Sanna Alikoski

Improving the monitoring of telephone network with Analytics

Helsinki Metropolia University of Applied Sciences

Bachelor of Engineering

Industrial Management

Thesis

23 May 2017

Metropolia

Author Title Number of Pages Date	Sanna Alikoski Improving the monitoring of telephone network with Analytics 37 pages 15 May 2017
Degree	Bachelor of Engineering
Degree Programme	Industrial Management
Specialisation option	International ICT-Business
Instructor(s)	Anna Sperryn, Senior Lecturer Markus Virtanen, Data Scientist

The objective of this thesis was to propose recommendations for Elisa on how to improve the monitoring of the network performance. These recommendations were made based on an analysis of data gathered from the network. The analysis was carried out based on research questions defined with Elisa.

The current state analysis was made based on discussions with the professionals responsible for parts of the Network includes to this thesis. Discussions concentrated in the functioning of the network and the forms of analytics practiced in the field so far.

Literature regarding analytics, its use in telecommunication, and different tools and processes for practicing it was explored to gain full understanding on how analytics could help the company in their network monitoring.

The analysis was made based on the process model described in the literature. The analysis was made from data gathered from the company's databases. Findings from the analysis were reflected on the research questions. On the basis of the answers gained, the proposal for monitoring the network with analytics was constructed. The proposal was to implement Real Time analysis for network monitoring, and invest in monitoring of the key network components.

The results from the analysis give Elisa vital information about the behavior of the network. With this information, the company can concentrate further studies on the challenges discovered. The proposal presents a direction for the implementation.

Keywords	Analytics, Telecommunications, Network Monitoring

Metropolia

Tekijä Otsikko Sivumäärä Aika	Sanna Alikoski Puhelinverkon valvonnan parantaminen analytiikan avulla 37 sivua 15. toukokuuta 2017
Tutkinto	(Insinööri AMK)
Koulutusohjelma	Tuotantotalous
Suuntautumisvaihtoehto	Kansainvälinen ICT-liiketoiminta
Ohjaajat	Lehtori Anna Sperryn Data Scientist Markus Virtanen

Insinöörityön tavoitteena oli tarjota Elisalle ehdotuksia siitä, kuinka yritys voi parantaa puhelinverkkonsa toimivuuden valvontaa analytiikan avulla. Suositukset pohjautuivat verkkodatasta tehtyyn analyysiin. Analyysi tehtiin Elisan kanssa määriteltyjen tutkintakysymysten pohjalta.

Nykytilan kartoitus tapahtui käymällä keskusteluja verkkoasiantuntijoiden kanssa. Nämä keskustelut kattoivat sekä verkkotekniikkaa, että tietoa analytiikan tähänastisesta roolista verkon valvonnassa.

Analytiikasta, sen soveltamisesta telekommunikaatio-alalla sekä työkaluja ja prosesseja sen harjoittamisesta tutkittiin kirjallisuutta. Tämän avulla pystyttiin saavuttamaan ymmärrys siitä, miten analytiikkaa voisi soveltaa verkon valvonnassa Elisalla.

Analyysi tehtiin pohjautuen kirjallisuudesta löytyvään prosessin. Analysoitava data saatiin Elisan tietokannoista. Analyysin tuloksia peilattiin aiemmin tehtyihin tutkintakysymyksiin, ja niiden pohjalta tehtiin kehitysehdotukset verkon valvontaan analytiikan avulla. Elisan tulisi implementoida ennakoivaa analytiikkaa verkon valvontaan. Implementointi tulisi aloittaa verkon keskeisimmistä komponenteista, jotta siitä saataisiin suurin hyöty.

Analyysin tulokset antoivat Elisalle tärkeää tietoa verkon toimivuudesta. Tiedon avulla asiaa pystytään tutkimaan lisää. Kehitysehdotukset taas antoivat Elisalle suunnan, josta aloittaa analytiikan hyödyntäminen verkon valvonnassa.

Keywords	analytiikka, telekommunikaatio, verkon valvonta

Metropolia

Contents

List of Abbreviations

ABR	Area Border Router. A router in a Fixed Network which operates as a gate when entering or exiting a certain area of the network.
ADSL	Asymmetric Digital Subscriber Line. Type of digital Subscriber Line that connects to the Internet.
AGN	Aggregation Node.
BI	Business Intelligence.
CEM	Customer Experience Management. A collection of processes the company uses to track, supervise every customer interactions.
LSP	Label Switched Path. Former established path for data packets traveling through routers in the network.
LTE	Long Term Evolution. Also known as 4G-technology.
MAG	Mobile Aggregation Node
MPLS	Multi-Protocol Label Switching. Network that routes each data packet by former established routes instead of routers making independent decisions where to forward the packet.
TWAMP	Two-Way Active Measurement Protocol. A protocol for measuring network performance between two devices connected to the network.
TN	Transfer Node.
WCDA	Wideband Code Division Multiple Access. Also known as 3G technology, or third generation mobile network technology.

List of Key Concepts

Analytics	Science of examining raw data to gain answers to certain questions by reaching different conclusions with the information gained.
Fixed Network	Part of phone network where all the devices are connected to each other with physical cables and wires.
Mobile Network	A radio network consisting of local transceivers called base stations. All these stations are together providing the network coverage for larger areas.
Telecommunication	A term for various technologies to transmit information as electrical signals.
Tele Operator	A company which is responsible for establishing and maintaining the information flow in the network.

1 Introduction

Analytics has been a hot topic for the past few years. Everybody is talking about digitalization and the importance of making decisions based on data. In fact, the amount of data gathered is doubled in every two years, and by 2020 we will have approximately 50 times more data in our hands than we do now.

Now companies are facing the large question of should they make exploiting of data part of their future strategy or not. If they do, the next step is to understand what kind of data is important for them, how they should analyze it and which assessments to make from it.

This thesis focuses on how Elisa Oyj can improve their customers' experience by getting better call reliability with better network monitoring. The thesis will be carried out by analyzing the data from different components of their network and making recommendations based on the findings.

1.1 Elisa Oyj

Elisa Oyj is a Finnish telecommunication, ICT and online service provider which has 2.3 million consumers, companies and public administration organizations as its customers. In Finland, Elisa is the market leading provider of mobile subscriptions. The main market area is Finland but the company also has operations in Estonia. (On Elisa, 2016)

Elisa has a long history, starting from 1882 when Daniel Johannes Wadén established the first telephone company in Helsinki. Throughout the long history, the company has expanded and developed to its current state. Elisa has been a forerunner for various telephone technologies. In fact, Helsinki was the first city in the world where the switchboards were automated in 1929, and the first GSM phone call was transmitted through the network of Elisa's subsidiarys Radiolinja. (Elisa History, 2016)

Elisa has grown from one man's telephone company to one of the Finland's leading ICT companies and in 2015 Elisa had 1.57 billion Euros revenue and the strength of the staff was 4,100. (On Elisa, 2016)

The organization consists of two business units which are consumer customers unit and corporate customers unit, production unit and support functions (Elisa organization, 2016).

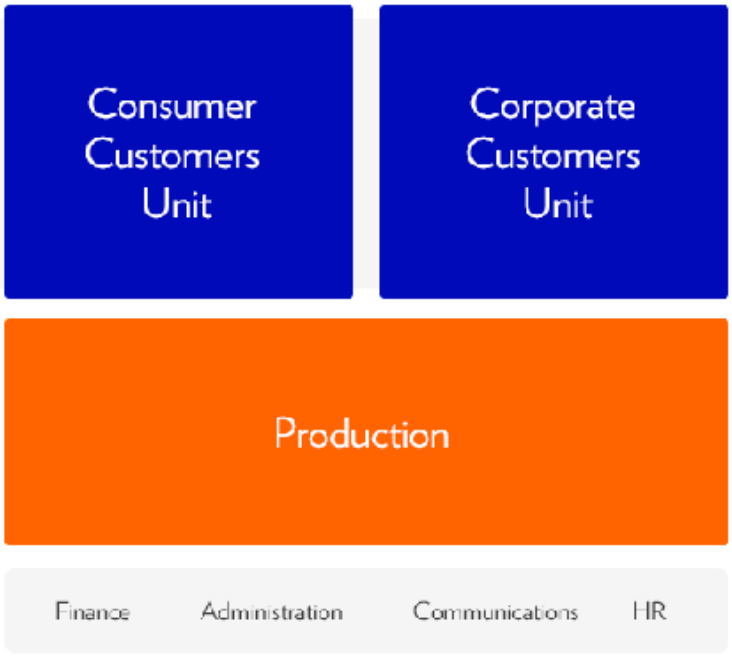

Figure 1. Elisa's organizational structure (Elisa organization, 2016).

This thesis is done for the production unit. This unit is meant to operate in a guidance of Elisa's two business units in achieving the common goals. The production unit is responsible for planning, building and sustaining telecommunication and IT networks. Responsibilities also include, for example, service maintenance, availability management, defect management, producing and implementing new services and logistics. Production unit consists of nine subunits which are Service Management, Mobile Network Services, Fixed Network Services, IT Platform and Hosting Services, Software Services, Delivery, Carrier Services, Technology & Architecture and Company IT. The unit that officially commissioned this thesis project is Fixed Network Services but it also includes factors from Mobile Network Services.

1.2 Business challenge

Elisa has the most comprehensive network throughout the whole Finland. This means that the network has a vast number of different components. All these different components are producing data. The company has its own unit to take care that every single one of these components works as they should. The network consists of fixed and mobile network, both equally important but support different technologies. Therefore, there are different professionals responsible for each of them.

To establish a phone call from one phone to another, there needs to be a data flow through both sides of the network. If something unexpected happens on the route it effects on customers experience by weakening the quality of the call. In large infrastructures like this, it is certain that sometimes something goes wrong. It is important for Elisa that all their customers can enjoy flawless connections at any given time, and they want to develop their network and processes continuously to ensure the best service for their customers.

In this thesis, the challenge is to gain knowledge about the network by data modeling methods and analyses. Through data modeling, one can assess different routes of the data packets to gain better understanding of the network and so on, improve the customer experience.

1.3 Objective of the Thesis

The objective of this thesis is to propose recommendations for Elisa on how to improve monitoring of the network performance. These recommendations are made based on an analysis of data gathered from the network.

The analysis will focus on discovering dynamic routes data packets are traveling in, but also errors and dysfunctions affecting their flow throughout the network. The analysis will be carried out based on the research questions defined in the table one.

Research questions of this thesis

1. **Which are the exact routes for each data packet traveling through the network, and how time consuming they were?**
2. **Is utilization level and number of errors increasing when the paths get longer?**
3. **Is there a connection between Two Way Active Measurement Protocol (TWAMP) measures and dysfunctions in Fixed Network devices?**
4. **Is there a certain type of equipment which is more likely to cause errors?**

1.4 Outcome of the thesis

The outcome of this thesis is a report answering the research questions stated earlier and recommendations for further actions. These recommendations are based on information found from the data analysis executed during this thesis.

In the report, Elisa can see which are the factors decreasing the efficiency of the network. This will help Elisa to give better experience to its customers by ensuring better call reliability.

1.5 Structure of the thesis

Second chapter of this thesis is for determining the methods of this thesis. The chapter will clarify how this thesis project has been carried out, and the actual process flow will

be showed. In the third chapter, the current state of the company will be explained as well as the reason they want this type of thesis to be conducted. To gain the best understanding of the current state, a number of meetings and interviews were held with professionals from each field of network technology regarding this thesis.

The fourth and fifth chapters of this thesis are focusing on the theory that helps to carry out the thesis. These chapters will explain theory about Telecommunication operators' operations and the best practices of analytics in the field of Telecommunications. On the basis of these best practices, the analysis will be carried out in chapter six. In the same chapter, the findings will also be reported. Recommendations on how to monitor Elisa's phone network better by analyzing data will be given in the seventh chapter. A conclusion of everything found and learnt in this thesis will be in the eighth chapter.

2 Method

In this chapter, the thesis method is introduced. The approach chosen for this research and the causes for the choice made are introduced in the first section. In the second section, the process flow chart is presented, and there will also be more information about how the thesis is going to be carried out. Finally, in the third chapter the reliability and validity of this thesis is assured.

2.1 Research Approach

This thesis project was carried out as a case thesis. A case thesis is a method for exploring a single phenomenon in a real-world perspective. The reason this method was chosen is that the thesis is about a unique case and the goal is to obtain deep understanding of it through discovering solutions to the research questions. (Collis & Hussey 2009: 82)

Best practices in analytics, company's internal data bases and the results of the analysis where used to gain the best understanding of the data and finding the best possible answers to the research questions.

2.2 Research Design

The process flow chart used in this thesis is presented below in Figure 2.

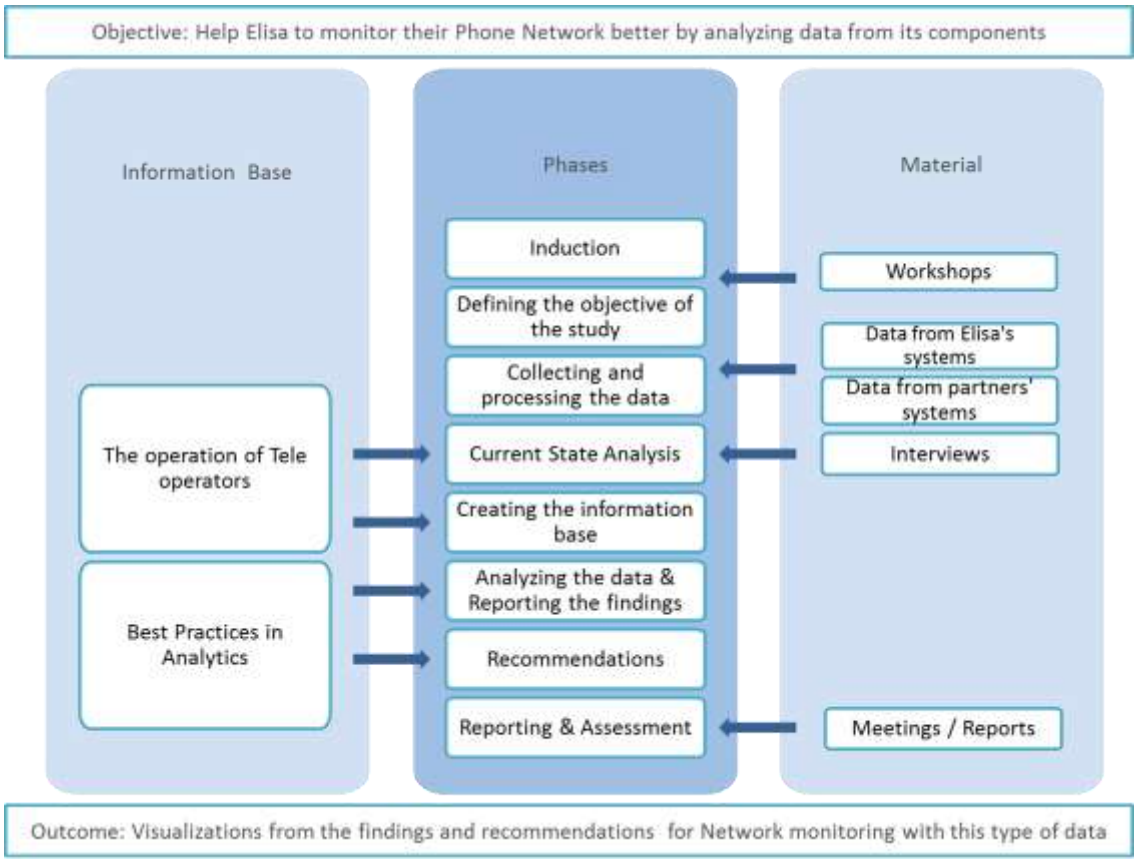

Figure 2. Process Flow of the Case Thesis.

The objective and outcome of the thesis is defined on top and bottom of the chart. On the left side, the information base is defined. The studied theory and best practices applied to the thesis is also displayed on the left side. On the right side of the chart, there is all the material used to execute the thesis. The actual process flow is displayed on the middle of the chart. It places all the steps needed to get the from objective to the outcome.

The thesis case was introduced in various workshops where professionals from different fields explained their part of the process. The objective for the thesis was also defined in these workshops. The actual data for the analysis were taken from the company's own information systems as well as their partners' Information Systems. After that, the data was structured, and research questions were created based on the data and the opinions of different professionals who are responsible for each area of data included to the analysis. The current state of this process was determined through different meetings where these experts expressed their requests and opinions.

Current state analysis led the way to research suitable theory and finding the best practices to carry out this thesis. The next phase was to analyze the data. This was carried out with an analytic tool by Microsoft, called Power BI. With Power BI, the relationships of different data sources were established and required calculations were made. The analysis itself were carried out by selecting data from different sources and making calculations with the software on the correlations, and making clear visual charts from them. These visualizations and the information gained from them were reported. Based on these results found through the analysis recommendations were made for Elisa on how they should monitor their network, and on which type of analysis they should make from the data gained from different components of the network.

2.3 Reliability and Validity

Reliability describes the reliability of methods and results of the thesis. Generally, the reliability is good and therefore the thesis is reliable when the same thesis would have the same result when done again. (Eskola, Kannas, Mustajoki, Välimaa, 2013) In this thesis, reliability is assured by making a mathematical model with the analytic software for the data analysis. The model is not dependent on the values of data and can be implemented to a same type of data with new values.

Validity of the thesis describes did the thesis actually measure what was suppose d to. (Eskola, Kannas, Mustajoki, Välimaa, 2013) In this thesis, validity is assured by defining the research questions based on different professionals' requirements.

3 Current State Analysis

This chapter has removed due to non-disclosure agreement.

4 Analytics

4.1 General

In todays' digitalized environment it is necessary for companies to comprehend the value of data. The transformation from old processes and methods to new ways of thinking is affecting every single industry. This means that the companies which are going to be successful in the future are the ones that embrace rich understanding of exploiting data in their business. (Markkula & Syväniemi, 2015)

To gain understanding from the data, the company should implement analytics in to their business processes. Analytics is a science of examining raw data in various methods.

The key of implementing analytics to a business is to understand that it is not an intrinsic value but a tool to obtain benefit when used right. The chance to succeed gets better when analytics is driven by the business need. (Markkula & Syväniemi, 2015)

The most common mistakes in exploiting analytics are paying too much attention to one technology or data and closing eyes from other options, putting too much effort to project that doesn't bring much value and trying to do too much at once.

4.1.1 Benefits

As said before, analytics works as a tool that helps a business to gain more value from data. Due to data can be analyzed in various ways, there are naturally different kinds of benefits from it.

The largegest benefit of analytics is a more efficient and reliable decision-making process, Due to it is based on facts under the information gained from the data instead of estimations.

Some other benefits gained from analytics are lower cost and higher efficiency, better risk management, ability to forecast market changes better, get customer insights and guide the business through challenging times.

4.1.2 Business Intelligence

Business Intelligence is generally described as the first era of analytics even though it is still widely used today. BI is mainly focused on reporting what happened in the past. The companies using Business Intelligence as their analytical method generally store their data in data warehouses or databases Due to the amount of data is still manageable. Usually also the structure of data is statistic enough to analyze, and does not require too much preparing.

This type of analytics is generally done manually and can take weeks or even months. Therefore, it often doesn't offer much explanations or predictions. (Thomas H. Davenport, 2013)

4.1.3 Big Data

Big Data is nowadays a widely used term. Everyone seems to be excited about it, and there are many definitions on what it means. SAS Institute describes it as follows:

Big Data is a term that describes the large volume of data – both structured and unstructured – that inundates a business on a day-to-day basis.

On the other hand, IBM describes it a little differently:

Big Data is being generated by everything around us always. Every digital process produces it. Systems, sensors and mobile devices transmit it.

In conclusion, Big Data is a term for large amount of data coming from different sources. One well known way of describing Big Data is four Vs: volume, velocity, variety and veracity. Volume means that the data is coming from various sources and the amount of data gathered every day is getting largeger all the time. Velocity means that the data is being gathered really fast and must also be dealt with in nearest real time as possible. Multiple different sources of data create variety, which means that the data can be in almost any format from easy read structured numerical data to more difficult unstructured data like text documents, video or audio. One more aspect of Big Data is veracity, which means the uncertainty of data. When there is a large amount of different data, it might

be difficult to actually know what it actually consists of, or where it comes from. This often causes problems to trust the data gathered. (Roponen, 2015)

Due to the large amount of various forms of data, the linking and matching of them becomes more laborious. Even so, it is very important to still manage the data and maintain connected relationships, hierarchies and linkages. Otherwise, the risk of losing control gets quickly higher.

To get the benefit from Big Data, the company must understand it is not the large amount of data itself that brings the value but knowing how to exploit it. Some ways of getting the advance from the data are real time analysis to find root causes of failures, issues and defects and detect frauds before they have a chance to affect the company. (SAS Institute, 2016)

4.2 Use in Telecommunication

In this chapter, the short history and some opportunities of analytics in telecommunication are presented. In addition, there will be an assessment of what the industry's stage in analytics is and what are the main challenges now.

4.2.1 Short History

Telecommunication industry is interesting when it comes to analytics Due to it has always operated around data. All the way from Morse's telegraphed message in 1844 to today's Facebook-chatting, it is all based on data flows. By 1990s, there was already more than one million servers connected to the Internet, and by 1998, there was a packet switching network created which was able to transfer voice, video and data simultaneously in a single phone line. In the past 20 years, the industry has developed in a vast phase and today the amount of information about users, usage, network operations etc. Tele Operators can access is overwhelming. (MastersInDataScience, 4/2016)

4.2.2 Opportunities

Dr.Volkmar Scharf-Katz, chief of IP Networks and head of Service Platform at Vodafone US Inc. said “We have a tremendous amount of data that is generated every single second in a network.” (Kelly Hill, 4/2016)

Due to the large amount of data tele operators are handling today, the current direction of analytics is moving from Business Intelligence to Big Data. Naturally the industry is old and this kind of large change can be slow. However, there are many opportunities waiting for the operators to reach in. In the figure 3. there are described how the operators can advantage from Big Data throughout the whole value chain.

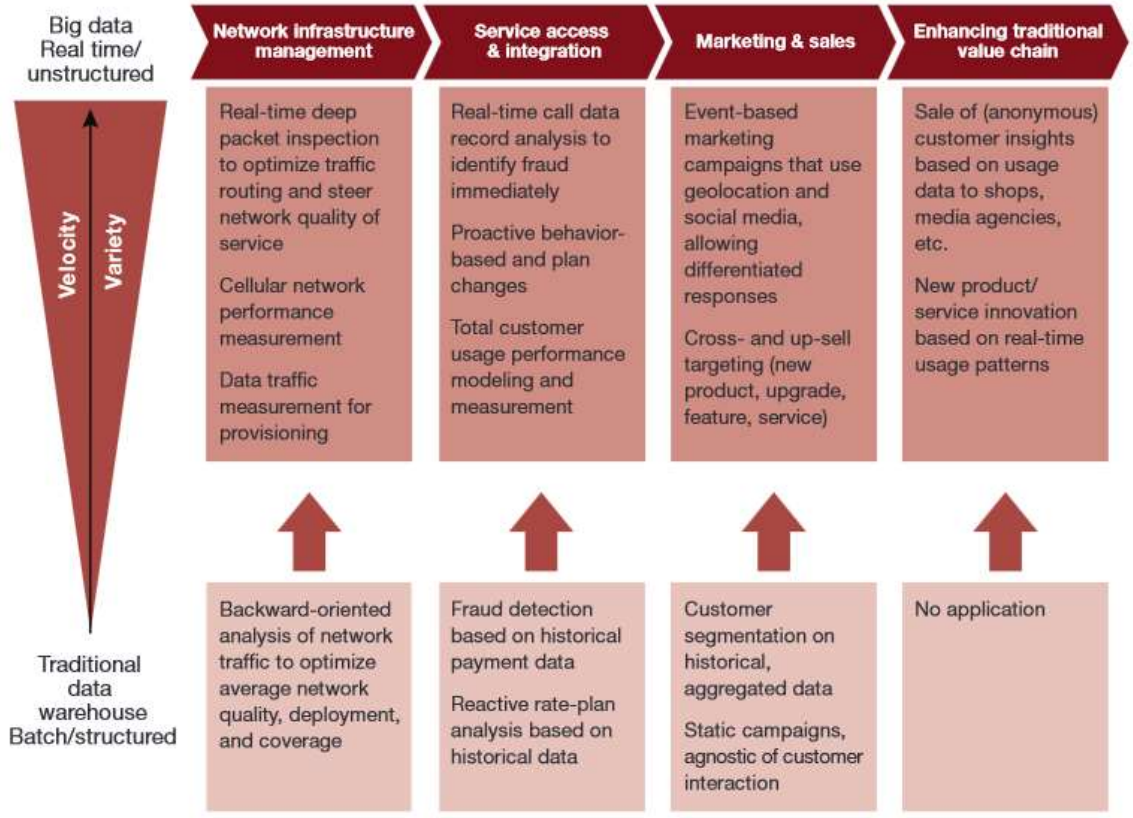

Figure 3. Benefits of Big Data Analytics through the Telecom value chain (Pwc, 2013)

At the bottom of the figure, the old ways of Business Intelligence are shown, and on the top the new opportunities opening with Big Data are presented. Networks Infrastructure Management has always been a fundamental part of Tele Operators business, Due to

the network is the core factor that enables the whole operation of the company. Therefore, there has always been analytics performed related to the network. However, in the large network which has thousands of components and continuous data flow, analytics that are concentrating on the past might not be that useful all the time. It is also a fact that there are loads of accessible data in the network waiting to be analyzed in real time. With Big Data Analytics, the operators are able to optimize their routing and measure their network performance and quality by analyzing the traffic flowing through their network in real time. (Pwc, 2013)

Real-time analyses from the network can also be used for damage control. For example, if the network goes down all the departments are able to react to it by finding the customers who have been affected from the failure and try to fix the problem immediately.

Operators can also prepare for these failures of the network by analyzing big data. For example, combining their internal data about the network capacity and external data about schedules of large music festivals, they can anticipate the possible dropping of network and prepare more capacity to that certain area for the time of the festival. (MastersInDataScience, 4/2016)

In the field of service access & integration, operators can use Big Data for detecting fraud in their system. The types of fraudulent behavior Operators face can be for example looping or call forwarding on hacked private branch exchanges or swapping of SIM cards. If this type of analysis is done afterwards it naturally slows down the company's ability to react to frauds which might even impact the company's revenue and even brand image. Therefore, they should exploit Big Data Analytics by combining their internal data like location, usage and account information and external data like credit reports. (PwC, 2013)

Operators can also use Big Data for customer usage performance modeling and proactively change plans based on customer behavior.

In the field of Marketing & Sales Operators can use Big Data for more efficient event based marketing. For example, returning to the same example of music festivals, the operator can use the geolocation of the visitors in the festival and make targeted marketing for them in social media.

Lastly, Operators can benefit from Big Data by enhancing the whole traditional value chain. This means creating new streams of income that were not available before the ability to use the data. An example for this kind of new revenue stream is to sell anonymous customer insights to different third-party companies and agencies. However, this is not the only way to create more income. The large amount of data can be a stepping stone for all kinds of new innovative services and even products. (PwC, 2013)

4.2.3 Aspects to consider

As said before, there are lots of different opportunities and ways for telecom operators to benefit from the large amount of data they are gathering all the time. Nevertheless, there are some factors needed to take into consideration when implementing Big Data analytics in to their businesses.

To get most from the data, Operators need to find a way to manage the complexity of it. The volume of the data is a factor itself but also the mix between structured and unstructured data is making analyses more complicated. The structured data is already quite well used for certain areas of business. For example, information about data usage is generally used in billing. However, this type of data could be used in other areas of business as well to create new value. (HP Enterprise, 2016)

When it comes to unstructured data, the analysis becomes more difficult Due to the complexity of the data gathered from different sources. Trying to combine data from different sources usually creates a bottleneck Due to the different data sources do not communicate. For example, there is a drop in a network and customer is not able to download a picture. Naturally the operator wants to find the root cause for the problem in real time but this is often challenging Due to the networks' size and data coming from different sources which are not correlated. The solution for this problem is to collect the data at the lowest level possible and then bring it all into the same data storage. Then, the correlation can be done for all data before the actual incident happens, making the analysis much easier and faster when needed. (Dr. Srinivasa Vegi, 2014)

One aspect that the operators should consider when making analytics from big data, especially when the data is related to customers, is privacy. Operators need to find a way to get the most from the data but at the same time be very careful of not leaking any

sensitive information in public. The privacy of this type of data is also strictly guarded by officials and making a violation may cause serious harm for the company and its brand. (MastersInDataScience, 4/2016)

Last aspect is the organizational structure. Generally, telecom operators are massive organizations with long history behind them. Sometimes they are also partly owned by governments. This can easily make the environment to be resistant for changes. Other challenge related to the industry is that usually the organizational structure is built on separate units, almost silos. To benefit the most from the data it should be made sure that it is connected throughout the organization. This means that the organizational structure should be changed from separate silos to a more transparent form where the information flows throughout the whole company. Data should be the rallying point between the units and play a leading role in every decision made. (Robert McCarroll, 2014)

4.3 Tools and processes

In this chapter some tools, methods and processes to take into consideration when doing analytics are presented.

4.3.1 Key Functions and Technologies

When doing an analytical project, there are some functions and technologies that are recommended to use in order to handle the project better and getting the wanted result.

Table 1. is gathering these functions and technologies. As seen in the table, there are six main function and five technologies supporting them. Each of them has its own role in the process of getting the best information out of raw data.

Table 1. The key functions in collecting & handling data and the related technologies (Markkula & Syväniemi, 2015)

	Collection	**Recording**	**Editing**	**Analyzing**	**Modeling**	**Interpretation**
Data Modeling	Defines which data will be collected	Creates a frame of preference for handling and using the data	Defines framework for analyzing the data	Defines rules for analyzing the data	Defines the types and dimensions of the data	Creates an understanding for overall vision
Data Base Technologies	Creates rules and boundaries for collection of the data	Defines the methods for recording the data	Enables editing of the data	Offers the data for exploiting	Guides modeling of the data with its features	
Data Handling	Executes collection of the data from data resources	Writes the data to an information base	Executes editing of the data	Executes quality ensuring for the data		
Analytics	Reads the data for analytical processes	Records the enriched data for analytical processes	Edits the data with analytical tools	Offers ways to find phenomenon from the data	Implements mathematical methods	Offers explanations, predictions and recommendations to support decision making
Reporting			Counts key digits	Offers concrete and visualized information for the end user		Realize data for decision making

4.3.2 CRISP-DM

CRISP-DM (Cross-Industry Standard Process for Data Mining) is an agile process model for analytics. It helps the organization to put analytics in the production in a systematic and business oriented way. The process has certain steps as showed below. However, the steps don't necessarily follow each but are often separate iterative segments.

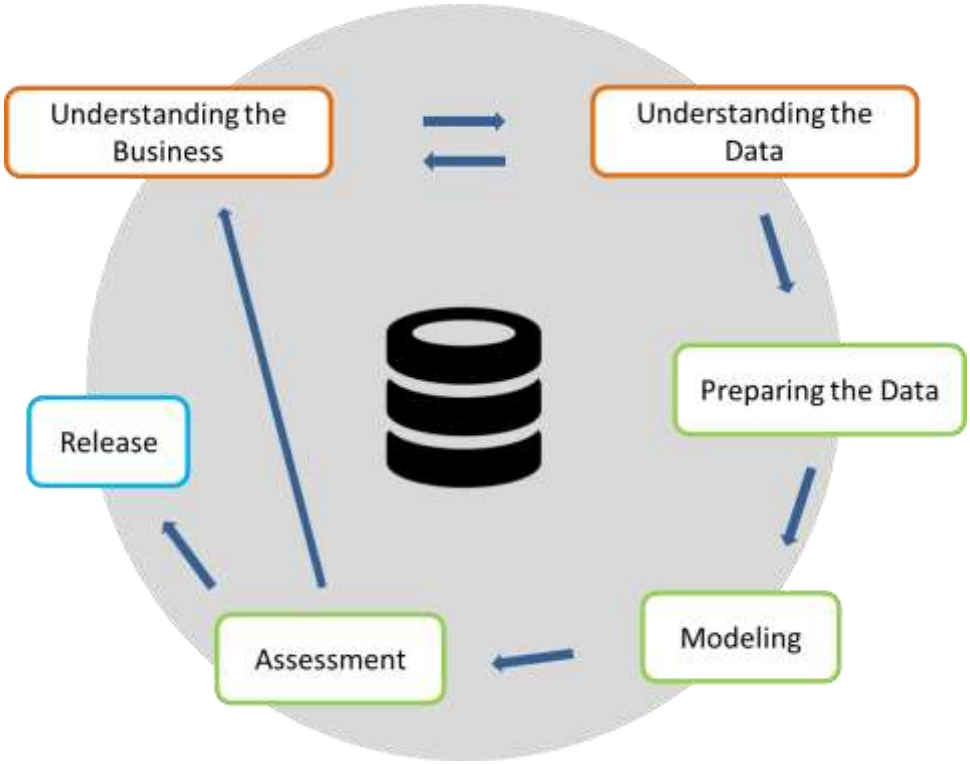

Figure 4. CRISP-DM Process Model (Markkula & Syväniemi, 2015)

Understanding the Business

The purpose of the phase of understanding the business is to define one or few questions or cases which are going to be solved during the process. It is important to define the goals in the early state of the process because it will give understanding of scope and cost, as well as, affect the end result of the project.

Understanding Data

The phase of understanding data consists of user case validation and assessment of forms, contents and quality of the available data.

Preparing Data

The phase of preparing the data is essential when ensuring the validity of it. In this phase the data is modified to fit the shape needed for modeling phase.

The time needed for this phase depends on the amount and structure of the data.

Nowadays the process from preparation all the way to visualization can be generally done with same software and it starts by bringing the data to this software from various sources. There are different types of software for analytics in the market, for example Microsoft's PowerBI, SAP HANA and SAS Enterprise Guide. These are the solutions from largest companies in the industry, and they all work under the same principles. PowerBI is a good tool for analytics because it flexible and can understand multiple data sources. It is easy to use, even for people not that familiar with analytics, Due to its clear and logical user interface. Microsoft is also offering a free license from it, so it doesn't require an investment to implement, at least in the early stages.

The first step is to bring the wanted data to the software from various data sources. For example, in PowerBI data can be brought from various sources like Excel, SQL Server Database, Oracle Data Base, Hadoop, Google Analytics and many other. When bringing the data to PowerBI, the software tries automatically to modify it little to fit the structure needed. It also has an option where the user can modify the data manually before it is downloaded to the software.

After bringing the data to PowerBI, it understands each set as a different table. To do modeling later on, there has to be relationships defined between the tables. To establish a relationship between two tables there should be one column in both tables which has, at least partly, the same data. When data is coming from multiple sources this step might become a challenge Due to it may be hard to find the connecting column between different tables. A quite common problem appearing at this stage is that some parts of the data are missing. This can be caused by different factors, for example the original data is put in the system by human and they forgot something or the measurement device has been changed but the dataset has not been updated. Sometimes a plank space can also mean zero value in other systems.

Other problem that might occur when gathering data from different sources is that the data is not consistent. For example, the same measure might be shown as a percentage value in one table and a numerical value in other, or there might be different types of identification methods from same data.

To solve these kinds of problems the data must be completely understood before so that the right queries can be used in the software. However, even the data were understood properly these remain to be complicated situations, and it might take many time to solve them. That is why it is very important to schedule enough time for this step.

Modeling Data

After the relationships are established the data is generally ready for modeling which means taking information out from the data. In this phase, the modified data is analyzed with statistical and mathematical methods to get results that solve a certain business problem.

The way to get information from the gathered data is usually summarize them. This means taking something out of various tables and calculating something from the data. When doing the summarizing it is essential that the preparation of the data is done properly. Even if it has, there still might be some challenges occurring at this point. For example, if the relationships are not established in the right way also results from the calculations will be false. Therefore, it is important to try different methods and execute the step in an iterative way.

Visualization

One important step is visualization. It is also the only part where the results are being presented. Therefore, it is an important part of the process which should not be under estimated.

When doing data visualization, the key to succeed is to actually not start from visualizations but paying attention to understanding the context for the need to communicate. When the context is properly understood first, it is easier to make the right visualizations.

When starting the analysis one should have understanding between exploratory and explanatory analysis. Exploratory analysis means one's ability to understand the data and picking up the pieces relevant. This method should be used for only discovering the data. The analysis showed to audience should be explanatory which means explaining a specific aspect or story to the audience with the information found from the data.

When making explanatory analysis, there are few things to take in to consideration. These things can be presented as questions who, what and how. Firstly, it is important to know to whom is the analysis for. The presentation needs to be understandable for the audience, and therefore modified according who is listening. The actual decision maker should be identified and the presentation narrowed down based on what she/he can understand.

Next, it should be clear what the presenter wants the audience to know or do. This is the point to ensure the presentation is relevant for the audience. Often the analysis is presented to experts of that certain field and it can make the presenter feel that the audience knows the topic better than the presenter. Usually this is a false assumption because the audience may know the concept but the presenter is the one that went through the data and therefore probably knows it best. Therefore, the presenter should be confident and make recommendations for the audience on how to act based on the results found. In fact, the presenter has a full control of the audience, and therefore he/she can determine what the audience sees and when.

Lastly, how can the data be used to help making the point from it. The role of the data is to be supporting evidence for the story the presenter is telling. However, it is important to understand that showing only data that validates a point the presenter wants to make might be misleading. Instead, there should be right amount of supporting and opposing data to make the presentation reliable and realistic. (Knaflic, 2015)

Assessment

The next phase is assessment which means reviewing if the result met the goal or not. Therefore, the key of this phase is to make sure that all the work done earlier is generating a result to the problem stated in the first phase. All the different techniques and methods of the modeling phase should also be assessed at this phase. If there are some

defects spotted at this phase or the wanted result has changed during the process, the cycle starts from the beginning.

Implementation

In the last step, the process is implemented to the business. This is also an ongoing process. It is fundamental to understand that analytics is not a onetime project but the measures and methods must be maintained during time. This is the way to ensure the reliability and relevancy of the information. (Markkula & Syväniemi, 2015)

4.4 Conceptual Framework

For clarifying how the theory is implemented to this process there is a conceptual framework chart below. On the left side of the chart the theory studied is presented. The implementation of the theory to different steps in the process is presented on the right side of the chart.

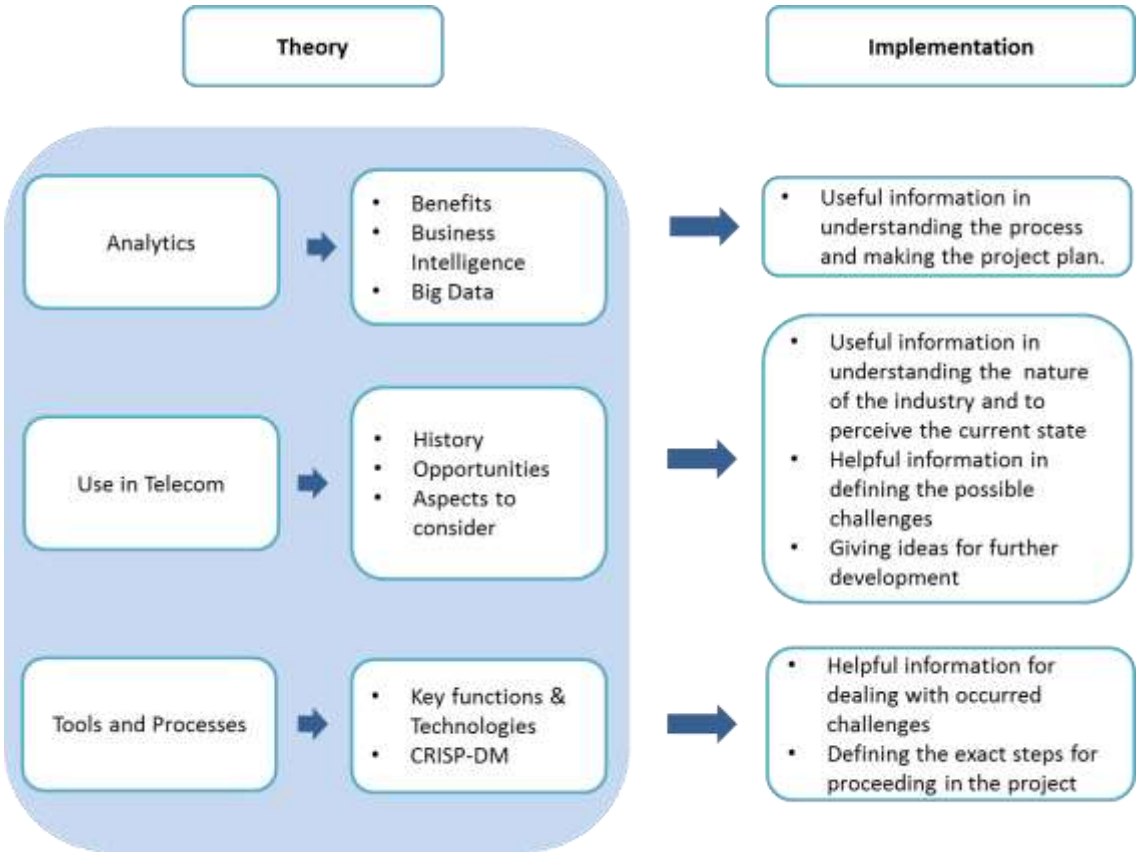

Figure 5. Conceptual Framework

The theory was chosen based on the current state analysis, and the best practices found are useful for both, understanding the process more deeply and so on implementation. It also helped to identify the major challenges on the way and how to deal with them.

5 Analyzing Process & Findings

This project was made based on CRISP-DM methodology. In this chapter each step of the process is described more closely and the findings from the analysis are revealed.

5.1 Process Steps

The first step of the process was to gain understanding from the business case, and the data available. This was done by interviewing professionals from various departments. In the interviews, they pointed out the problems in the network and explained the nature of the data they have. The challenge in this phase was to combine all these different parts of the network and understand the large picture.

The second step, preparing and modeling the data, was the most challenging and time-consuming phase of the process. Data gathered from different systems in the network had many variety and it took time to organize and combine it in the correct way. In the end, there were eight different data sources where the data for the analysis was taken from.

The data was given in Excel-sheet reports from different databases and systems. There were two reasons for using Excel-sheets instead of granting access to all databases and systems. Firstly, waiting for user rights for different systems would have consumed too much time, and therefore ready data sets were noted to be a better option. Secondly, these systems contain various network data, and being able to search a certain part requires further knowledge about using them. Naturally, getting familiar with these systems was not the goal of the thesis, and therefore it was noted to take too much time.

TWAMP and utilization measurements, and error counts were given in separate sheets. Values from TWAMP measures were presented in percentage by every Base Station. TWAMP percentage indicates the response rate between Base Station and Radio Network Controller. It was decided that Base Stations with only less than 100 % response rate were included to the thesis, Due to the ones working flawlessly were not relevant.

Utilization measures were presented in percentage by trails between different devices. Utilization percentage indicates the amount of traffic between two devices. In the analysis, the utilization level between each device of a single route were summed up, and then a total average and maximum value was counted from them. Errors were measured in two different ways, in and out errors. In errors occur when a data packet is entering the device, whilst out errors occur when the packet is exiting the device. In this thesis, only in errors were taken into consideration. This decision was made Due to the routes were known and therefore the out error of one device would be approximately same than the in error of the next device on the route.

As mentioned before the tool used for this process was MS Power BI. It was quickly discovered that this software is mainly created for analyses with simpler, data and in this thesis, it required extra effort to make the given data suitable for it. There was difficulties occurring when using Power BI's own functions, and therefore there was a need to apply programming with M-query language in order to do calculations from the data. Learning to implement the M-query language and finding the most suitable queries to get the wanted results caused delay in the process.

Together preparing and modeling data took approximately 70 % of time and effort put in to this project. On the other hand, this phase gave the most information and understanding about the nature of this industry and the challenges one will face when handling large amounts of various data.

Next step on the process was to make visualizations from the data. As stated earlier in the theory chapter, the analysis should be explanator, which means that it should give the audience relevant information about the thesis. In this case the audience were a group of network experts with extensive experience from the field. Therefore, instead of explaining the logics of the network, the visualizations focused on measurements results and the conclusions made from them.

In conclusion, one might think that the analyzing of data is the most important step of the process. However, it was seen during the thesis that the step that affects the whole process most is the one where the data prepared and modelled. If the preparation has been done poorly, it is impossible to make the right conclusion out of the data in the end of the process.

5.2 Findings

This chapter has removed due to non-disclosure agreement.

6 Recommendations for Improvement

This chapter has removed due to non-disclosure agreement.

7 Conclusion

The purpose of this thesis was to propose recommendations for Elisa on how to improve their network performance by doing an analysis based on their network data.

Elisa has a wide network, and its continual functioning and development is the core of their business. The company saw the need to thesis the relation of Fixed and Mobile Network to gain better understanding of dysfunctions and utilizations of throughout the whole network.

This section measures whether this thesis project met the objectives set for it and the needs of the company. It also evaluates the reliability and the validity of this thesis. Lastly, reflection on the thesis is done to look back upon the work.

7.1 Evaluating Success

The evaluation was made based on how well this thesis project met the needs of the company as well as the objectives stated in the beginning of the thesis. Also, the recommendations made must fulfill the needs of the company and give clear guidelines for further actions.

7.1.1 Company Feedback

The results of the analysis as well as the recommendations were presented to the professionals of Mobile and Fixed Network. After the presentation feedback was given by the experts.

The thesis proved aspects of the network these experts were predicted before and therefore gave them a valid reason to thesis this field more. There was a group of experts gathered, and a project started based on the result of this thesis. In addition, the data

gathered and the models of structuring data were introduced to this group and saved for further investigations.

7.1.2 Achievement of Objectives

The objective of the thesis was to perform an analysis which gives answers to the research questions stated in the beginning of the thesis. On the basis of the results from these questions, recommendations for further improvement were made.

The analysis gave answers to all the asked questions, and there were two recommendations made for further actions. With these actions Elisa can improve their network monitoring and performance.

7.1.3 Reliability and Validity

Due to this thesis was based on data analysis, it is dependent on the reliability of the given data. The data was accurate Due to it was gathered from the company's information systems. The amount of data used for the analysis had to be reduced to one tenth of the total amount due the capacity limitations of the analyzing software. Nonetheless, the reduction of the data was made by deleting every 10^{th} row of the table, and therefore the variety of the data was ensured.

The result of this thesis can be considered as valid for Elisa. The professionals of the field agreed with the results and decided to thesis the area more. The network topology is complex, and there can be other factors affecting the performance which was not considered in this thesis.

7.2 Reflection on thesis work

The subject of this thesis was interesting, yet challenging. When making an analysis, it is important to thoroughly understand the field analyzed. In this case, the field was technical and therefore it took effort to gain a true understanding of it. The experts of the studied fields were involved closely in the thesis to ensure the correct understanding of

the topology of the network. However, it took time to reach an understanding deep enough to execute the analysis.

Therefore, the most important learning points of the thesis were to gain the understanding of the network operations and how fatal small errors and dysfunctions can be. It was very interesting to find out the amount of engineering work and complicated infrastructure is required to make a single phone call or to use internet browser on one's mobile phone. In addition, learning to use an analyzing software was interesting and rewarding.

7.3 Summary

At the beginning of this thesis, the company which this thesis was made for was introduced. The introduction of the company was followed by stating of the business challenge, objectives, outcome and structure of the thesis. In addition, case thesis was introduced as a thesis method used for this thesis.

In the third chapter, the company's current state of analytics in network monitoring was presented. The understanding of the current state was gained by interviewing the field experts about the background of the company and the models of making analysis throughout the different business units. The current state indicated that there have not been analyses covering both Fixed and Mobile Networks made before and this stated the need for this thesis to be executed.

The theory covered in this thesis focused on analytics. Examples on how analytics has been used in telecommunications traditionally and ideas of how it could be used in the future were presented. In addition, different tools and methods of implementing analytics were assessed.

The analysis was made from the data gathered by Elisa and by the best practices learnt in the theory sections. The analyzing model used was CRISP-DM.

Recommendations for further actions were made from the analysis including the answers gained to early stated research questions. From these recommendations, Elisa got ideas how to improve their network monitoring and increase the network efficiency.

References

Booz & Company, Benefiting from Big Data – A new Approach for the Telecom Industry, 2013. Available at: http://www.strategyand.pwc.com/media/file/Strategyand_Benefiting-from-Large-Data_A-New-Approach-for-the-Telecom-Industry.pdf [Accessed 3 May 2016]

Collis Jill & Hussey Roger, Business Research (A practical guide for undergraduate & postgraduate students), PALGRAVE MACMILLAN 2009

Davenport Thomas H, Analytics 3.0, 2013. Available at: https://hbr.org/2013/12/analytics-30 [Accessed 15 May 2016]

Elisa, History, 2016. Available at: http://corporate.elisa.com/on-elisa/history/ [Accessed 25 April 2016]

Elisa, Introduction, 2016. Available at: http://corporate.elisa.com/?_ga=1.32406397.2112600947.1458553922 [Accessed 25 April 2016]

Elisa, Organization, 2016. Available at: http://corporate.elisa.com/on-elisa/organisation/ [Accessed 25 April 2016]

Elisa, Strategy, 2016. Available at: http://corporate.elisa.com/on-elisa/strategy/ [Accessed 25 April 2016]

Harri Pietarinen, Tutkimus: Datan määrä räjähtää käsiin - it-osaaminen ei riitä, 2011. Available at: http://www.itviikko.fi/uutiset/2011/06/29/tutkimus-datan-maara-rajahtaa-kasiin---it-osaaminen-ei-riita/20119157/7 [Accessed 13 May 2016]

HP Enterprise, Transforming Telecommunications with Big Data & Analytics. Available at: https://www.hpematter.com/issue-no-4-spring-2015/transforming-telecommunications-large-data-analytics [Accessed 07 May 2016]

Hill Kelly, Telco Case Thesis: Vodafone and Argyle Data on using Big Data to combat fraud, RCR Wireless News, 2014. Available at: http://www.rcrwireless.com/20141014/large-data-analytics/telco-case-thesis-vodafone-argyle-data-

tag6#prettyPhotohttp://www.rcrwireless.com/20141014/large-data-analytics/telco-case-thesis-vodafone-argyle-data-tag6#prettyPhotohttp://www.rcrwire-less.com/20141014/large-data-analytics/telco-case-thesis-vodafone-argyle-data-tag6#prettyPhoto [Accessed 08 May 2016]

IBM, What is Big Data, 2016. Available at: http://www.ibm.com/large-data/us/en/ [Accessed 07 May 2016]

Kannas Lasse, Eskola Kari, Välimäki Pia, Mustajoki Pertti, Virtaa – Kertauskirja TEK, Atena Kustannus Oy 2013

Masters in Data Science, Data Science in Telecommunications, 2014. Available at: http://www.mastersindatascience.org/industry/telecom/

McCarroll Robert, Bridging the Telecom and Data Divide, March 2014. Available at: http://blog.integratelecom.com/bridging-telecom-data-divide/ [Accessed 07 May 2016]

Nussbaumer Knaflic Cole, Storytelling with Data (A data visualization guide for business professionals), Wiley 2015

Roponen Jukka, Analytics with Big Data and Internet of Things, IBM 2015

SAS, Big Data, 2016. Available at: http://www.sas.com/en_us/insights/large-data/what-is-large-data.html [Accessed 02 May 2016]

Vegi Srinivasa, Analytics may be in place, but there is more to the puzzle, October 2014, read 20.4.2016. Available at: http://www.telecomstech-news.com/news/2014/oct/31/analytics-technology-may-be-place-there-more-puzzle-technology-implementation/ [8 May 20