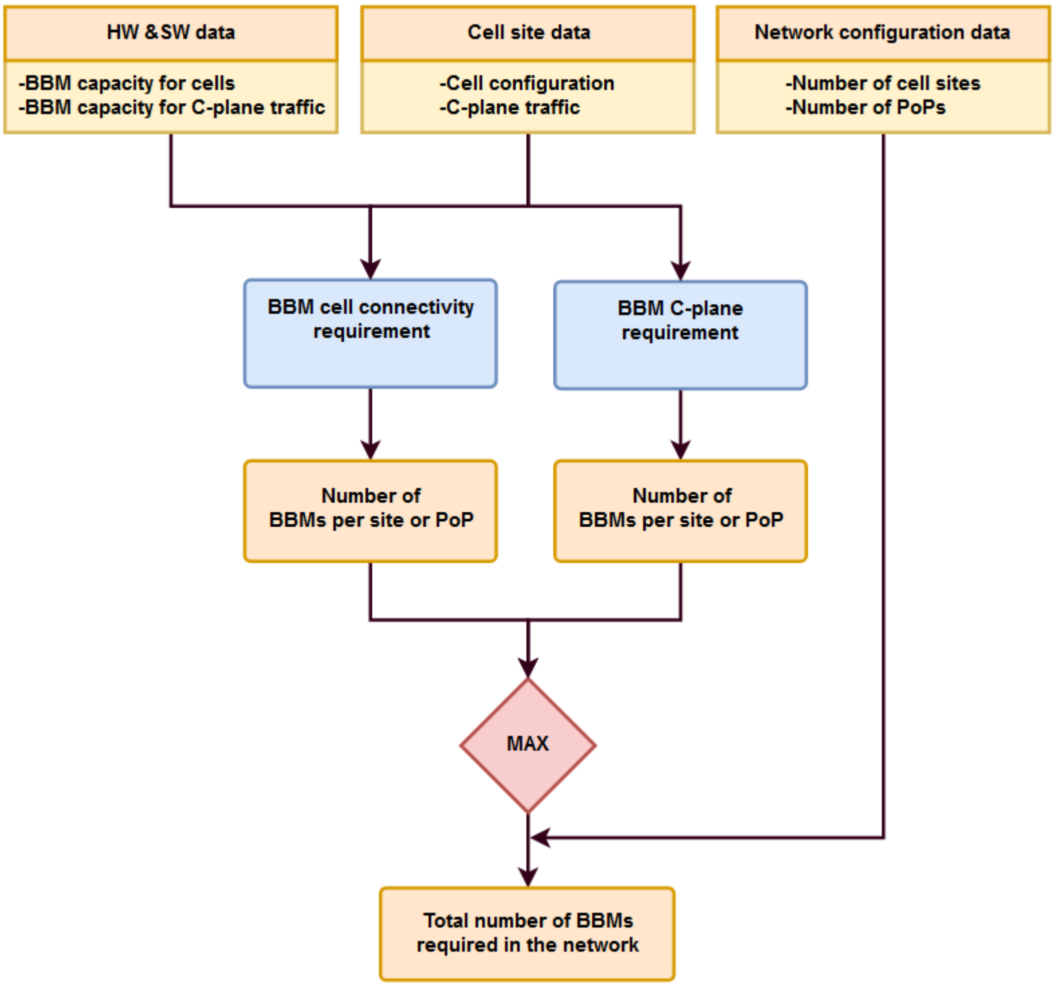

Figure 18: Baseband system model dimensioning flow chart.

Table 6: Cell connectivity requirement per cell site or PoP

BBM unit	Cell connectivity requirement
Capacity unit	Cells have different processing requirement depending on the used bandwidth and MIMO. Higher bandwidth and MIMO require more resources from the CAP unit and the number of CAP units per site must satisfy the total processing requirement by the cells.
Common unit	CAP units are mainly responsible for cell connectivity requirement, but at least one COM unit is required per cabinet.
Cabinet unit	CBN unit is responsible for hosting the other two units. The number of required CBN units is limited by the maximum capacity for hosting the CAP and COM units.

Table 7: C-plane traffic requirement per cell site or PoP

BBM unit	C-plane traffic requirement
Capacity unit	CUE per cell and C-plane load per CUE determine the C-plane traffic on sites. CAP units have certain capacity for handling the C-plane traffic and the number of CAP units must satisfy the total C-plane traffic processing requirement.
Common unit	COM units have two configurations, which are the single and double deployment. The configuration is chosen on depending on the C-plane traffic.
Cabinet unit	The number of required CBN cabinets is dependent on how many cells can be supported per cabinet (determined by how many cells one CAP unit can support) and how many COM units are required to meet the C-plane traffic demand.

Cell connectivity and C-plane traffic requirements are calculated for all three BBM units (CAP, COM, CBN). Depending on which BBM unit is being dimensioned, each requirement outputs a number of either CAP, COM or CBN BBM units required per cell site or PoP. After that, the maximum number of units chosen out of the cell connectivity and C-plane traffic requirements, so that both conditions are met. Total number of BBM units required in the network is then the number of BBM units required per cells site or PoP multiplied by the total number of cell sites or PoPs in the network.

Output of the BBM dimensioning is bill of material for each year consisting of all the required BBM hardware and software licenses in bare-metal and Cloud RAN architectures.

4.2.2 Cloud server dimensioning

In Cloud RAN architectures, part of the baseband processing is virtualized to VNFs (virtualized BBU pools) operating on top of COTS cloud servers. Purpose of the cloud server dimensioning is to calculate the number of cloud servers required by the VNFs for the virtualized baseband processing.

Cloud RAN VNFs can consists of different VMs that handle the different parts of the virtualized baseband processing. In this model it is assumed that the following VMs are deployed to handle the NRT baseband function processing in the cloud. The deployed VMs are User VM (UE VM), Cell VM, Central eNB VM (CVM) and Operation and Management VM (OAM VM). Functions of these VMs are described in table 8. VNF requires at least one of these VMs to be operational, and to guarantee high availability redundant CVMs and OAM VMs might be deployed as back-ups on the VNF in case of failure. VMs require cloud server's processor cores as resources for their processes, which is why it is required to calculate the total number of VMs needed in the network to solve the number of required cloud servers for Cloud RAN.

Table 8: Cloud RAN base station VM description

VM	Function
Cell VM	Cell VM is responsible for running the cell related functions in the VNF.
UE VM	User VM is responsible for running the user related functions in the VNF.
CVM	Central eNB VM runs the centralized functions and the eNB in the VNF.
OAM VM	Operations, Administration and Management VM handles the OAM applications of the VNF.

Cloud server dimensioning follows the flow chart in figure 19. Dimensioning begins by calculating the required network data from cell site and network configuration data. HW & SW data defines the VM and VNF capacity limits and VM processor core requirements as well as the number of available server processor cores.

First part of the cloud server dimensioning is calculating the required VMs in the network, which is done in the Cell VM and UE VM dimensioning. The number of required OAM VMs and Central VMs is assumed to be equal to the number of VNFs so they do not require any dimensioning processes.

After the total number of VMs in the network is known, the number of VNFs can be calculated in the VNF dimensioning. Finally, the total number of cloud servers required in the network can be calculated by using the outputs from Cell VM, UE VM and VNF dimensioning. Cell VM, UE VM, VNF and server dimensioning are explained in more detail in the sections 4.2.3, 4.2.4, 4.2.5 and 4.2.6.

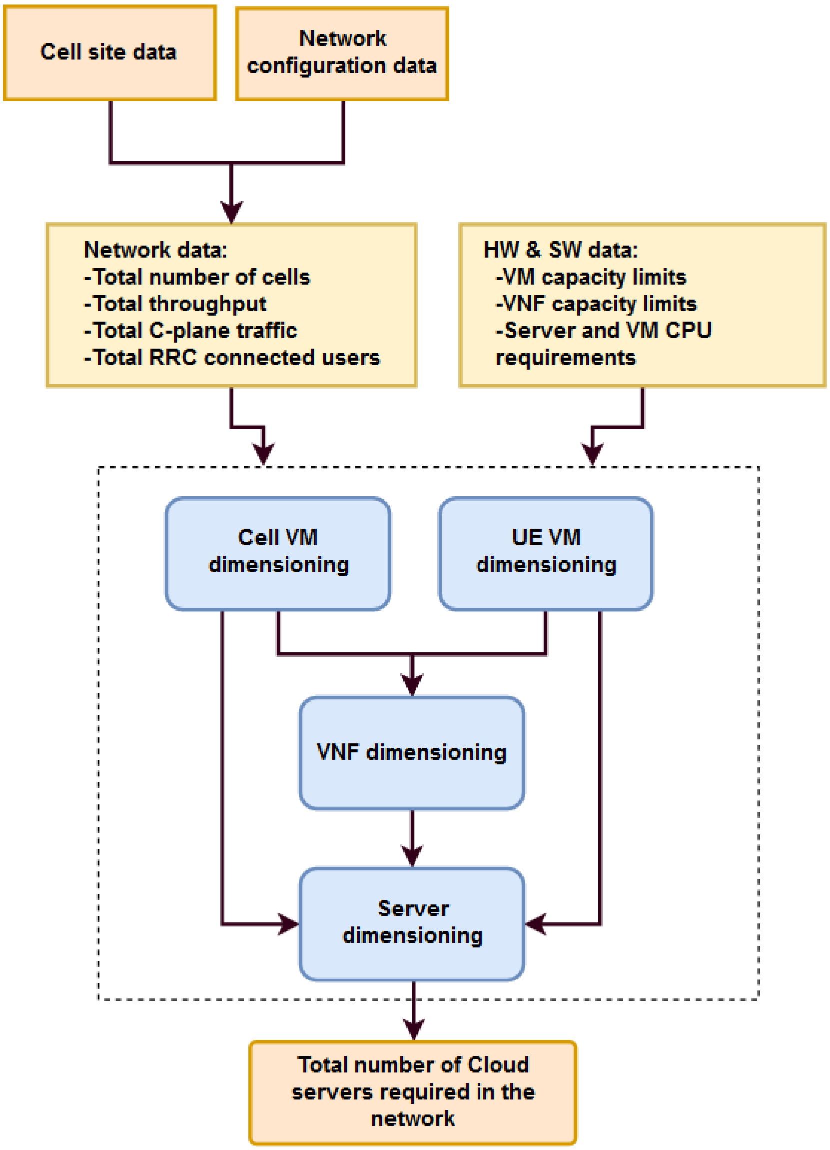

Figure 19: Cloud server dimensioning flow chart.

4.2.3 Cell VM dimensioning

Cell VM dimensioning follows the flow chart in figure 20. Cell VMs are responsible for cell related functions, therefore the Cell VM dimensioning is done based on the total number of cells, number of simultaneous RRC connected users and C-plane traffic in the network. HW & SW data defines the Cell VM's capacity limits for each of these parameters, which then form the Cell VM dimensioning requirements (table 9).

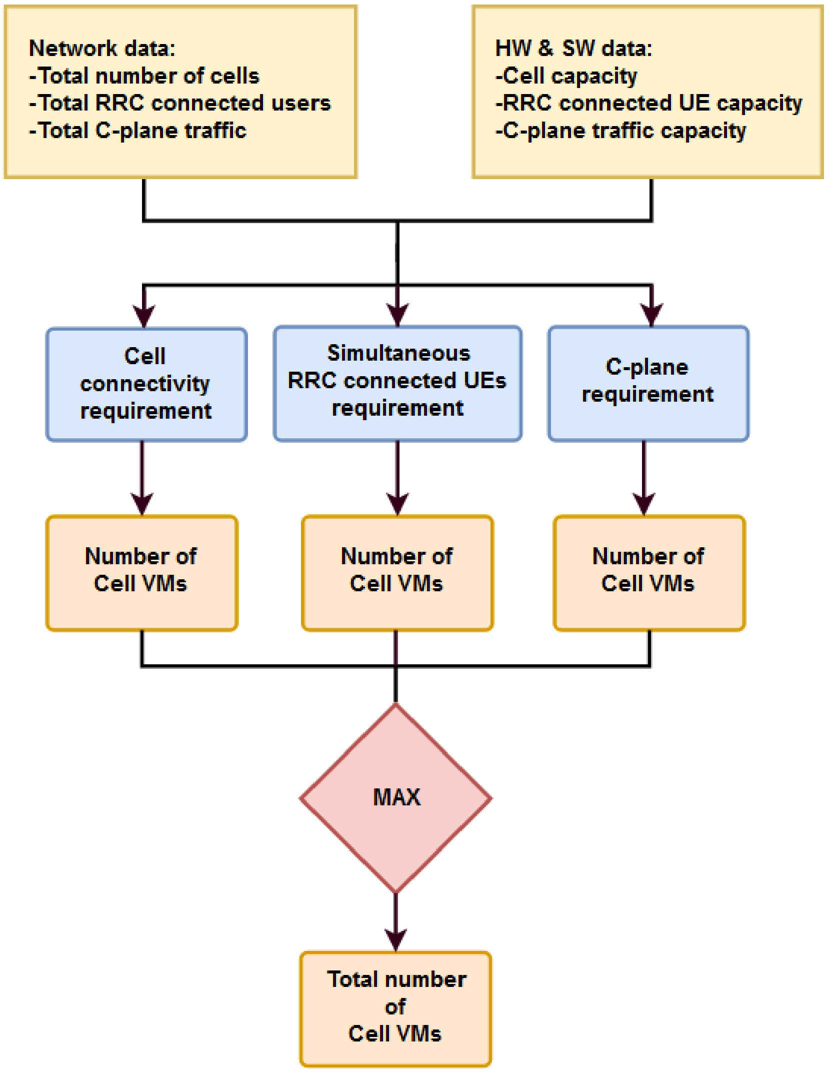

Figure 20: Cell VM dimensioning flow chart.

Each requirement in the Cell VM dimensioning outputs a number of Cell VMs required for full-filling that requirement. The maximum number of Cell VMs is chosen out of these requirements, thus the total number of Cell VMs in the network supports all the dimensioning requirements. The output of the dimensioning is the total number of Cell VMs required in the network.

Table 9: Cell VM dimensioning requirements

Requirement	
Cell connectivity	Number of Cell VMs needed to support the total number of cells in the network.
Simultaneous RRC connected UEs	Number of Cell VMs needed to support the total number of RRC connected UEs in the network.
C-plane traffic	Number of Cell VMs required to support total C-plane traffic in the network.

4.2.4 UE VM dimensioning

UE VM dimensioning follows the flow chart in figure 21. UE VMs are responsible for user related functions, therefore the UE VM dimensioning is done based on the total uplink and downlink throughput traffic, simultaneous RRC connected users and C-plane traffic in the network. HW & SW data defines the UE VM's capacity limits for each of these parameters, which then form the UE VM dimensioning requirements (table 10).

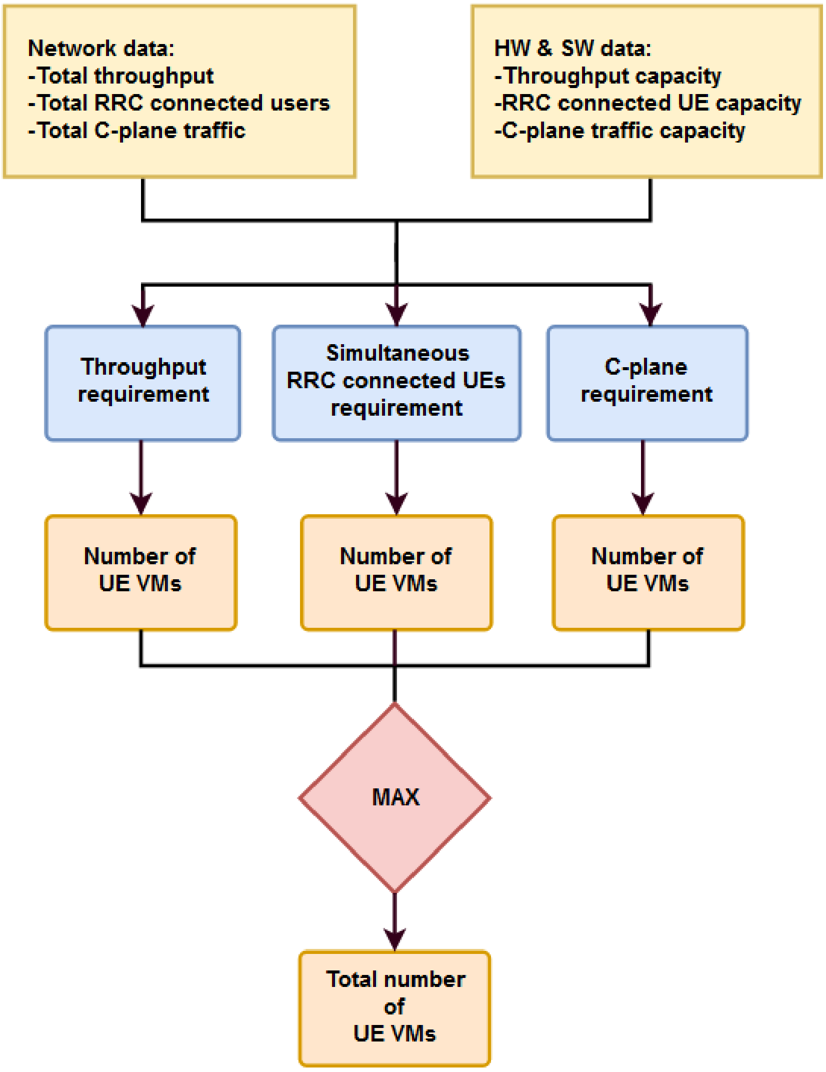

Figure 21: UE VM dimensioning flow chart.

Each UE VM dimensioning requirement outputs a number of UE VMs that are required for full-filling the requirement. The maximum number of UE VMs is chosen out of these requirements, thus the total number of UE VMs in the network supports all the dimensioning requirements. The final output of the UE VM dimensioning is the total number of UE VMs required in the network.

Table 10: UE VM dimensioning requirements

Requirement	
Throughput	Number of UE VMs needed to support the total downlink and uplink throughput from cell sites in the network.
RRC connected UEs	Number of UE VMs needed to support the total number of RRC connected UEs in the network.
C-plane traffic	Number of UE VMs required to support total C-plane traffic in the network.

4.2.5 VNF dimensioning

VNF dimensioning follows the flow chart in figure 22. VNFs are entities consisting of different VMs, but they also have a limitations for the cell connectivity and C-plane traffic. Therefore, the number of VNFs required in the network depends on total number of cells, C-plane traffic, Cell VMs and UE VMs in the network. HW & SW data defines the VNF's capacity for each of these parameters, which then form the VNF dimensioning requirements (table 8).

Table 11: VNF dimensioning requirements

Requirement	
Cell connectivity	Number of VNFs needed to support the total number of cells in the network.
C-plane traffic	Number of VNFs required to support the total C-plane traffic in the network.
Cell VM support	Number of VNFs required to support the total number of needed Cell VMs in the network.
UE VM support	Number of VNFs needed to support the total number of needed UE VMs in the network.

Each VNF dimensioning requirement outputs a number of VNFs required for full-filling that requirement. The maximum number of VNFs is then chosen out of these requirements, thus supporting all dimensioning requirements for the network.

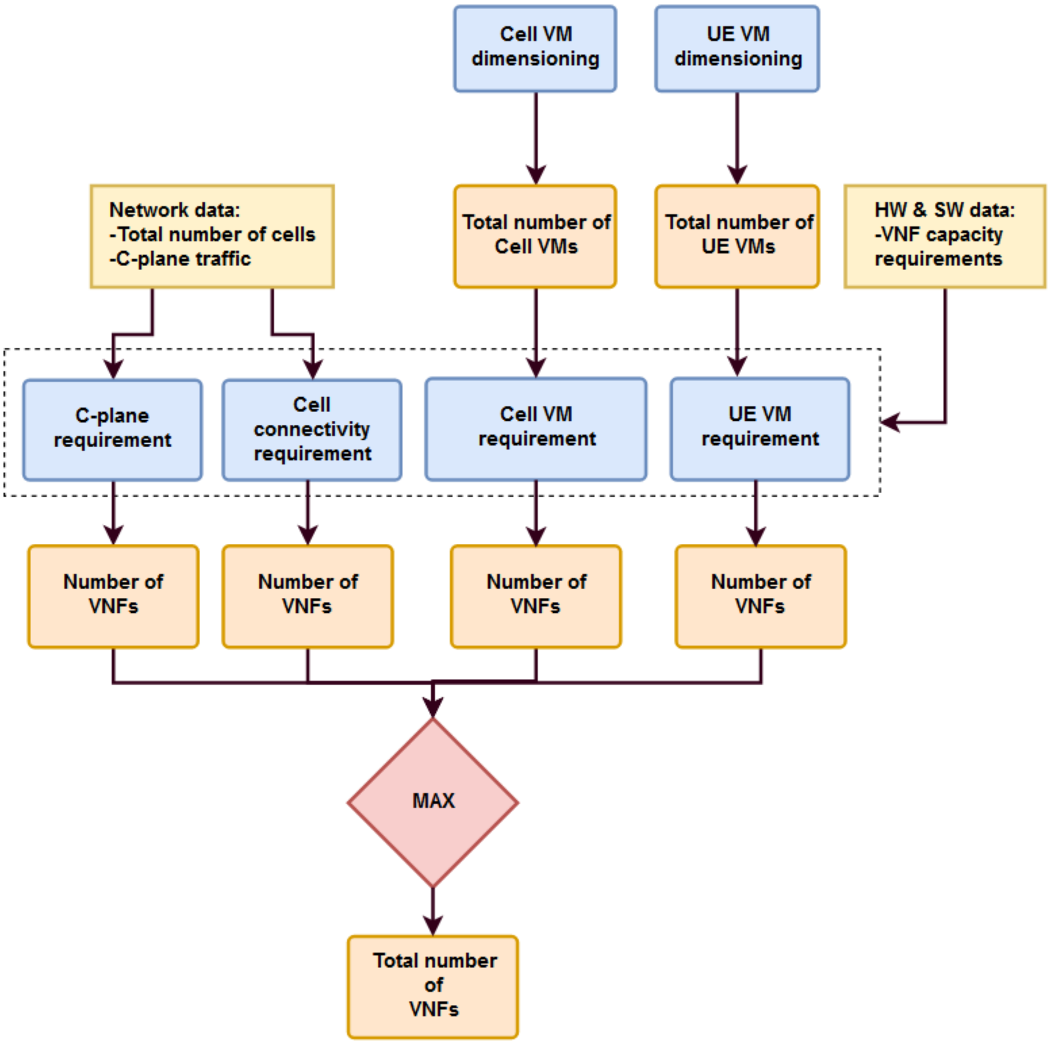

Figure 22: VNF dimensioning flow chart.

4.2.6 Server dimensioning

Server dimensioning is the final part in Cloud server dimensioning and it solves the total number of cloud servers required in the network. First part of the server dimensioning is allocating the required Cell VMs and UE VMs to VNFs to approximate the VM configuration per VNF.

Number of Cell VMs per VNF (VNF_{CellVM}) can be estimated by:

$$VNF_{CellVM} = \frac{N_{CellVM}}{VNF_{total}} \tag{1}$$

Where N_{CellVM} is total number of Cell VMs in the network and VNF_{total} is the total number of VNFs required in the network.

Number of UE VMs per VNF (VNF_{UEVM}) can be estimated by:

$$VNF_{UEVM} = \frac{N_{UEVM}}{VNF_{total}} \quad (2)$$

Where N_{UEVM} is total number of UE VMs in the network and VNF_{total} is the total number of VNFs required in the network.

After the Cell and UE VM configuration per VNF is solved, the number of cloud servers required per VNF (S_{VNF}) can be estimated with:

$$S_{VNF} = \frac{VNF_{CellVM} \times C_{CellVM} + VNF_{UEVM} \times C_{UEVM} + C_{CVM} + C_{OAMVM}}{C_{server}} \quad (3)$$

Where VNF_{CellVM} and VNF_{UEVM} are the total number of VMs per VNF. C_{CellVM}, C_{UEVM}, C_{CVM} and C_{OAMVM} are the number of CPUs required by the different VMs and C_{server} is the number of available CPUs per server.

Finally the total number of cloud servers ($N_{servers}$) required in the network is then calculated with:

$$N_{servers} = S_{VNF} \times VNF_{total} \quad (4)$$

Where S_{VNF} is number of cloud servers per VNF and VNF_{total} is the total number of VNFs in the network.

After the number of cloud servers required in the network is solved, all the other required hardware elements such as switches, controllers and server racks can also be calculated into the bill of material.

4.3 Cost calculations

Cost calculations part of the Cloud RAN TCO model calculates the CAPEX, OPEX and TCO for all four architectures follwing the flow chart in figure 23. The calculations are based on each year's bill of material, network configuration data and financial data. The CAPEX and OPEX calculations consist of the the costs required for setting up and operating RAN. The final output of the Cloud RAN cost modeling is TCOs for the evaluated architectures for the given subscriber traffic and network configuration scenario during the specified time period of 1-5 years.

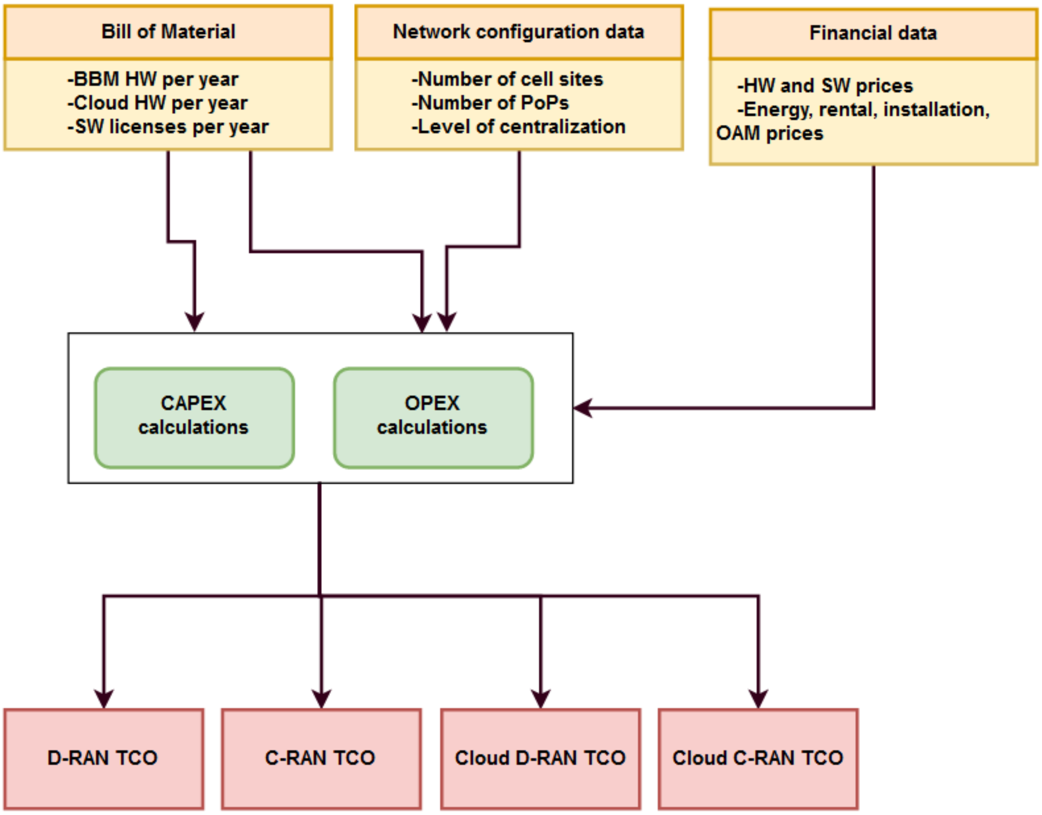

Figure 23: Cost modeling flow chart.

4.3.1 CAPEX calculations

CAPEX in this model includes investments into BBM hardware, cloud server hardware and software licenses. Calculating CAPEX is done by using the bill of material, which contains all the required hardware and software items for each year and the financial data, which contains the pricing information for the required items. Calculations follow a simple formula of multiplying each year's required hardware items and software licenses listed in the bill of material for the chosen architecture with the corresponding prices from financial data. CAPEX is calculated on yearly basis and the total network CAPEX is then calculated by summing each year's costs together.

Total CAPEX

Total cumulated CAPEX for the network architecture is therefore:

$$CAPEX = \sum_{i=0}^{i=5}(C_{BBM} + C_{BBSW} + C_{CHW} + C_{CSW} + C_{RFSW}) \quad (5)$$

Where C_{BBM} is the cost of BBM hardware , C_{BBSW} is the cost of baseband software, C_{CHW} is the cost of cloud hardware, C_{CSW} is the cost of cloud software and C_{RFSW} is the cost of RF software in year i.

4.3.2 OPEX calculations

OPEX includes energy cost, rental costs, baseband capacity upgrade costs, Operation and Maintenance (OAM) costs and hardware and software maintenance fees. Calculating the OPEX for the network is done by using the bill of material from network dimensioning, network configuration data and financial data. OPEX is calculated for each of the architectures on yearly basis starting from year 1 and the total cumulated OPEX for the network is retrieved by summing each year's costs together.

Energy cost

Network's yearly energy cost is calculated by summing the total energy consumption of BBM units and cloud servers. BBMs and cloud servers do not require any external cooling equipment, therefore only their energy consumption is taken into account in the calculations.

Yearly power consumed by the baseband processing hardware (P_{tot}) can be calculated with:

$$P_{tot} = \sum_{i}(P_i \times n_i) + \sum_{j}(P_j \times n_j) \quad (6)$$

Where **i** represents the total number of CAP, COM, and CBN units in the network, **j** represents the total number of power consuming elements in cloud servers (compute nodes, switches) and **P** is the power consumption of the BBM unit i or cloud server hardware element j.

Yearly energy consumption (E_{tot}) in kilowatt hours is therefore:

$$E_{tot} = \frac{P_{tot}}{1000} \times 24h \times 365d \quad (7)$$

Where P_{tot} is yearly power consumed by the baseband processing hardware.

Yearly energy cost (C_{energy}) is then calculated with:

$$C_{energy} = E_{tot} \times C_{kWh} \quad (8)$$

Where E_{tot} is yearly energy consumed by the baseband processing hardware and C_{kWh} cost of energy per kWh.

Site rental cost

Site rental costs consist of yearly cell tower rental cost for the radio hardware and antennas, cell site floor rental cost per BBM cabinet in distributed architectures, PoP floor rental cost per BBM cabinet in centralized architectures and data center cabinet rental cost per server rack in Cloud RAN architectures.

Yearly site rental cost for the network can be calculated with:

$$C_{rent} = R_{tower} \times n_{RAP} + R_{BBsite} \times n_{Dcab} + R_{BBPoP} \times n_{Ccab} + R_{DC} \times n_{rack} \quad (9)$$

Where **R** represents the rental costs for the cell tower, cell site floor space rental cost per cabinet, PoP floor space rental cost per cabinet and data center rental cost per server rack. And **n** represents the number of cell sites/radio towers, BBM cabinets in D-RAN configuration, BBM cabinets in C-RAN configuration and the number of server racks in the network.

Baseband capacity upgrade cost

Baseband capacity upgrade costs consists of baseband processing hardware installation costs at cell sites and PoPs. New equipment is required to be installed each year, when the network traffic is expected to exceed the current capacity of the baseband processing hardware. Installation costs are assumed to be less at C-RAN PoPs due to centralization. Installing new hardware at fewer locations decreases the cost of traveling to different locations and reduces the hardware installation time.

Total baseband capacity upgrade cost for the network per year is:

$$C_{BBup} = U_{D-up} \times n_{D-BBM} + U_{C-up} \times n_{C-BBM} + U_{DC-up} \times n_{server} \quad (10)$$

Where **U** represents the baseband capacity upgrade cost for D-RAN sites, C-RAN PoPs and data centers. And **n** represents the number of new BBM units required to be installed to D-RAN sites, PoPs and new cloud servers required to be installed at data centers.

Operation and Maintenance cost

Operation and Maintenance (OAM) costs consists of network trouble management, preventive maintenance, performance optimization, capacity analysis and planning and software release upgrades. OAM costs are very unique and differ depending on the MNO's operations and maintenance processes. The OAM costs used in the model are cell site specific estimations for D-RAN, C-RAN, Cloud D-RAN and Cloud C-RAN provided by Nokia Bell Labs study.

Hardware and software maintenance fee

Hardware and software maintenance fees are calculated as fixed percentages of the total cumulated CAPEX spent on hardware and software each year.

Total OPEX

Total cumulated OPEX for the network architecture is therefore:

$$OPEX = \sum_{i=i}^{i=5} (C_{energy} + C_{rent} + C_{BBup} + C_{OAM} + C_{maint}) \tag{11}$$

Where C_{energy} is the energy cost, C_{rent} is the rental cost, C_{BBup} is the baseband capacity upgrade cost, C_{OAM} is the OAM cost and C_{maint} is the hardware and software maintenance cost in year i.

4.3.3 Total Cost of Ownership

Finally, the Cloud RAN TCO model combines the results from CAPEX and OPEX calculations for each of the architectures to solve and compare the TCO figures. The results given by the Cloud RAN TCO model can be used to understand in which scenarios Cloud RAN could be more economical and what are the cost effects to MNOs using the Cloud RAN and companies offering the solution.

5 Analysis

This chapter features an example cost modeling case for Operator A using the Cloud RAN TCO modeling tool to analyze costs of different architectures. The example case is based on data given by a mobile network Operator A. The modeling is done for 5-year period, where year 0 represents the initial state of the network in the model. Operator A has an existing Radio Access Network using the D-RAN architecture in year 0, which is then either kept as D-RAN or transformed to C-RAN, Cloud D-RAN or Cloud C-RAN. Each option is evaluated in terms of TCO in the 5-year period to solve what is the most economical RAN architecture.

This chapter will first introduce the assumed inputs from the Operator A's network deployment, that were used for the modeling and feature the cost assumptions for the case. The C-plane traffic growth forecast plays a significant role in the success of the Cloud RAN solutions. Therefore, two different scenarios are created based on the expectations of the C-plane traffic growth, which are the low and high C-plane traffic growth scenarios. The four architectures are then compared and analyzed in terms of costs, required hardware and energy consumption in both scenarios.

5.1 Inputs and Assumptions

5.1.1 Network configuration

Operator A's network configuration is shown in the figure 24. Total number of cell sites consisting of radio functionalities is the same in all RAN architectures and the number is expected to increase from 2762 to 6158. Both C-RAN and Cloud C-RAN architectures are gradually centralized from the year 0's D-RAN architecture. Baseband processing is being transitioned from cell sites to C-RAN PoPs and data centers and the number of C-RAN PoPs and data centers will be increased from 0 to 10. Current Cloud RAN solutions were also assumed to be able support 96 cells per VNF.

D-RAN/RAPs	Year 0	Year 1	Year 2	Year 3	Year 4	Year 5
Total number of cell sites (D-RAN sites or RAPs)	2762	3227	3901	4567	5283	6158
C-RAN and Cloud C-RAN						
Share of D-RAN sites in C-RAN	100 %	80 %	60 %	50 %	40 %	20 %
Number of D-RAN cell sites in C-RAN	2762	2582	2341	2284	2114	1232
Number of C-RAN cell sites in C-RAN	0	646	1561	2284	3170	4927
Number of C-RAN PoPs	0	2	5	7	10	10
C-RAN sites served per PoP	0	323	313	327	317	493
Cloud RAN						
Number of regional PoPs/DCs in Cloud RAN	0	1	1	1	1	1
Number of Cloud C-RAN PoPs	0	1	4	6	9	9
Share of sites served by DC	0 %	50 %	20 %	15 %	10 %	10 %
Share of sites served by PoPs	0 %	50 %	80 %	75 %	90 %	90 %
Total number of C-RAN sites served by DCs	0	323	313	343	317	493
Total number of C-RAN sites served by PoPs	0	323	1249	1713	2853	4435
# of Cloud C-RAN sites served per DC	0	323	313	343	317	493
# of Cloud C-RAN sites served per PoP	0	323	313	286	317	493

Figure 24: Network configuration for operator A.

5.1.2 Cell site configuration

Cell configuration affects the cell connectivity requirements in the network dimensioning. Adding more cells to cell sites, upgrading to higher bandwidth and MIMO will require more resources from the baseband processing hardware. Operator A's cell site cell configuration consists of FDD type 5/10 MHz 2x2 MIMO, 5/10 MHz 4x4 and 15/20 MHz 4x4 MIMO cells. The cell configuration over the 5-year period is shown in the figure 25, where in the year 0 cell sites consists of nine 5/10 MHz 2x2 cells and during the years more cells are added and some are replaced by cells with higher bandwidth and MIMO.

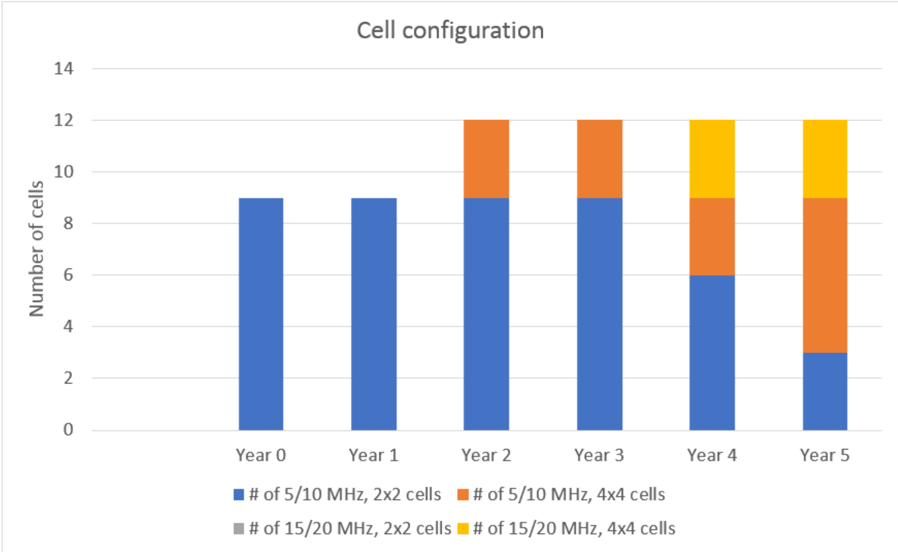

Figure 25: Operator A's cell configuration on cell sites.

Busy hour throughput and cell site data volume affect the cloud server dimensioning and the required software licenses. Cell site busy hour throughput and daily data volume is expected to follow the figure 26 for the five-year period. Busy hour share of the traffic per day is expected to be 8 %, uplink share of the total throughput is expected to grow from 10 % to 16 % and average busy hour throughput per site is expected to grow from 9,20 Mbps to 14,80 Mbps.

		Year 0	Year 1	Year 2	Year 3	Year 4	Year 5
BH share of traffic per day per site	%	8,00 %	8,00 %	8,00 %	8,00 %	8,00 %	8,00 %
UL share of total throughput (DL+UL)	%	10 %	10 %	11 %	12 %	14 %	16 %
Avg. BH throughput per site	Mbps	9,20	10,10	11,10	12,20	13,40	14,80
Data volume per site	GB/day	50,54	55,48	60,97	67,02	73,61	81,30

Figure 26: Cell site throughput and data volume for operator A.

C-plane traffic affects the C-plane requirements in the network dimensioning for both proprietary baseband hardware (BBMs) and COTS cloud hardware (cloud servers). C-plane traffic growth is measured with two main parameters, which are the maximum number of Connected User Equipment (CUE) per cell site and C-plane traffic load per CUE. Maximum CUE per cell site measures the maximum number of connected devices per cell site, which includes mobile phones, IoT sensors, connected cars, etc. Maximum CUE per cell site is expected to grow according to figure 27. C-plane traffic load per CUE measures the frequency of signaling messages sent by these devices. C-plane traffic load per CUE is assumed to be 0,33 messages per second (1 message in 3 seconds) and not to increase in the 5-year period.

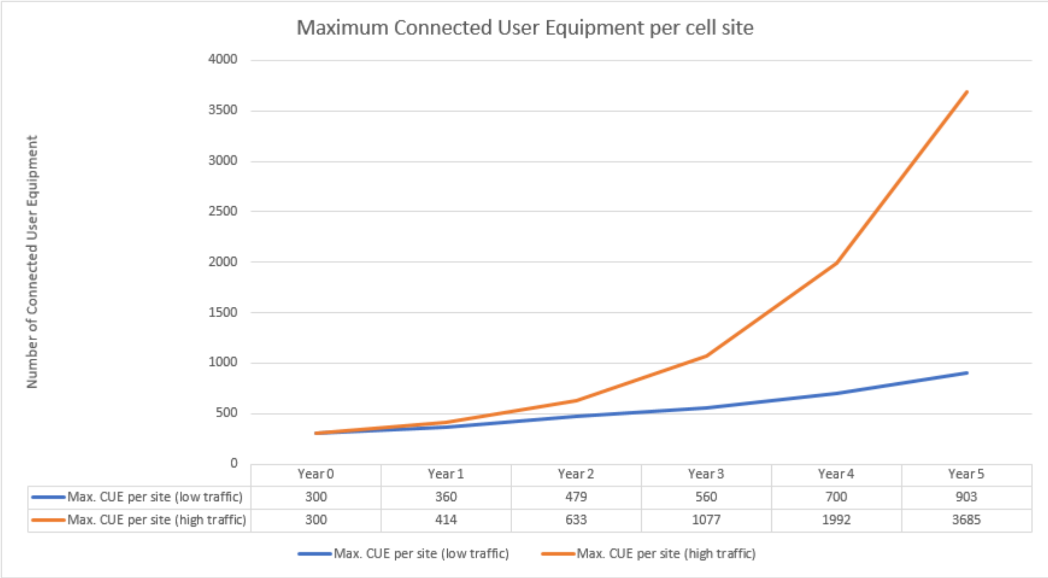

Figure 27: Operator A's expected CUE per cell site traffic growth in the low and high C-plane traffic growth scenarios.

5.1.3 Cost assumptions

The measurement scale used to compare the costs in TCO, CAPEX and OPEX is using example units to show the differences between the costs and does not represent any real-life currency. The following assumptions were made on the CAPEX and OPEX costs.

- Pricing of RF and baseband software is assumed to be approximately the same or more in Cloud RAN due to lower amount of baseband hardware needed in the network and moving towards cloud based baseband processing.
- Rental costs for cell tower, cell site floor space for the baseband hardware, PoP floor space rental and data center cabinet rental were estimates provided in the example case. Operator A has its own data centers so data center cabinet rental cost was set to zero.
- Capacity installation costs for installing new baseband capacity at D-RAN sites is assumed to more expensive due to upgrades have to be made each site separately, which will take more time and require more work.
- OAM costs for network trouble management, preventive maintenance, RF and performance optimization, capacity analysis and planning and software release upgrades are taken from Nokia Bell Labs study case to estimate the costs per cell site in each architecture.

5.2 Scenario Analysis

Based on the C-plane traffic growth expectations, two different traffic growth scenarios are evaluated for Operator A. In these two scenarios maximum CUE per cell site determines the C-plane traffic growth, as the C-plane traffic load per CUE is assumed to stay the same. This is due to the expectation that number of devices will grow, but the signaling frequency of these devices is will stay the same in the Operator A's network.

Scenario 1 consists of the low C-plane traffic growth scenario (figure 27), where the maximum number of CUE per cell site is assumed to increase from 300 to 998 with 25 % Compound Annual Growth Rate (CAGR) between the years 0 and 5. Expectation for scenario 1 is that Cloud RAN architectures will not be able to leverage the advantage of better capacity scaling with low C-plane traffic growth to achieve better TCO than bare-metal architectures.

Scenario 2 consists of the high C-plane traffic growth scenario (figure 27), where the maximum number of CUE per cell site is assumed to increase from 300 to 3685 with 65 % CAGR between the years 0 and 5. Maximum number of CUE per site growth is assumed to be slower in the first years and accelerate in the later years. Expectation for scenario 2 is that Cloud RAN architectures are expected to offer much better scaling for the traffic demand than the bare-metal architectures, thus resulting into lower TCO.

5.2.1 Scenario 1 analysis

In the low C-plane traffic growth scenario, the C-RAN architecture turned out to be the most economical architecture in terms of TCO (figure 28). Compared to D-RAN, C-RAN achieved 13,6 % lower TCO, Cloud D-RAN 3,4 % higher TCO and Cloud C-RAN 8,4 % lower TCO. Both, Cloud D-RAN and Cloud C-RAN architectures failed to achieve better TCO than their corresponding bare-metal architectures and in-fact Cloud D-RAN turned out the be the most expensive architecture.

The cost positions of the architectures were quite stable during the 5-year time period (figure 29). During year 4, Cloud C-RAN became more economical than D-RAN. C-RAN stayed as the most economical architecture and Cloud D-RAN as the most expensive during the evaluated time period.

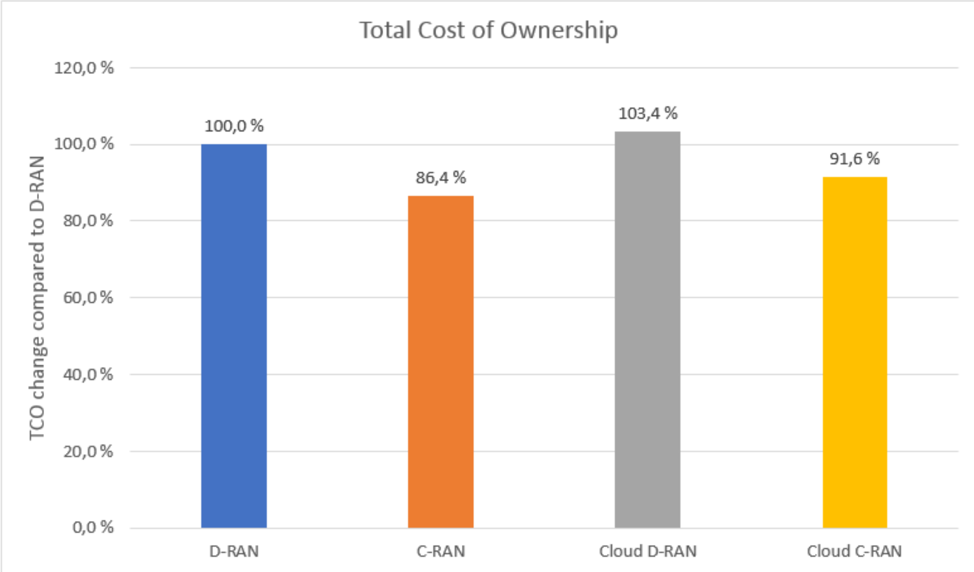

Figure 28: Scenario 1 cumulative TCO in year 5 for the architectures.

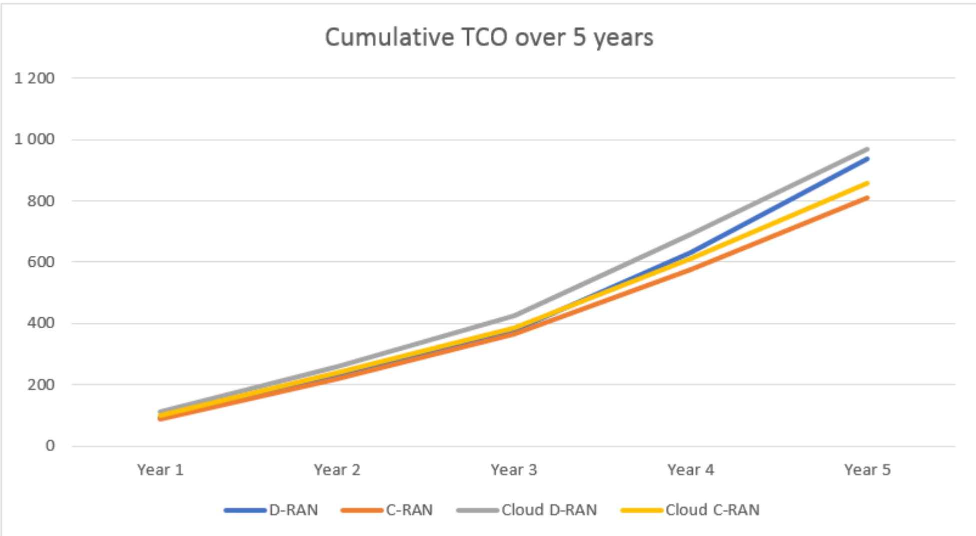

Figure 29: Evolution of cumulative TCO for the architectures during the 5-year period in scenario 1.

Baseband hardware and energy consumption

Determining the reasons why Cloud RAN architectures failed to achieve lowest TCO in scenario 1 can be discovered by analyzing the required baseband processing hardware between the architectures (figure 30). Comparing the corresponding bare-metal architectures with the respected Cloud RAN architectures shows that the amount of BBMs is reduced only slightly, but new cloud servers have to be added to the infrastructure.

Cloud C-RAN can utilize the baseband processing hardware much more effectively than D-RAN. On the other hand, when comparing Cloud C-RAN to C-RAN, the difference is not that big with the addition of cloud servers. Even though Cloud RAN can leverage better pooling gains to utilize the hardware more efficiently, with low C-plane traffic growth Cloud RAN is not able to take advantage of cloud hardware's better scaling for the C-plane traffic.

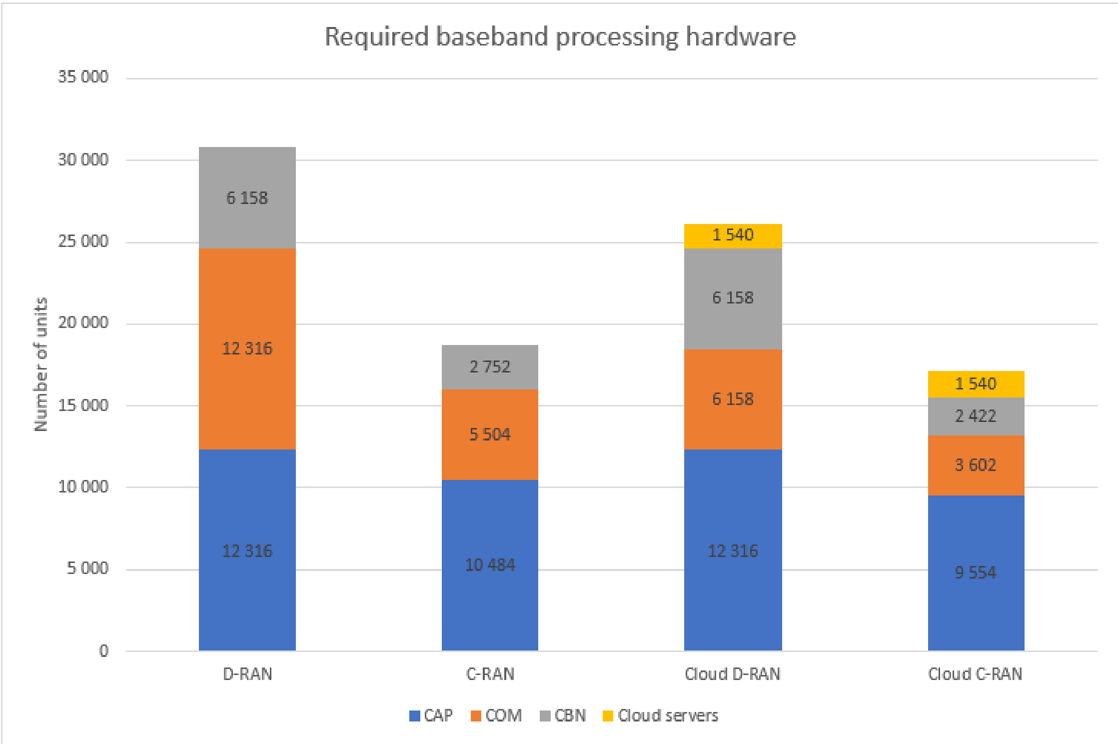

Figure 30: Scenario 1's required baseband hardware by the different architectures in year 5.

Studying the cumulative energy consumption of the baseband processing hardware (figure 31) in the different architectures it can be noticed that C-RAN is also the most ecological architecture in terms of energy consumption. COTS cloud hardware has higher energy consumption than the proprietary hardware and due to only minor difference in the required baseband processing hardware Cloud RAN is not able to have better energy consumption than C-RAN.

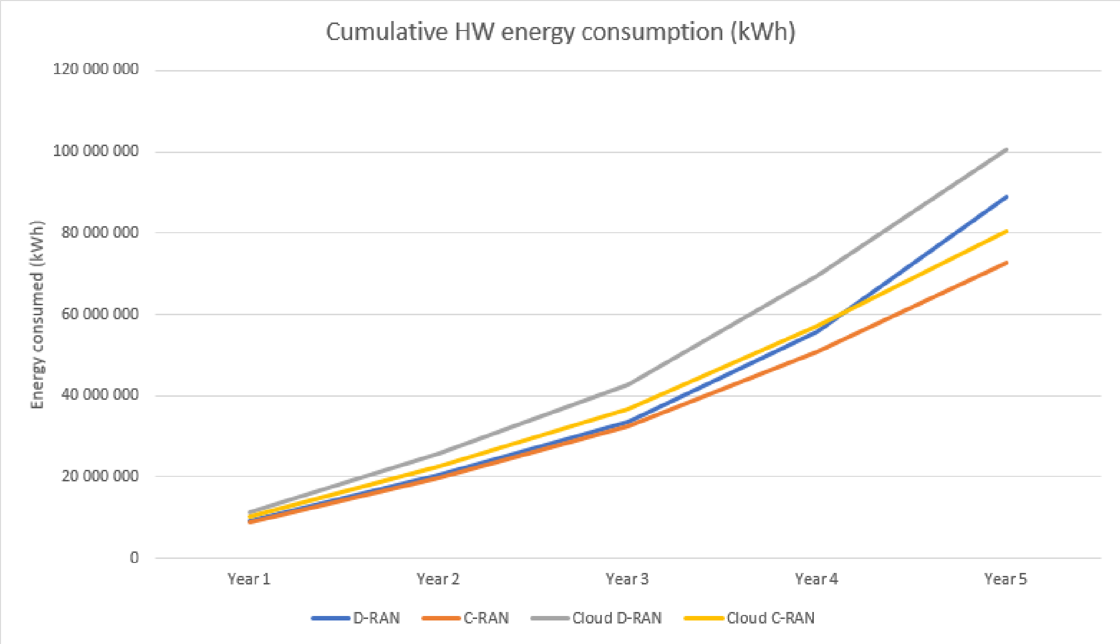

Figure 31: Cumulative energy consumption by the baseband processing hardware in scenario 1.

CAPEX

Analyzing Scenario 1's low C-plane traffic growth in terms of CAPEX (figure 32) shows that C-RAN has the lowest CAPEX with 10,1 % lower costs than D-RAN. C-RAN was able to achieve the lowest CAPEX, because it was able to reduce the number of required BBMs without having to invest into additional cloud servers. Both Cloud RAN architectures turned out to have higher CAPEX than D-RAN, as Cloud D-RAN had 12,3 % and Cloud C-RAN had 2,1 % higher costs. In Cloud RAN architectures, the addition of cloud servers and software was more expensive than the savings in BBM hardware.

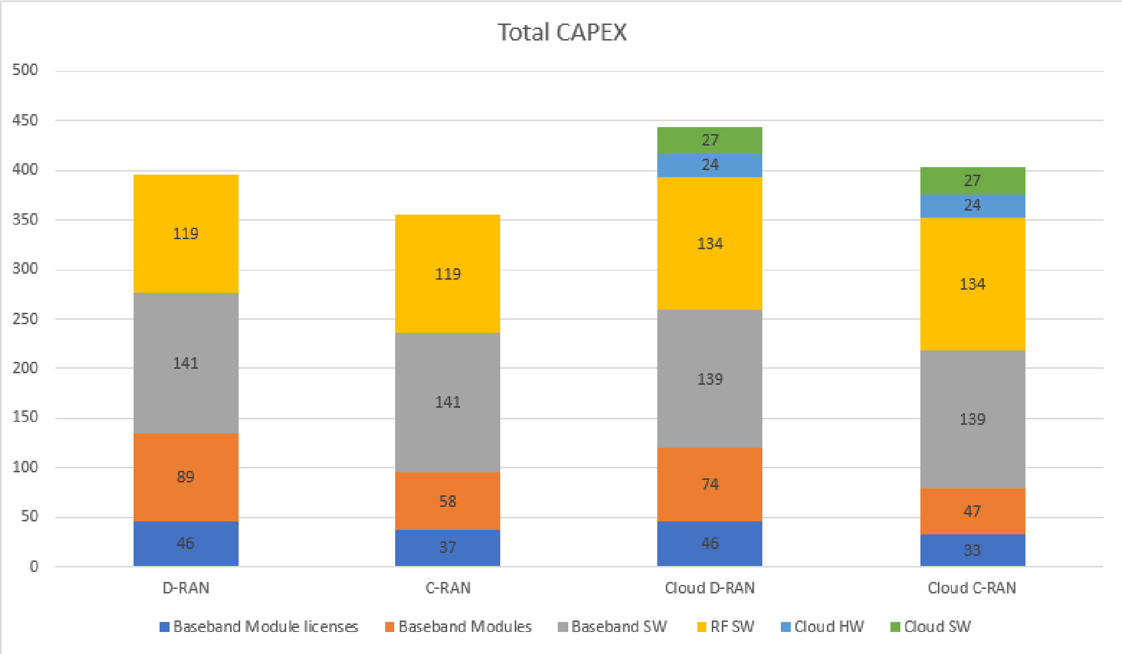

Figure 32: Scenario 1's cumulative CAPEX in the fifth year for the architectures.

OPEX

OPEX analysis of scenario 1 (figure 33) shows that C-RAN has the lowest costs compared to D-RAN with 16,0 % lower OPEX. Cloud RAN architectures were also able to achieve lower costs compared to D-RAN, as Cloud D-RAN had 3,2 % and Cloud C-RAN had 16,1 % lower OPEX. C-RAN achieving the lowest OPEX really shows the benefits of centralizing the baseband processing, as baseband capacity upgrade costs and site rental costs are reduced significantly. Cloud RAN architectures were also able to reduce costs, because of lower baseband capacity upgrade costs. Moving the baseband processing to cloud servers reduces the baseband capacity upgrade costs, as less site visits are required and the better scaling of cloud hardware means baseband capacity upgrades are required less frequently.

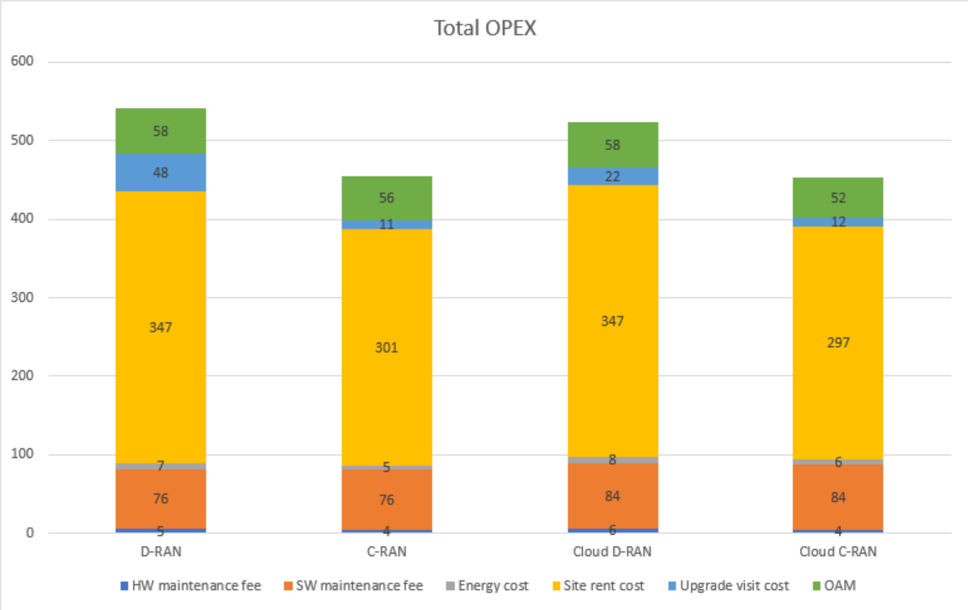

Figure 33: Scenario 1's cumulative OPEX in the fifth year for the architectures.

5.2.2 Scenario 2 analysis

In the high C-plane traffic growth scenario, the Cloud C-RAN turned out the be the most economical architecture in terms of TCO (figure 34). Compared to D-RAN, C-RAN achieved 6,6 % lower TCO, Cloud D-RAN 10,3 % lower TCO and Cloud C-RAN 19,0 % lower TCO. In the high C-plane traffic growth scenario, the Cloud RAN architectures were able to achieve lower TCO than their corresponding bare-metal architectures.

The cost position of the architectures during the 5-year time period can be seen in figure 29. There were no significant differences between the costs until year 4, when the CUE per site growth starts to accelerate, which causes the costs in bare-metal architectures to rise more rapidly than in Cloud RAN architectures. Due to this, Cloud D-RAN is able become more economical than C-RAN and Cloud C-RAN becomes clearly the most economical architecture.

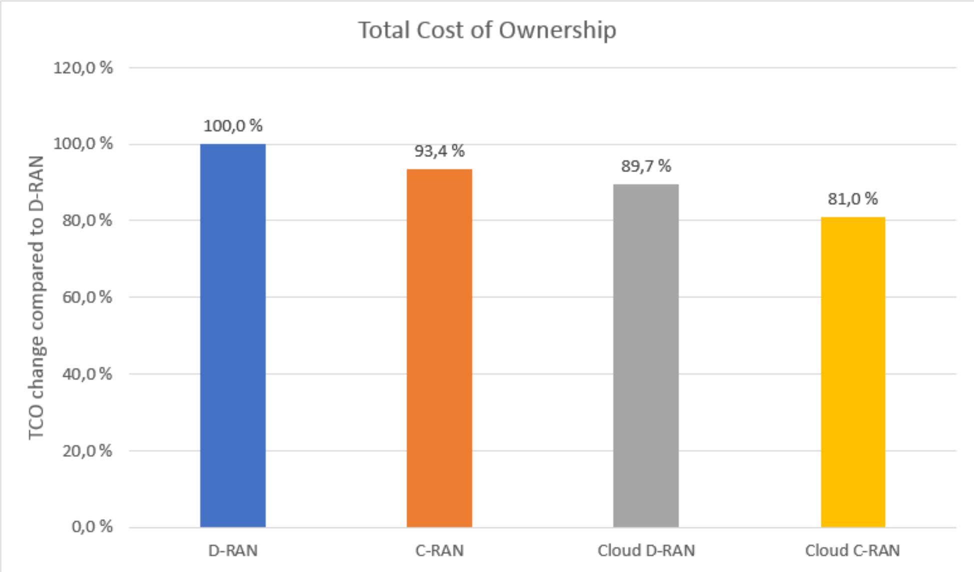

Figure 34: Scenario 2 cumulative TCO in year 5 for the architectures.

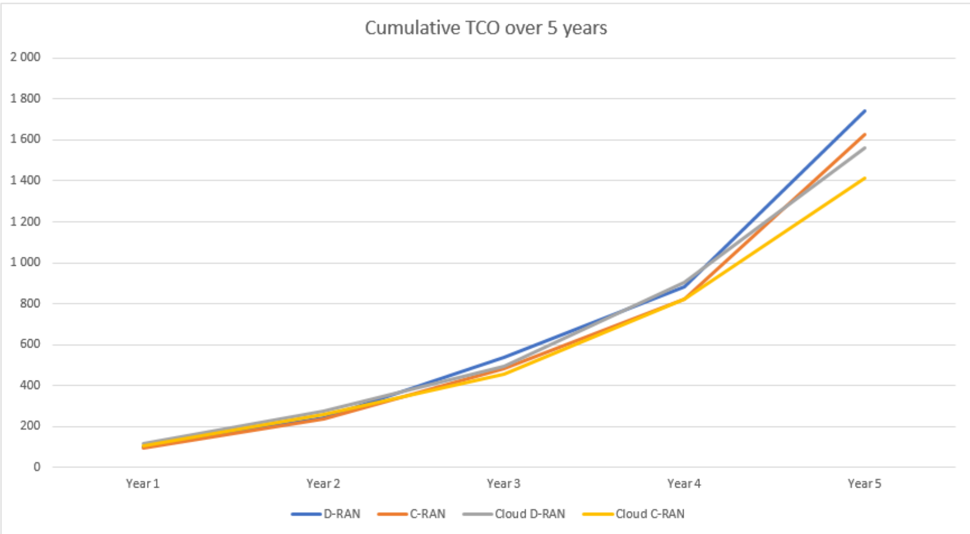

Figure 35: Evolution of cumulative TCO for the architectures during the 5-year period in scenario 2.

Baseband hardware and energy consumption

Explanation for Cloud RAN's success in the scenario 2 can be found by analyzing the required baseband processing hardware between the architectures (figure 36). Comparing the corresponding bare-metal architectures with the respected Cloud RAN architectures shows that scaling of the cloud servers is much better than the proprietary BBM units in the high C-plane traffic growth scenario. When comparing Cloud RAN architectures to D-RAN, Cloud D-RAN can reduce the required BBM units by 50 % and Cloud C-RAN can reduce the amount of required BBM units by 60-80 %. Cloud RAN's ability to leverage pooling gains in the network dimensioning and more rapidly growing C-plane traffic allows Cloud RAN to fully take advantage of better hardware scaling for the C-plane traffic.

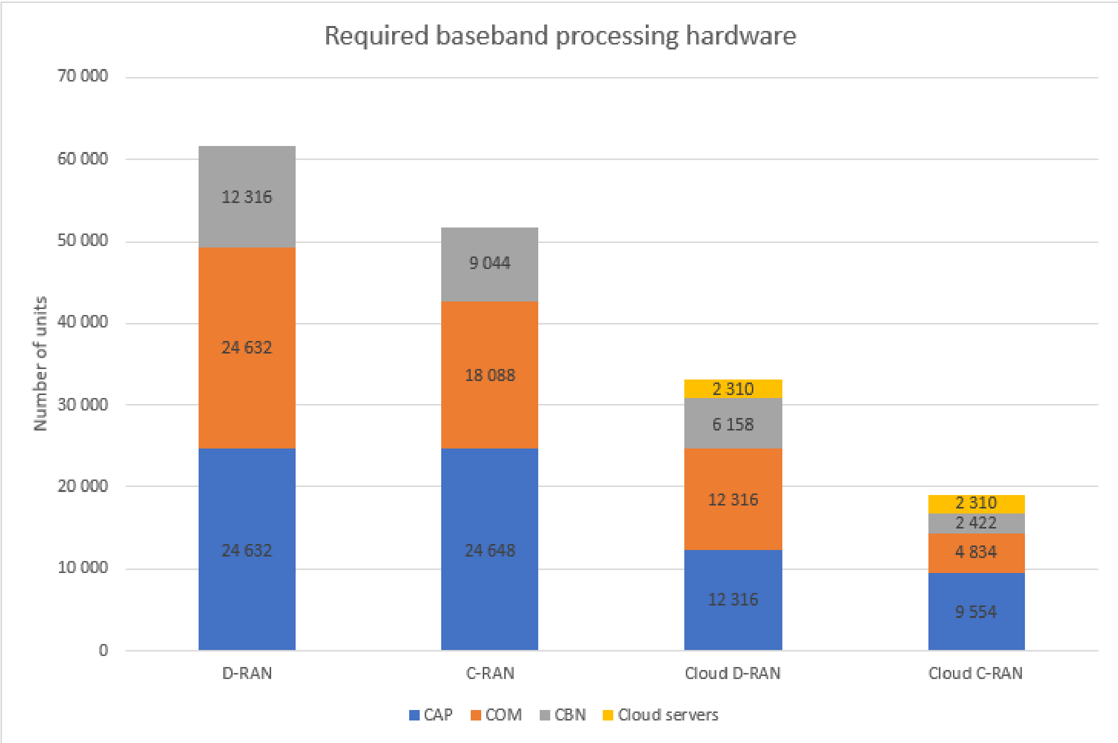

Figure 36: Scenario 2's required baseband hardware by the different architectures in year 5.

Studying the cumulative energy consumption of the baseband processing hardware in the different architectures (figure 37) shows that Cloud C-RAN is the most ecological architecture in terms of energy consumption. Even though the cloud hardware consumes more energy than the proprietary BBM units, the difference in the total required baseband processing hardware is so significant that Cloud D-RAN can achieve 20 % lower energy consumption and Cloud C-RAN 40 % lower energy consumption.

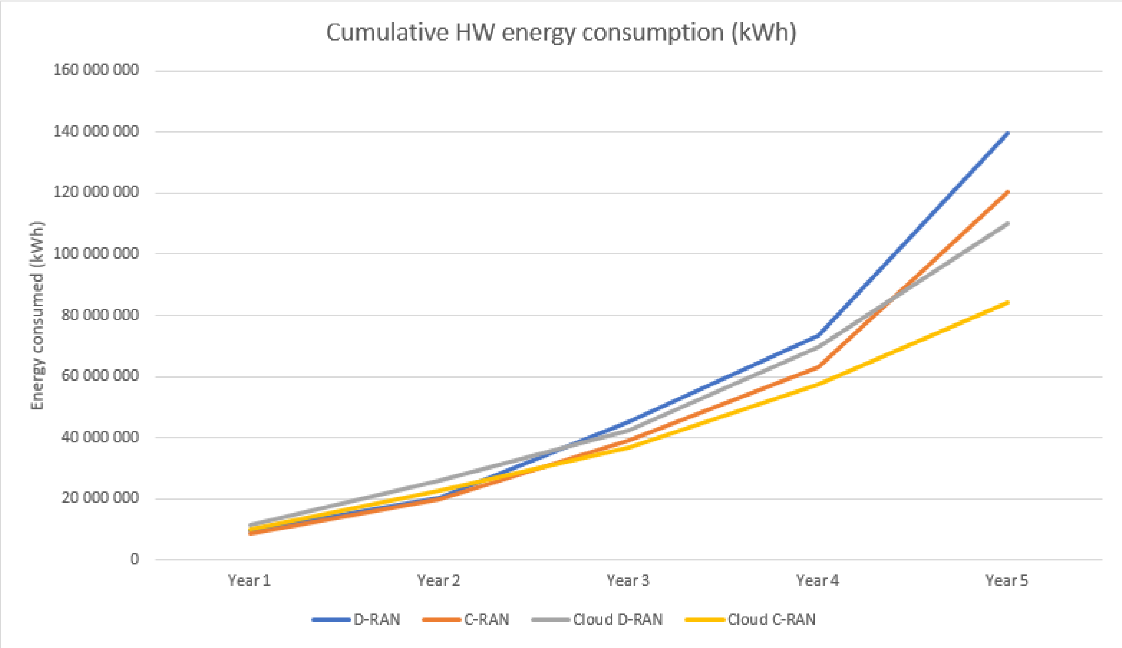

Figure 37: Cumulative energy consumption by the baseband processing hardware in scenario 2.

CAPEX

Analyzing Scenario 2's high C-plane traffic growth in terms of CAPEX (figure 38) shows that Cloud C-RAN had the lowest CAPEX with 14,6 % lower costs than D-RAN. Followed by Cloud D-RAN with 9,5 % and C-RAN with 1,8 % lower CAPEX than D-RAN. Cloud RAN architectures were able to achieve larger CAPEX savings due to cloud hardware's better scaling. Increasing the network's C-plane traffic capacity is much more economical with cloud hardware than with BBMs in scenario 2 as the savings in BBM units is enough to offset the costs caused by the addition of cloud servers and software.

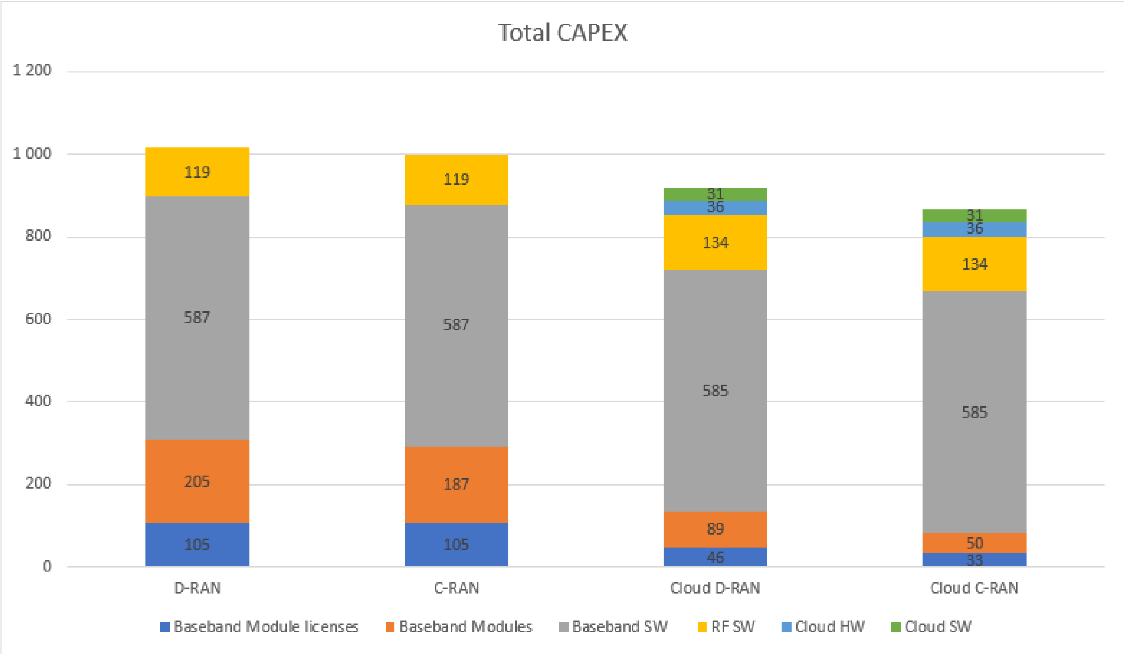

Figure 38: Scenario 2's cumulative CAPEX in the fifth year for the architectures.

OPEX

OPEX analysis of scenario 2 (figure 39) shows that Cloud C-RAN had the lowest costs compared to D-RAN with 25,2 % lower OPEX. C-RAN was able to lower the OPEX by 13,4 % and Cloud D-RAN by 11,5 % compared to D-RAN. Cloud C-RAN achieving the lowest OPEX shows the benefits of virtualizing and centralizing the baseband processing in high C-plane traffic growth scenario. Cloud C-RAN was able to lower costs in almost all areas except having higher software maintenance fee due to more expensive software in Cloud RAN. Baseband capacity upgrade costs, rental costs and energy costs were reduced in C-RAN, Cloud D-RAN and in Cloud C-RAN compared to D-RAN. This was due to less required hardware and the addition of cloud servers did not increase the costs significantly in Cloud RAN architectures. Cloud D-RAN was not able to surpass C-RAN with lower OPEX, because the distributed architecture turned out to be more expensive in terms of OAM costs, baseband capacity upgrade costs and rental costs.

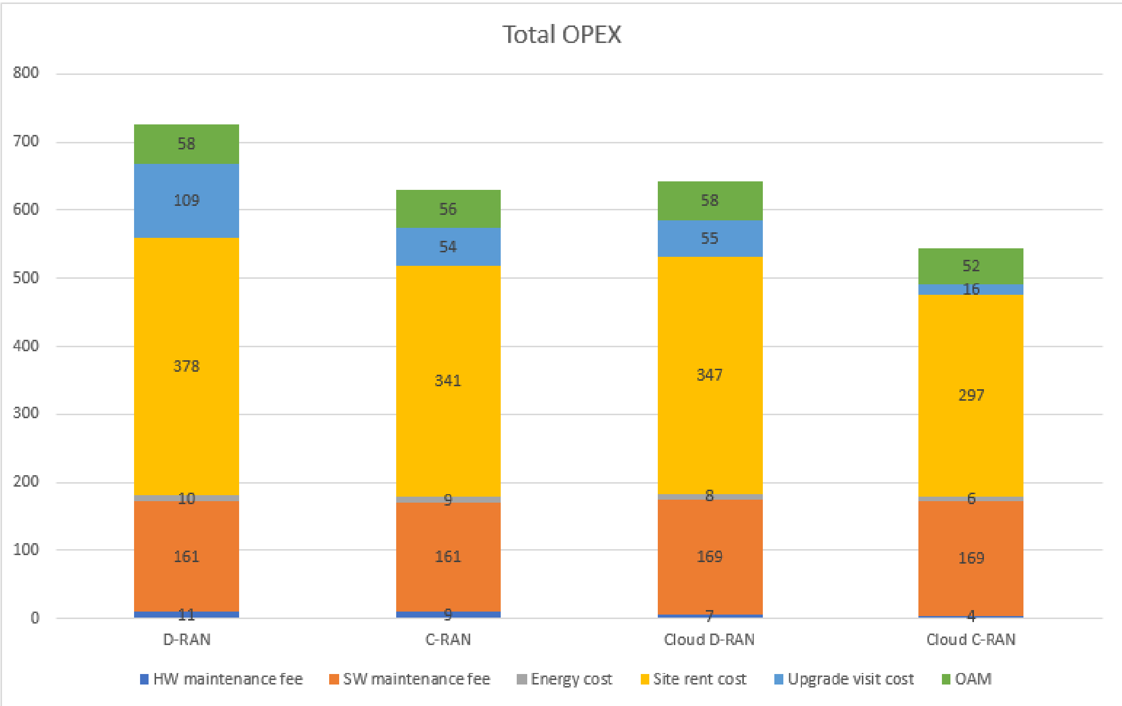

Figure 39: Scenario 2's cumulative OPEX in the fifth year for the architectures.

6 Conclusions

6.1 Results

Results of the scenario 1 show that the Cloud RAN architectures are not the best options in terms of costs with low C-plane traffic growth, which was expected. Cloud D-RAN turned out to be even more expensive than the current D-RAN architecture. On the other hand, Cloud C-RAN was able to offer lower TCO compared to D-RAN, but it was not able to achieve better results than the bare-metal C-RAN. With low C-plane traffic growth, Cloud RAN was not able to leverage the benefits of better hardware scaling for the C-plane traffic. Therefore the benefits of Cloud RAN's baseband processing could not offset the cost of adding cloud servers to the architecture. Scenario 1's 25 % CAGR for CUE per cell site is moderate expectation for the growth rate of devices as technologies such as IoT is still on the horizon. Therefore, when considering only costs, moving to Cloud RAN architecture is questionable.

Results for the scenario 2 show that the Cloud RAN architectures turned out to be the most economical solutions with high C-plane traffic growth expectation. Cloud RAN architectures were able to achieve much better scaling for traffic growth than the bare-metal architectures, which lead to lower costs in CAPEX and OPEX. Due to better hardware utilization, Cloud RAN was also able to achieve much lower energy consumption and proving the capability to offer eco-friendlier RAN. Scenario 2's 65 % CAGR for CUE per cell site is high expectation for the growth rate, but it can be realistic in the future. The number of connected devices is rising constantly and with the forthcoming of IoT the number of connected devices can increase rapidly. In terms of costs, moving to Cloud RAN can offer clear benefits with high C-plane traffic growth.

Breakeven points were also identified through iterative process for Operator A, to solve when Cloud RAN architectures become more economical than its corresponding bare-metal architectures. For Cloud D-RAN the breakeven point was discovered at 59 % and for Cloud C-RAN it was discovered at 46 % CAGR CUE per cell site. At these breakeven points in the network dimensioning, the C-plane traffic becomes a limiting factor instead of the cell connectivity. This then allows the more economical scaling of COTS cloud servers over the proprietary BBMs in Cloud RAN. Results of the cost modeling indicate that moving to cloud based architecture can offer lower costs, when the C-plane traffic is expected to have relatively high growth.

6.2 Assessment of Results

Results of the cost modeling were mostly affected by the C-plane traffic growth and the cell configuration on sites. In the network dimensioning, the C-plane traffic and the cell connectivity requirements determine the amount of baseband processing hardware required in the network. When the C-plane traffic becomes the dominant parameter in the dimensioning, Cloud RAN offers better scaling for the C-plane traffic. In this model the C-plane traffic was modeled with CUE per cell site and C-plane load per CUE. Therefore increasing or decreasing these input parameters changes the output of the model the most.

6.3 Exploitation of Results

Cloud RAN's cost benefits for MNOs are currently dependent on the C-plane traffic growth expectation. Analyzing Cloud RAN in terms of costs is just one perspective for studying all the potential benefits. It does not take into account how Cloud RAN's benefits such as shorter time to market for new software features, improved Quality-of-Experience or new revenue opportunities can affect the business benefits.

Some might argue that there is no need to invest into developing the RAN until the IoT, M2M communications and AR/VR applications are overwhelming the current network. However, playing catch-up in highly competitive marketplace is risky and applications such as Pokémon GO have already emerged, which can put the current RAN under heavy pressure. Investing into the RAN to meet the demands of these new generation applications can also speed-up the application development and end-user adoption. The development of 3G and 4G networks to offer better Quality-of-Service with higher bandwidth and lower latency was one of the reasons why the popularity of mobile applications has exploded. The same thing can happen with these new generation applications if the underlying infrastructure is capable to meet their demands.

For the Network Equipment Providers (NEPs) Cloud RAN threatens to cannibalize the sales of their hardware business. With Cloud RAN, MNOs will require less of the proprietary baseband hardware and they are free to buy the COTS cloud servers from anywhere. Cloud RAN architecture moves the business away from being hardware focused towards software focused business. The future RAN solutions will most likely be mixture of different architectures consisting of D-RAN, C-RAN and Cloud RAN working together in different areas. Therefore, the NEPs hardware business will not disappear, but it is reasonable to expect that the revenues will shrink.

Solution for NEPs to maintain the same or higher revenue while offering Cloud RAN would require them to move towards more software focused business model and gain more revenue through the software business. This could be achieved by raising the price of the software offered with Cloud RAN or by innovating new additional proprietary software features and services that MNOs would want to buy. Raising the price of software is probably the easiest solution, but it might be challenging to justify why MNOs should suddenly need to pay more for their software. On the other hand, innovating new proprietary software features and services in Cloud RAN can potentially offer higher revenues, while giving value for the MNOs at the same time. These new features and services could offer benefits such as lower OAM, energy and baseband capacity upgrade costs to reduce the MNOs OPEX. However, innovating new software features and services is never an easy task. Focusing on lowering the MNO's OPEX with Cloud RAN is more attractive than offering CAPEX benefits for NEPs since then Cloud RAN would not be cannibalizing their own hardware revenues.

Cloud RAN will also affect businesses that will be using digital platforms, mobile applications and IoT devices through the mobile network as part of their business. Laying the foundation for MEC, massive IoT and M2M communication with higher bandwidth, lower latency and available computing resources on the mobile edge Cloud RAN has the potential to accelerate innovation and development of new services, applications and products. These third party applications and services play critical role in creating a market pull for the Cloud RAN. Interacting and co-operating with different businesses to develop new applications and services that are made possible by the Cloud RAN can help to create a demand, which will then push MNOs to invest into Cloud RAN.

6.4 Future research

The cost modeling of Cloud RAN in this thesis provides an approximation of costs between the different architectures, which then can be used to understand Cloud RAN's value proposition. The model focuses on analyzing the costs and TCO of the different architectures, which is just one perspective for analyzing Cloud RAN's possible benefits. It leaves out other value adding features such as better QoS for the users and new possible revenue sources that can come with Cloud RAN deployment. The model also does not consider transport costs for renting or building a fiber network for the Cloud RAN, which were left out due to the costs being really MNO and region specific. Future work on understanding Cloud RAN's value proposition could include research on how these things affect the costs and what possible new revenue sources could be realized in Cloud RAN.

References

[1] Cisco Visual Networking Index. “Global mobile data traffic forecast update, 2016-2021”. In: *Cisco white paper* (2016).

[2] Kavi Bains and Mark Giles. “Expanded financials coverage shows mixed fortunes for developed and developing mobile markets”. In: *GSMA Intelligence research note* (2015). URL: https://www.gsmaintelligence.com/research/2015/09/expanded-financials-coverage-shows-mixed-fortunes-for-developed-and-developing-mobile-markets/523/.

[3] GSMA Intelligence. “The Mobile Economy 2017”. In: *Research report* (2017). URL: http://www.gsma.com/mobileeconomy/.

[4] China Mobile. “C-RAN: the road towards green RAN”. In: *White Paper, ver* 2 (2011).

[5] Yasir Zaki. *Future mobile communications: LTE optimization and mobile network virtualization*. Springer Science & Business Media, 2012.

[6] Theo Dunnewijk and Staffan Hultén. “A brief history of mobile communication in Europe”. In: *Telematics and Informatics* 24.3 (2007), pp. 164–179.

[7] Jörg Eberspächer et al. *GSM-architecture, protocols and services*. John Wiley & Sons, 2008.

[8] Qualcomm. “The Evolution of Mobile Technologies: 1G, 2G, 3G, 4G LTE”. In: *Qualcomm.com* (2014). URL: https://www.qualcomm.com/media/documents/files/the-evolution-of-mobile-technologies-1g-to-2g-to-3g-to-4g-lte.pdf.

[9] Peter Curwen and Jason Whalley. *Fourth Generation Mobile Communication: The Path to Superfast Connectivity*. Springer Science & Business Media, 2013.

[10] Harri Holma and Antti Toskala. *LTE for UMTS: Evolution to LTE-advanced*. John Wiley & Sons, 2011.

[11] Fujitsu. “Centralized Radio Access Network (C-RAN) Transport”. In: *Fujitsu Application Note* (2017). URL: http://www.fujitsu.com/us/Images/Centralized-Radio-Access-Network-C-RAN-Transport-Application-Note.pdf.

[12] Nick Marshall and Hoffman Joe. “NFV: RADIO VIRTUALIZATION IN THE RAN”. In: *ABI Research white paper* (2015).

[13] Peter Rost et al. “Cloud technologies for flexible 5G radio access networks”. In: *IEEE Communications Magazine* 52.5 (2014), pp. 68–76.

[14] Nokia. “Evolution to centralized RAN with mobile fronthaul – Technology overview”. In: *White Paper* (2016).

[15] GSMA Intelligence. “Understanding 5G: Perspectives on future technological advancements in mobile”. In: *White paper* (2014), pp. 1–26.

[16] Michael J Marcus. “5G and" IMT for 2020 and beyond"[Spectrum Policy and Regulatory Issues]”. In: *IEEE Wireless Communications* 22.4 (2015), pp. 2–3.

[17] Ericsson. "5G Radio Access capabilities and technologies". In: *Ericsson White paper* (2016). URL: https://www.ericsson.com/assets/local/publications/white-papers/wp-5g.pdf.

[18] Ericsson. "Cloud RAN". In: *White Paper* (2015).

[19] Aleksandra Checko et al. "Cloud RAN for mobile networks—A technology overview". In: *IEEE Communications surveys & tutorials* 17.1 (2015), pp. 405–426.

[20] Nokia. "Nokia AirScale Cloud RAN". In: *White Paper* (2017).

[21] Huawei. "Cloud RAN Introduction". In: *The 4th CJK International Workshop –Technology Evolution and Spectrum* (2011).

[22] Intel Cor. "Intel heterogeneous network solution brief". In: *Solution brief, Intel Core Processor, Telecommunications Industry* ().

[23] ZTE. "ZTE Green Technology Innovations White Paper". In: *White Paper* (2011). URL: http://www.zte.com.cn/mi_imgs/global/investor_relations/353156/P020120918593482919117.pdf.

[24] Tom Flanagan et al. "Creating cloud base stations with TI's KeyStone multicore architecture". In: *Texas Instruments White Paper* (2011).

[25] J. Huang et al. "Overview of cloud RAN". In: *2014 XXXIth URSI General Assembly and Scientific Symposium (URSI GASS)*. 2014, pp. 1–4. DOI: 10.1109/URSIGASS.2014.6929243.

[26] Lifeng Wang et al. "Secrecy and energy efficiency in massive MIMO aided heterogeneous C-RAN: A new look at interference". In: *IEEE Journal of Selected Topics in Signal Processing* 10.8 (2016), pp. 1375–1389.

[27] Georgios Kardaras and Christian Lanzani. "Advanced multimode radio for wireless & mobile broadband communication". In: *Wireless Technology Conference, 2009. EuWIT 2009. European*. IEEE. 2009, pp. 132–135.

[28] Bernd Haberland et al. "Radio base stations in the cloud". In: *Bell Labs Technical Journal* 18.1 (2013), pp. 129–152.

[29] I Chih-Lin et al. "Recent progress on C-RAN centralization and cloudification". In: *IEEE Access* 2 (2014), pp. 1030–1039.

[30] CPRI Specification. "V6. 0". In: *Common public radio interface (CPRI)* (2013).

[31] OBSAI. "Open Base Station Architecture Initiative (OBSAI) BTS System Reference Document Version 2.0". In: (2006).

[32] Aleksandra Checko. "Cloud Radio Access Network architecture. Towards 5G mobile networks". PhD thesis. Technical University of Denmark, 2016.

[33] Aleksandra Checko et al. "Evaluating C-RAN fronthaul functional splits in terms of network level energy and cost savings". In: *Journal of Communications and Networks* 18.2 (2016), pp. 162–172.

[34] 3GPP. "Study on new radio access technology: Radio access architecture and interfaces". In: *Technical report, Specification 38.801 release 14* (2016).

[35] VMWare. “Virtualization Overview”. In: *VMWare White Paper* (2006). URL: https://www.vmware.com/pdf/virtualization.pdf.

[36] Telecommunication engineering center FN Division. “Network Function Virtualization (NFV) Its impact on Future Telecom Networks”. In: *Department of Communications, Government of India. White Paper* (). URL: http://tec.gov.in/pdf/Studypaper/Network_Function_Virtualization%20.pdf.

[37] Tieto Intel. “Realizing the Benefits of Network Functions Virtualization in Telecoms Networks”. In: *Tieto Intel White Paper* (2014).

[38] Rui Wang, Honglin Hu, and Xiumei Yang. “Potentials and challenges of C-RAN supporting multi-RATs toward 5G mobile networks”. In: *IEEE Access* 2 (2014), pp. 1187–1195.

[39] Open Networking Foundation (ONF). “Software-Defined Networking: The New Norm for Networks”. In: *ONF White Paper* (2012). URL: https://www.opennetworking.org/images/stories/downloads/sdn-resources/white-papers/wp-sdn-newnorm.pdf.

[40] Nokia. “Multi-Layer and Cloud-Ready Radio Evolution Towards 5G.” In: *White Paper* (2016).

[41] Kevin Murphy. “Centralized ran and fronthaul”. In: *White Paper, Ericsson* (2015).

[42] Gabriel Brown. “Mobile Edge Computing use cases and deployment options”. In: *Heavy Reading and Juniper Networks White paper* (2016). URL: https://www.juniper.net/assets/us/en/local/pdf/whitepapers/2000642-en.pdf.

[43] Yun Chao Hu et al. “Mobile edge computing—A key technology towards 5G”. In: *ETSI White Paper* 11.11 (2015), pp. 1–16.

[44] Intel. “Real-World Impact of Mobile Edge Computing (MEC)”. In: *Intel White paper* (2016). URL: https://builders.intel.com/docs/networkbuilders/Real-world-impact-of-mobile-edge-computing-MEC.pdf.

[45] Fujitsu. “The Benefits of Cloud-RAN Architecture in Mobile Network Expansion”. In: *Fujitsu White paper* (2014). URL: http://www.fujitsu.com/downloads/TEL/fnc/whitepapers/CloudRANwp.pdf.

[46] Statista. “Google Play: number of Android app downloads 2010-2016”. In: *Statista.com* (2017). URL: https://www.statista.com/statistics/281106/number-of-android-app-downloads-from-google-play/.

[47] XRAN.org. “The mobile access network beyond connectivity”. In: *XRAN.org White paper* (2016). URL: https://static1.squarespace.com/static/57e95f5abe65943a7a209f26/t/57fd5746ebbd1a584a57ff7e/1476220770104/XRAN+%E2%80%93+Mobile+Access+Network+Beyond+Connectivity+20-161011.pdf.

[48] Nokia. "Finding Mewtwo. What the Pokémon Go phenomenon means for mobile operators". In: *Nokia White paper* (2016). URL: https://resources.ext.nokia.com/asset/200545.

[49] Nokia. "Rethinking your radio". In: *Nokia E-Book* (2016). URL: https://pages.nokia.com/1753.Airscale-Ebook.html.

[50] ABI Research. "ABI Research Projects 5G Worldwide Service Revenue to Reach $247 Billion in 2025". In: *ABI Research report* (2016). URL: https://www.abiresearch.com/press/abi-research-projects-5g-worldwide-service-revenue/.

[51] Christos Bouras, Panagiotis Ntarzanos, and Andreas Papazois. "Cost modeling for SDN/NFV based mobile 5G networks". In: *Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), 2016 8th International Congress on*. IEEE. 2016, pp. 56–61.

[52] MM Rahman, Charles Despins, and Sofiene Affes. "Analysis of CAPEX and OPEX benefits of wireless access virtualization". In: *Communications Workshops (ICC), 2013 IEEE International Conference on*. IEEE. 2013, pp. 436–440.