Tomi Tenhunen

Improving Network Integrity of Finnish Permanent GNSS Network FinnRef

Helsinki Metropolia University of Applied Sciences

Master's Degree

Multimedia Services and Networks

Master's Thesis

8 May 2014

Helsinki Metropolia University of Applied Sciences

Author	Tomi Tenhunen
Title	Improving Network Integrity of Finnish Permanent GNSS Network FinnRef
Number of Pages	104 pages + 1 appendix
Date	8 May 2014
Degree	Master of Engineering
Degree Programme	Multimedia Services and Networks
Instructor	Ville Jääskeläinen, Principal Lecturer

The Finnish permanent GNSS (Global Navigation Satellite System) network FinnRef has been the backbone of the Finnish coordinate system for over twenty years. It has been also used for research purposes in the Finnish Geodetic Institute (FGI), which operates the system. The system consists of 13 stations, which are located around Finland.

FinnRef has been suffering from reliability problems during the recent years due to aging equipment and infrastructure, and now the original system is about to be renewed. The renewal concerns the specialized equipment, as well as the infrastructure including the network. This thesis gives an answer to the research question “What can be done to improve the reliability of Finnish permanent GNSS network FinnRef?”, and the results will be implemented for the use of the FinnRef system.

This thesis approaches the system upgrade from the network analysis view point offered by James D. McCabe. The process is very comprehensive and seizes the details together within the larger picture, which many network analysis processes seem to forget to do. The problems of the system combined with the requirements set by the users, applications, devices and network are reviewed. Through the analysis of those requirements the answer to the research question is found in the form of a set of recommendations. The analysis also produces the requirements list, data flow map and more reliable network.

The thesis describes the analysis and renovation process and the results. The thesis and its results were reviewed by the FGI’s IT manager to ensure the validity.

The whole system along with its problems is reviewed and the transferred data is studied to understand purpose of the system to define the requirements. The reliability is increased due to many improvements. These improvements are for example station power improvements, remote power control and the core network between stations and FGI, which has been changed from VPN (Virtual Private Network) over Internet to more robust and simple MPLS (Multiprotocol Label Switching) network.

This thesis also discusses the importance of the sufficient human resources in the operation and maintenance of the system, together with the well-organized documentation, project leading and operation procedures. All of this has a major impact on the system reliability, since it keeps the system in a constantly evolving state and if the system encounters a critical (or non-critical) issue, there is a clear vision on what should be done to limit and solve the problem with minimum disturbance to the overall system reliability.

Keywords	GPS,GNSS,Network analysis, FinnRef

Contents

Glossary of Acronyms:

3G	Third generation mobile telecommunications technology
ADSL	Asynchronous Digital Subscriber Line
BDS	BeiDou System; Chinese global satellite navigation system
C/A-code	Coarse/Acquisition code
DGNSS	Differential GNSS
DOS	Disk Operating System
EFEC	Earth-Centred, Earth-Fixed
ESP	Encapsulating Security Payload
DGNSS	Differentiated GNSS
DGPS	Differentiated GPS
FGI	Finnish Geodetic Institute; Finnish research institute
FTP	File Transmission Protocol
HDD	Hard Disk Drive
HP	Hewlett-Packard
HRT	Human Response Time
HTTP/1.1	Hypertext Transfer Protocol
GLONASS	Global´naya Navigatsionnaya Sputnikovaya Sistema
GNSS	Global Navigation Satellite System
GNSS-SMART	GNSSS - State Monitoring And Representation Technique
GPS	Global Position System
ICMP	Internet Control Message Protocol
IKE	Internet Key Exchange
INTD	Interaction Delay
I/O	Input/Output
IOPS	I/O Operations Per Second
IP	Internet Protocol
IP-sec	Internet Protocol Security
LAN	Local Area Network
MPLS	Multiprotocol Label Switching
MTBCF	Mean Time Between Critical Failures of network
MTBNCF	Mean Time Between Non-Critical Failures of network
MTTR	Mean-Time-To-Repair
NKG	Nordic Geodetic Commission
NTRIP	Networked Transport of RTCM via Internet Protocol

P-code	Precise-code
PPP	Precise Point Positioning
RINEX	Receiver Independent Exchange format
RMA	Reliability, Maintainability and Availability
RPM	Revolutions Per Minute
RTCM	Radio Technical Commission for Maritime Services
RTCM-104	Real-time GPS data format, also called RTCM
SBAS	Satellite Based Augmentation System
SLA	Service-level agreement
SMS	Short Message Service
SNMP	Simple Network Management protocol
SSD	Solid State Drive
SSH	Secure Shell
TCP	Transmission Control Protocol
UPS	Uninterruptable Power Supply
UTC	Coordinated Universal Time
VPN	Virtual Private Network
WAN	Wide Area Network
WiFi	Wireless Fidelity
WiMAX	Worldwide Interoperability for Microwave Access

1 Introduction

Finnish Geodetic Institute (FGI) has been operating the Finnish permanent GNSS (Global Navigation Satellite System) network FinnRef® for over twenty years now. FinnRef is going through its greatest change, since the equipment and the network infrastructure were getting old and the system needed urgent renovation. The old infrastructure and devices have caused the reliability of the network to decrease year by year. The two greatest challenges are the network and the GPS (Global Position System) receivers, since the original Ashtech receivers can store only up to three days data. When the memory is fills up, the receiver stops storing the data and thus loses all data came after the memory came full. So in case of a network outage, there is always a big chance to lose valuable research data. The data connections are also hard to maintain due to the remoteness of most of the stations.

This thesis aims at giving an answer to the research question “What can be done to improve reliability of FinnRef network?”. The answer comes in the form of a set of recommendations and updated network, and those will be achieved through the collection of information from the initial network, and gathering of the requirements for the renovated network, and then analysing the data. The analysis is done by using a Network Analysis, Architecture and Design approach, which the recognized network professional James D. McCabe has developed.

The thesis is mainly concentrated on the analysis part, but also lightly covers the architecture and design parts. The renovation of the system and the network was already started in 2012 and thus most of the improvements have been already implemented. The renovation also included the constructing of new stations. The renovations have been following the recommendations and ideas made on the research done for this thesis, and they also have been implemented as the renovation process has needed them to be implemented.

As mentioned, many improvements were already installed as the analysis process stage went forward. The two greatest improvements were the new GNSS receivers and the new core network solution. All stations, including the six new stations, have the new receiver already, and the new core network is installed on all the new stations and some of the original stations, but it will be eventually installed to all of the stations.

Since Sonera was FGI's current operator, the chosen core network was Sonera's product DataNet, which uses the MPLS (Multiprotocol Label Switching) technique. It should increase reliability, as MPLS does not need as much computing power and management as previously used IPsec VPN (Virtual Private Network) does. The MPLS core network will be installed over various connection types; 3G mobile network, ADSL (Asymmetric Digital Subscriber Line) and fiber optics. Also other important improvements were installed to the stations; better lightning protection and remote electricity control. These actions along with the new network and with the improvements in the set of recommendations of this thesis will make the data gaps shorter or even prevent them entirely.

1.1 Finnish Geodetic Institute

The Finnish Geodetic Institute, which was established in 1918, is an expert and research institute of spatial data infrastructure governed by the Ministry of Forestry and Agriculture. Finnish Geodetic Institute provides a scientific basis for Finnish maps and geospatial information, carries out research and development on methods for the measurements, data acquisition, processing and exploitation of geospatial information, provides knowledge and value-added information for public-sector bodies and other providers and users of spatial information, and co-operates with other governmental organizations, industry and universities. FGI has five departments: Geodesy and Geodynamics, Geoinformatics and Cartography, Remote Sensing and Photogrammetry, Navigation and Positioning, and Administration Services. The Finnish Geodetic Institute employs of 84 people. (Finnish Geodetic Institute, 2012: 3)

1.2 FinnRef

Finnish permanent GNSS (Global Navigation Satellite System) network (FinnRef), formerly known as FinnNet, is a part of the Nordic Permanent GPS (Global Position System) Network. The Nordic Permanent GPS Network was established in 1993 by the Nordic Geodetic Commission (NKG), as response to the initiative of the directors of the Nordic Mapping agencies. (Koivula, 2006: 1-2)

Finnish Geodetic Institute made a decision to build 12 permanent GPS stations in 1992 and the first stations were built in 1993. The rest of the stations were built between

1994 and 1996, except one extra station was added to the network in 2005, so the station count at the initial point of this thesis was 13. One of these FinnRef stations is a part of the International GNSS Service (IGS) network and four of the stations, which also include that mentioned IGS station, are a part of the European Euref Permanent Network EPN. (Koivula, 2006: 4 & 15-17)

FinnRef is used for creating a backbone for the Finnish coordinate system EUREF-FIN, and to provide a connection to international and older national coordinate systems. FinnRef is also used to study the postglacial rebound in Finland and Fennoscandian region. (Koivula, 2006: 37-46)

The FinnRef system was renewed during the years 2012 and 2013. The initial renewal plan can be seen in Figure 1, where the original stations and the possible new stations are shown.

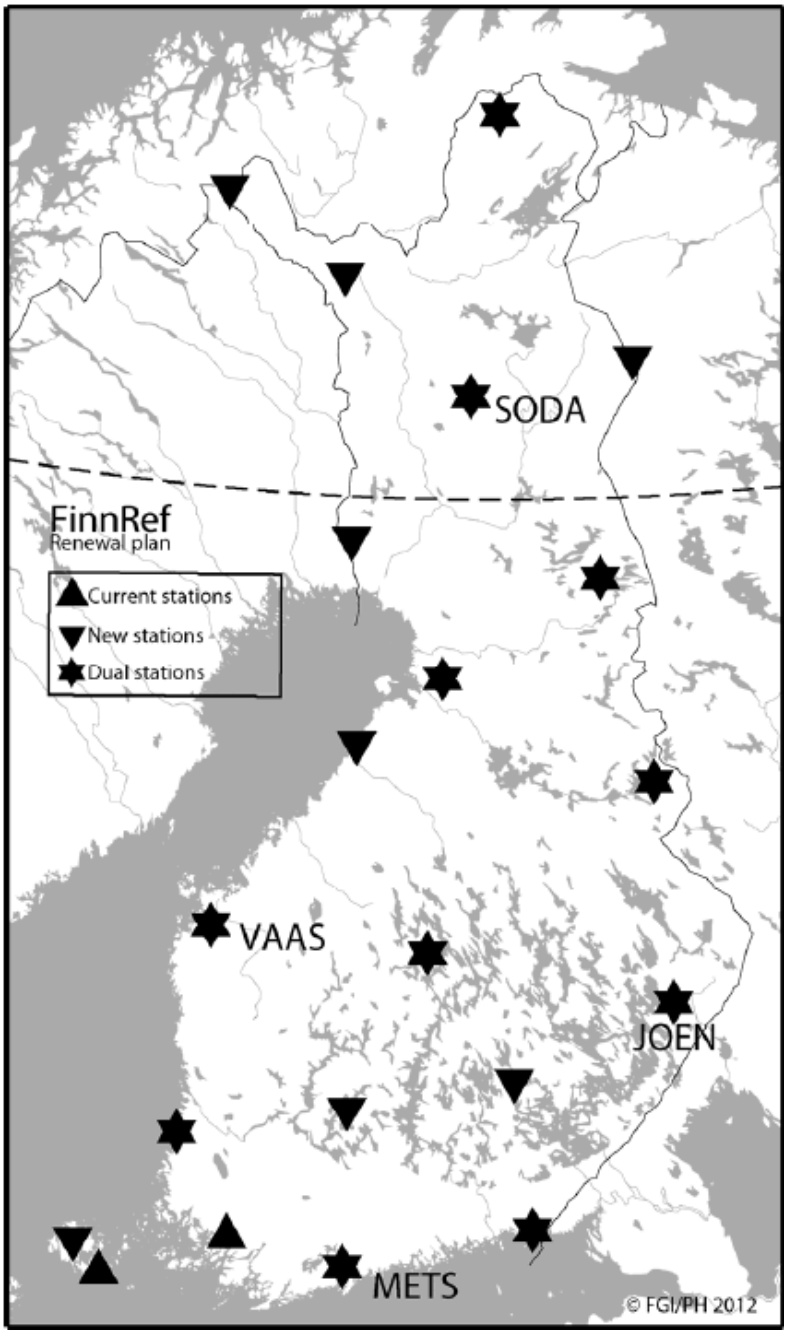

Figure 1. The renewal plan map of FinnRef in 2012. (Koivula & al., 2012)

The renewal included the whole system; new stations were built, new core system equipment was bought and installed, and many reliability improvements were made. (Koivula & al., 2012)

1.3 Research Method

The answer to the research question “What can be done to improve reliability of FinnRef network” is achieved through the study of the theory books concerning network analysing and designing, and then applying those methods to the FinnRef network.

The network analysis part of the thesis produces requirements and data flow information, which are based on the discussions with the network users and managers along with the system documentation and measurements. Also a user questionnaire results along with internal documentation and publications are used to obtain information about the system.

The analysis process results leads to the renovation process, which is done along with the FinnRef system renovation. The renovation process applies the results of the analysis process to the network. The renovation process results along with the network analysis results are used for deriving the set of recommendations to be applied later.

The network administrator reviews and validates the thesis to provide the results are consistent with the assignment. The overall process will produce four different outcomes; an answer to the research question, a renovated network and the set of recommendations. The process flow is illustrated in Figure 2.

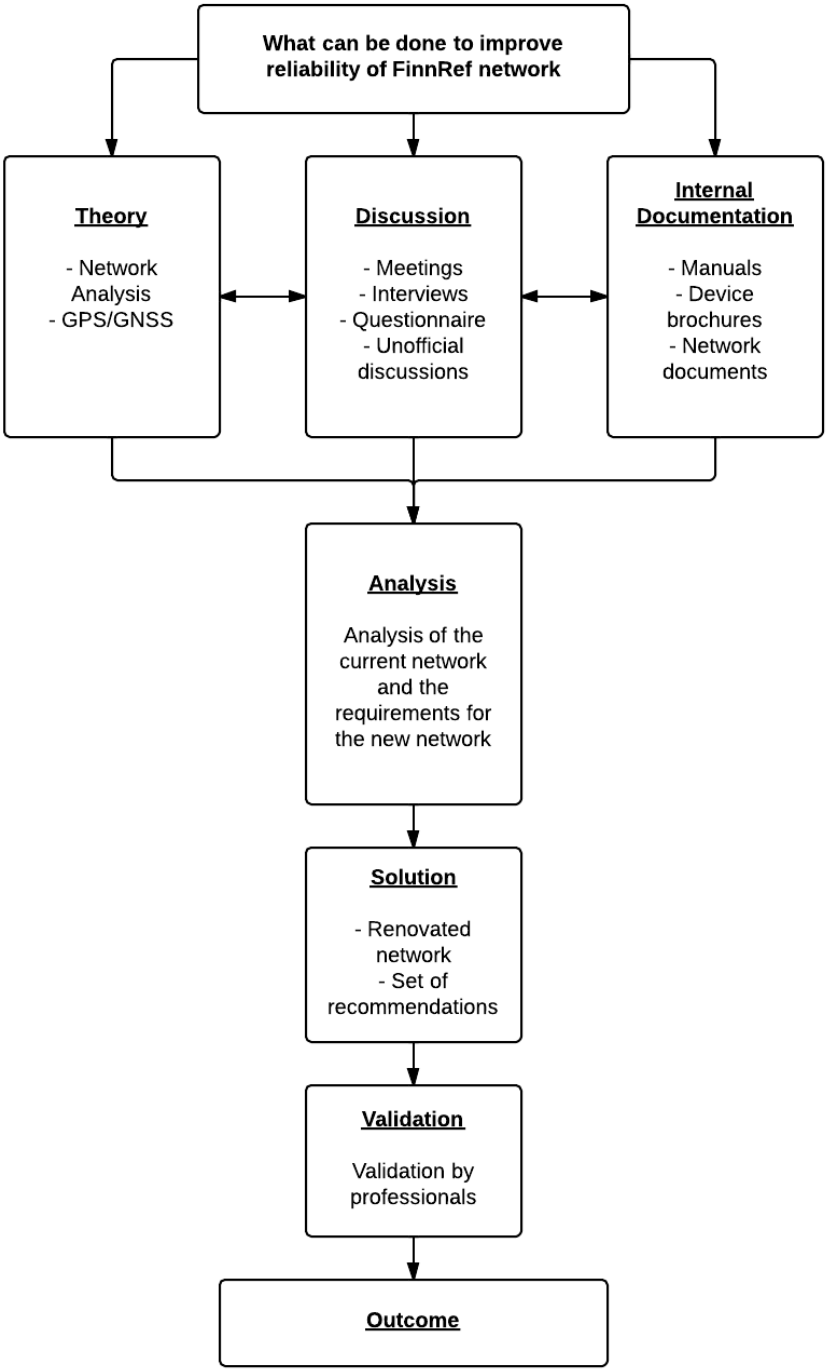

Figure 2. The research design and process flow of the thesis

The method of analysis is to gather information about the existing FinnRef network and gather the requirements for the new FinnRef network and then derivate from that data a set of recommendations for the new network. Some of the results of analysis were applied to network when the GNSS network was renovated.

1.4 Material

A single method and the material for the network analysis part were chosen after the comparison of four different network analysis theory books. The compared sources were; Ye's Secure Computer and Network Systems: Modeling, Analysis and Design (2008), Liotine's Mission-Critical Network Planning (2003), Lucas' Network Flow Analysis (2009) and McCabe's Network Analysis, Design and Architecture (2007).

Each of the sources has a slightly different approach to network analysis; Ye have concentrated on the security issues, Liotine on ensuring the reliability of the mission-critical services and Lucas purely on analysing and optimising the network data flows. McCabe has covered all of those aspects in his book quite comprehensively and he has also created a very proficient process in his book which carries on throughout from the start of the project until the end of it.

2 Principles of Satellite Navigation and Position Systems

To fully understand how the FinnRef network works and what its purpose is, the understanding on how Global Position System (GPS) and other global position systems works is essential. In this chapter GPS and other Global Navigation Satellite Systems (GNSS) are discussed and the idea on how the positioning happens in practice. Also the way how the GNSS data is handled and how it is formatted is important to know, since it is the product of the FinnRef network.

2.1 GPS and Other GNSS Systems

GPS was originally developed in 1973 by the United States for military use in order to determine instantaneous position, velocity and precise time of the military troops anywhere on Earth, although it was in 1983 declared to be open for civilian usage. All of the 24 satellites were operational in 1993 and the system was declared fully operational in 1995. (Hofmann-Wellenhof, Lichtenegger & Wasle, 2008: 309-310)

Russian GNSS system GLONASS (Global´naya Navigatsionnaya Sputnikovaya Sistema) was declared operational in 1993, although all of its 24 satellites were not operational until 1996. GLONASS was originally designed for military use only, but it was also declared open for civilian use in 1996. (Hofmann-Wellenhof, Lichtenegger & Wasle, 2008: 341-342)

European Union started their program to achieve their own independent open GNSS system in 1999. The system was named Galileo. The system should be in a fully operational stage in 2020, although the first position fix using only the Galileo was acquired on March 2013 and it will start to offer early services at the end of 2014 (Hofmann-Wellenhof, Lichtenegger & Wasle, 2008: 365-366; ESA, 2013)

China also had started to build their open GNSS system back in 1994, the system has changed its name a few times, from BeiDou to COMPASS and back to BeiDou System (BDS). At the end of the year 2012, BDS had achieved operational status at regional coverage with 16 satellites. By the end of 2020 the system should be fully operational with global coverage. (China Satellite Navigation Office, 2013)

2.2 Satellite Navigation and Position

The principle of the satellite-based positioning in systems like the Global Position System (GPS), is based on the idea seen in Figure 3, where on the given moment the distance (range) between the receiver and the visible satellites (vector ϱ) can be accurately measured by using the time what the satellite's coded signal takes to arrive to the receiver and the vectors ϱ^s and ϱ_r. Vector ϱ^s is the satellite's relative distance to the earth centre, which can be calculated from the satellite ephemerides (orbit) information broadcasted by the satellite and vector ϱ_r is the receiver location related to the earth centre. Using three satellites range, each of them forming a sphere around their location related on the ground, the location of the receiver can be narrowed to intersection of these three spheres. This is true when the receiver is equipped with an ideal clock, which is set precisely to the system time. (Hofmann-Wellenhof, Lichtenegger & Wasle, 2008: 3-4)

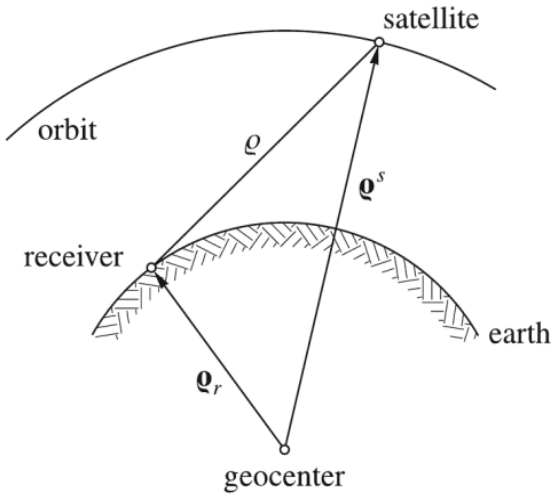

Figure 3. The principle behind the satellite-based positioning (Hofmann-Wellenhof, Lichtenegger & Wasle, 2008: 3)

But in practice, the receiver's clock has a small time offset, because of its inexpensive crystal clock, which tend to drift. (Poutanen, 1998) And because of that bias (and some other biases like atmospheric delay), the range measurement is not precise. The range together with the error caused by the biases is called pseudorange. (Gurtner and Estey, 2009/2012: 4) To resolve the clock bias, a fourth satellite is needed to get the position solution. (Hofmann-Wellenhof, Lichtenegger & Wasle, 2008: 3-4)

2.3 Positioning with GPS Signals

The GPS satellites are transmitting two carrier wave signals (L1 and L2) in two different frequencies. Carrier waves have two different kinds of pseudorandom signals modulated in them; C/A-code (Coarse/Acquisition code) and P-code (Precise code). The C/A-code is carried only on the L1 frequency, but the P-code is carried by both of the frequencies L1 and L2. Both of those frequencies have also the satellite ephemeris (orbital) data modulated in them. The P-code is encrypted and it is available only for the United States (US) and allied militaries along with US government. (Poutanen, 1998: 9; Küpper, 2005: 166-167)

The calculation of a pseudorange is done by using the C/A-code to calculate the code phase range (code-pseudorange) or by using the carrier wave to calculate the carrier phase range (phase-pseudorange). In practice, the code-pseudorange is obtained from the signal travel time calculation, and the phase-pseudorange is obtained by calculating how many full wavelengths there is between satellite and receiver. Also the Doppler-shift of a carrier wave signal frequency can be used to obtain the range between the satellite and the receiver. (Poutanen, 1998: 121-123)

Even though the P-code is encrypted, it still can be used in a position calculation, together with the C/A-code or alone. One possible way is that the receiver produces a copy of the P-code and correlates it with the received signal. It takes advantage from the fact that the encryption lasts longer than the signal. Although the P-code is usable in that way, it has worse signal to noise ratio than a real decoded P-code. (Poutanen, 1998: 155)

2.4 GPS Data Processing

To get more accurate position with the GPS, it is possible to collect data with multiple GPS receivers and process the collected data afterwards to remove the errors (biases). There are various ways to process the static GPS data, but only a few are used these days; precise point positioning (PPP), which is based on the use of the precise navigation satellite information obtained from the Internet, and the more traditional and most commonly used processing method, which is differentiating, which can be used with various methods and algorithms. (Chassagne, 2012; Koivula, 2006: 19)

Differentiating can be also done on real-time services. Such services are called DGPS (Differential GPS) or DGNSS (Differential GNSS). These systems are based on the use of a reference station or a network of them (wide area DGPS), which is/are sending its/their location to server which uses that reference data to remove the biases from the GPS data and the broadcasts it again to be obtained by selected users. This data is usually very accurate, but it is dependent on the user's distance from nearest reference station. (NATO, 2008; FGI, 2014)

In GPS data post-processing, the software commonly uses the following observables collected by the receivers; carrier-phase measurement (at one or both carrier frequencies (L1, L2)), the code pseudorange measurement and the observation time of receiver. (Gurtner and Estey, 2009: 2)

The carrier-phase measurement measures the range between the satellite and the receiver by the means of cycles in a satellite signal's carrier frequency. In practice it is the measurement of phase difference between the receiver generated reference frequency and the phase of received satellite signal carrier frequency. (O'Driscoll and Petovello, 2010; Koivula, 2006: 19)

The pseudorange (or code measurement) is measuring the distance to a satellite, by differentiating the receiver's time of receiving the signal to the satellite's time of sending the signal. The observation time is the receiver's exact time of observation of valid pseudorange or carrier-phase measurement. (Gurtner and Estey, 2009: 2)

2.5 RINEX (Receiver Independent Exchange) Format

The raw binary files generated by the GPS receivers differ depending on the receiver manufacturer, thus the usage and distribution of the data was difficult before RINEX was developed. Driven by that fact, together with the big data amounts generated by the large European GPS campaign EUREF 89, the Astronomical Institute of the University of Berne developed Receiver Independent Exchange Format RINEX. The first version was approved in 1989, and the version 2.00 a year later (Gurtner and Estey, 2009: 2-3). RINEX 2 and its subversions have been used as a de facto standard for more than twenty years (Hatanaka, 2008: 1-2). The current subversions which are used for GNSS data exchange by the major GNSS networks like the International GNSS Service (IGS) and European Permanent Network (EPN) is RINEX 2.10 and

2.11. More recent RINEX versions (2.12, 3.00 and 3.01) are also collected, but only for testing purposes. (ROB, 2012). Since part of FinnRef network belongs to IGS and/or EPN networks, it produces RINEX 2.10, but the renovated system produces also RINEX 3.01 (Koivula et al, 2012: 4).

As defined in The RINEX 3.01 specification (Gurtner and Estey, 2009: 3), the format consists of three ASCII file types:

1. Observation data File
2. Navigation message File
3. Meteorological data File

In the older versions, like RINEX 2.11, there were more file types (Gurtner and Estey, 2007/2012), but those three are the most commonly used ones (Hofmann-Wellenhof, Lichtenegger & Wasle, 2008: 193). All data files have a header at start the of the file, which includes file specific information, for example, time (in Coordinated Universal Time (UTC) format), date, used software and the version of RINEX used (Gurtner and Estey, 2009: 3). From those three, the observation data, as the name indicates, includes the necessary GNSS observation data; carrier phases, code ranges, Doppler measurements and the signal to noise ratios (Hofmann-Wellenhof, Lichtenegger & Wasle, 2008: 490). The observation data file can include observations from more than one receiver or antenna, but that is not recommended (Gurtner and Estey, 2009: 3).

The Navigation message file contains the ephemerides and almanac data, satellite number and the satellite's health status (Hofmann-Wellenhof, Lichtenegger & Wasle, 2008: 49-51 and 449). The Meteorological data file includes the weather data, which usually contains the temperature, humidity and the air pressure, recorded by an external weather station connected to the receiver. The meteorological data is used to correct the atmospheric error in the satellite signals (Paros and Yilmaz, 2002: 1).

As noted, the observations collected by receiver are the following; the carrier-phase measurement, the pseudorange (code) measurement and the observation time of receiver. Those observations are also in the used data processing. The processing software usually needs also the station name, antenna height and other station specific information along with the observation data. The naming convention of the files is also described in the RINEX format. (Gurtner and Estey, 2009: 2, 5)

2.6 RTCM and NTRIP Formats

RTCM (officially RTCM-104 or RTCM SC 104) is a real-time GPS data format, which was developed as an answer to the demands of industry standard for a real-time differential GPS correction messages. The standard was developed by the RTCM (Radio Technical Commission for Maritime Services) organization's special committee 104, from which the official name of the format is derived from. The RTCM format version 2.x is based on the structure of GPS navigation message format, and thus is a binary format.

Currently the most commonly used version is RTCM format version 2.3, which carries the data in several different message types, from which the most important types are the following; type 1 is carrying the differential GPS corrections (pseudorange and velocity, with a maximum of 12 satellites), type 2 carries delta-differential GPS corrections, or in other words, the pseudo-range corrections, referring to previous orbit data records (maximum 12 satellites), type 3 the reference station Earth-Centred, Earth-Fixed (EFEC) coordinates in X,Y and Z directions. Types 18 and 19 are for the uncorrected raw measurements of the pseudorange and differential, and types 20 and 21 are the same measurements with corrections. Types from 22 to 24 are carrying the reference station information and its used antenna type and serial number. (Heo et al., 2009: 4-12)

NTRIP (Networked Transport of RTCM via Internet Protocol) is a stateless protocol used to transfer GPS/GNSS differential correction data (like RTCM) over the Internet. It is based on the HTTP/1.1 (Hypertext Transfer Protocol), with three types of applications; NtripClients, NtripServers and NtripCasters. The NtripCaster is used as a HTTP server application, when the NtripClient and the NtripServer are used as a HTTP clients. NTRIP is capable of distributing hundreds of streams to thousands of users without the problems caused by firewalls or proxies. (Germany. Federal Agency for Cartography and Geodesy (BKG), 2013) The NTRIP version 1.0 was accepted by the RTCM Committee as a standard for a packet-based communications. The version 2.0 will have a full HTTP compatibility and a possibility to use the User Datagram Protocol and IP (UDP/IP) in network connections. (Heo et al., 2009: 3)

3 Network Analysis

As computer networks are becoming more and more complex and the used applications are more dependent on reliability and delay than the capacity of the network, and many of the currently used network designs are coming to the end of their road, so new approaches have to be considered. The Network Analysis, Architecture and Design –approach, as McCabe (2007) defines it, is very flexible, informative and executable on almost all networks. The approach is straightforward; each step (or process) produces all needed information for the next step, as the Figure 4 shows. (McCabe, 2007: 3-10)

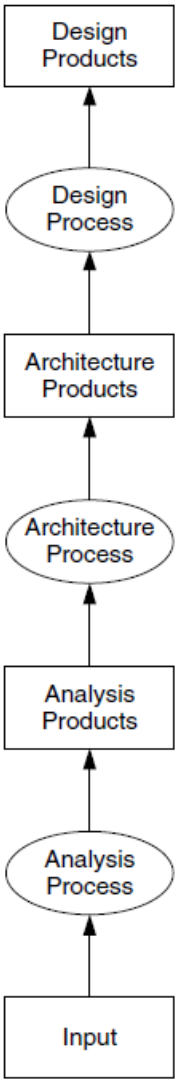

Figure 4. The process flow of the Network Analysis, Architecture and Design –approach (McCabe, 2007: 10)

The present study uses the forementioned approach to point out the problems and find the best possible solution for the FinnRef's network problems. In this chapter the Network Analysis, Architecture and Design –approach is discussed and the basic idea of it is introduced.

3.1 Concept of Network Analysis, Architecture and Design

Improvements to FinnRef's network reliability follow the Network Analysis, Architecture and Design –approach, concentrating on the network analysis. The concepts of network architecture and network design are also discussed to get a clear vision on why the analysis is done.

The network analysis process is done to achieve a better understanding of the current network, its usage and the system itself. With sufficient understanding of the system, it is easier to detect the problems and see what there must be done to achieve a better working and more manageable network.

As Figure 5 illustrates, the inputs for the network analysis process are the state and the problems of the existing network together with the requirements of the users, devices and applications. When input material has been processed through the network analysis process, the outputs will be the descriptions about the requirements and problems of the network, description about the mapping of applications and devices of the network and the descriptions of traffic flows and the potential risks. (McCabe, 2007: 6-7)

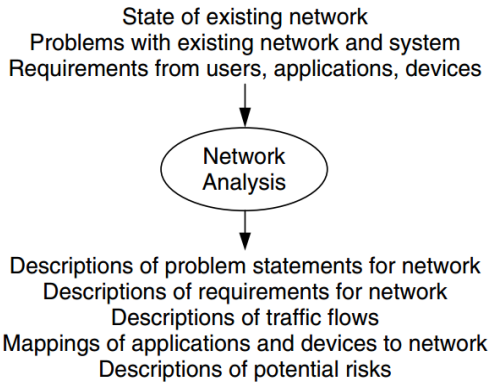

Figure 5. Inputs and outputs of network analysis process (McCabe, 2007: 7)

When the network analysis process is finished, it is time for the network architecture process (seen in Figure 6). It is about creation of the architectural concept model from the output of network analysis process. The model should define the network structure from end to end on a high level. During the network architecture process, the network topology and technology, together with the equipment classes are chosen. The relationships between the network functions, such as performance, addressing, security and management are determined. Also the optimization between the network functions is considered. (McCabe, 2007: 7-8)

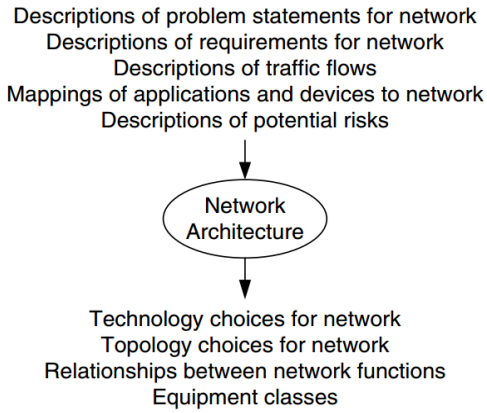

Figure 6. Inputs and outputs of network architecture process (McCabe, 2007: 8)

After the network architecture process, the next step is the network design process (as seen in Figure 7), which will generate the final plans and the specific information about the devices, vendors and service providers from the output of the network architecture process.

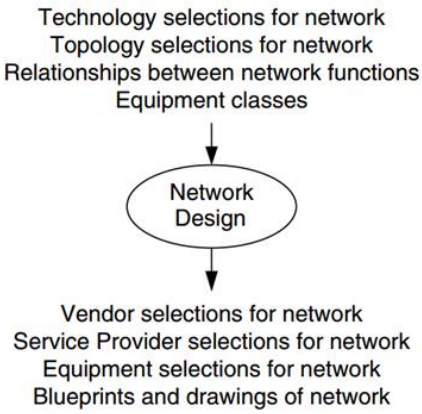

Figure 7. Inputs and outputs of network design process (McCabe, 2007: 8)

During the network design process, the design goals are set and the proposed designs are evaluated against them. Also the weighing between features like cost versus performance is done here. (McCabe, 2007: 8-9)

3.2 Network Analysis Fundamentals

The network analysis process output is used as a base for the network architecture and design, so making it is the most important phase in the overall process. The analysis is based on gathering and analysing the network requirements, or in other words, the requests of features, performance and functions generated by various system elements which are necessary for the network to operate as it should. These elements are the users, applications and devices, from which the requirements are gathered and derived. The requirements form up a base for the network analysis and the process as a whole, since they define customer expectations and satisfaction. (McCabe, 2007: 18)

The analysis of the requirements will grant better understanding of the network. The outputs are; performance requirements, definition of low and high performance applications, and identification of the services used in the network. (McCabe, 2007: 61-62)

3.2.1 Definition of System and Its Requirements

When defining the properties and problems of the network, a convenient way is to think that all of the individual components in the network are a part of a system; including the network itself with its users, applications and devices. And as a part of the system, the network offers a service to the rest of the system.

The network service can be defined in two ways; as levels of performance and functions in the network or as sets of requirements set by the users, devices, applications or other system components. The levels of performance can be broken down in three different main categories; capacity, delay and RMA (reliability, maintainability and availability), while the functions can be divided to a wider set of categories, including security, accounting and management among others. With these categories, or characteristics, along with the system requirements, the network service to the system can be defined in detail. (McCabe, 2007: 27-33)

The system requirements are defined gradually by the system components. The definition process starts from the user requirements, which tend to be subjective and at a general level. The application requirements expand and refine the user requirements. The device requirements in turn refine and expand the user requirements, and the device requirements does the same to the network requirements. As the process outputs the requirements to each system component, it also produces the network element requirements, which defines what kind of devices and settings the network needs as a final product.

The system requirements also help to define the service metrics, which are used to achieve measurable and comparable values from the system. The service metrics can then be used to define the service characteristics; delay, capacity, RMA and security for example. The service characteristics are used to describe the service levels of the network. The service metrics are also used in the engineering and optimization of the network, and they also serve as a tool when monitoring and verifying that the network is working the way it was designed to.

System functions along with the other system characteristics should also be counted, since the functions may also have requirements. Such functions are, for example, network monitoring and management, security, and accounting. Because all of the characteristics are affecting each other, the designing of the network architecture must be done with thought to avoid creating bottlenecks in the network or causing poor manageability. (McCabe, 2007: 33-37)

3.2.2 Service Requirements

The service requirements are defined by the service requests, which a user, application or device has generated. They are categorized by their predictability in three categories; best-effort, guaranteed and predictable.

The best-effort service is unpredictable and unreliable, which means that the other components of the system must adapt the state of the network at any given moment. Such service request either does not have specific requirements for the performance of the network or is based on the estimate of capacity.

The guaranteed service is an opposite of the best-effort service; the service is predictable and reliable. In guaranteed service, the service provider and user usually sign a contract, where the requirements and limits for the network operation are defined. If service is not offered in those limits the service provider may be obligated to pay compensation to the user.

Between the best-effort and guaranteed service falls the predictable services, which are offering some level of predictability, but are not guaranteed or accounted. The measurability and verifiability along with configurability are crucial when offering predictable or guaranteed services, since user's request and service provider offering must be consistent and they are based on the same set of requirements.

The set of requirements is usually defined with the help of the performance metrics to set the thresholds and limits to the service. Thresholds are used to set warning boundaries and the limits are used to set underachievement boundaries for the performance characteristics. These boundaries are used for monitoring service quality and for helping with traffic management and control. The thresholds can also define the high and low performance boundaries, which can be used for monitoring that the agreed service level is achieved. (McCabe, 2007: 38-45)

3.2.3 Performance Characteristics

As mentioned in Chapter 3.2.1, the network service performance can be described and measured with three characteristics; capacity, delay and RMA. These characteristics are used in the definition and monitoring of the network service.

Capacity measures how well the system can transfer the information from one point to another in the system. The capacity is expressed in amount of data per second, for example bits per second. Each user, device and application associated with the network takes a cut from the total capacity. Basically, the more system components there are sharing the total capacity, the less capacity there will be for each component.

Delay measures the time, which the transmission of the information takes to travel through one of the system's points to another. Delay can also measure how long it takes for an application or device to complete its process. The third thing delay can measure is jitter, which is describing the delay variation, which is important to keep at

minimum in real-time applications. Delay can be measured in two ways, only in one direction, which is called end-to-end time, or both directions, which is called roundtrip time. Delay is expressed in seconds. Delay is a good describer of the network behaviour, since it can measure the performance of the system components.

RMA, referring to reliability, maintainability and availability, is a characteristic which gives information about the availability of the service. Reliability is based on the statistics, and denotes how frequently the critical failures occur in the network and its components. Reliability also represents the confidence how the users believe the system and network will fulfil their requirements. Maintainability is also based on the statistical information, and it indicates the time, that it takes to restore the operation on to an acceptable level after a critical failure. This amount of time is commonly referred as a mean-time-to-repair (MTTR). The repair process for the failure is straightforward, starting from the detection of failure, moving on to isolation to find the replaceable component; then the replacement component will be transported to the location of the failed component and then finally the installing and testing of the replaced component, which leads to full restoration of the services. Availability, as mentioned, is the core of the RMA, since it describes the relationship between reliability and maintainability. The following equation clarifies that relationship:

$$A = \frac{MTBCF}{MTTR+MTBCF} \text{ or } A = \frac{MTBNCF}{MTTR+MTBNCF}$$

As the equation shows, the calculation of availability (A) is done by dividing the mean time between critical failures (MTBCF) or non-critical failures (MTBNCF) with the sum of mean-time-to-repair (MTTR) and the mean time between critical failures (MTBCF) (or non-critical failures (MTBNCF)). (McCabe, 2007: 45-50)

3.3 Network Requirements Fundamentals

The network analysis process begins with gathering and analysing the requirements. The requirements are gathered from every component in the system; users, applications, devices and also from the network. The user requirements are the most subjective and the least technical, and when moving towards the network level, the relationship of technicality and subjectivity changes so, that the network requirements are most technical and least subjective. The goal is to create requirements which are as objec-

tive, quantitative and technical as possible. The requirements should also be categorized and prioritized. (McCabe, 2007: 59-60)

3.3.1 User Requirements

The first requirements, which are gathered, are the user requirements. The gathering starts with the discussions with the different kinds of users of the system, network personnel and management. Figure 8 shows different kinds of user requirements, but the following general requirements are the most significant:

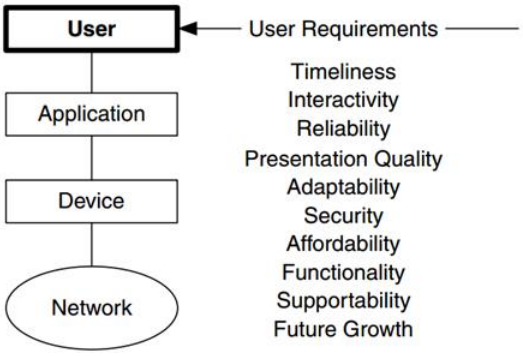

Figure 8- Types of user requirements (McCabe, 2007: 63)

Timeliness, which describes the amount of time, which the user tolerates waiting to be able to access, transfer or modify information. Next one is interactivity, which measures the system response time in interactive applications. Both can be measured by using the round trip delay measurement.

Reliability, in this case, is the availability of the network service from the view of its users. The users' view of availability is wider than the RMA defines; it is defined by all of the performance characteristics, including capacity and delay. Security, from the users' viewpoint is concerning the security of the user's data and personal information and secure access to them. It can primarily be measured with reliability, but also delay and capacity has impact to it.

Affordability is a requirement, which describes how much the users and management can afford to do purchases for the network without exceeding budget.

Supportability covers what kind of support the users want or need from the network staff, and their ability to affect the network configuration after its implementation. Supportability also covers what kinds of support applications the network staff will have to provide for the users, and what applications to use for identifying and troubleshooting the network problems. Future Growth describes if and when the users will be expecting new applications and/or devices to be added to the network.

And in addition to those requirements defined by the users, the user requirements include also the amount of users expected to be on the network with their locations. Also the future growth in the user amounts should be estimated, at least for the next three years. (McCabe, 2007: 61-66)

3.3.2 Application Requirements

Application Requirements are more technical than the user requirements, but might still be subjective, since it is also based on the users experience from the use of those applications along with the other information about the applications.

As can be seen in Figure 9, the application component, and thus the application requirements, is placed between the user component and the device component in the system. This bi-directional placement causes it to be the point where many of the requirements are generated, since the users are using the applications, and applications are using the devices and the network beneath it.

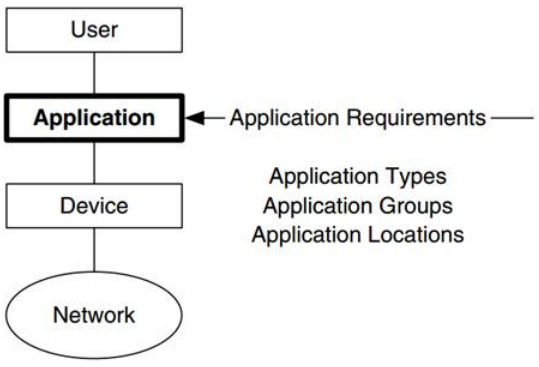

Figure 9. Types of application requirements (McCabe, 2007: 67)

Since the user requirements of timeliness, interactivity, reliability and security are affecting on the application requirements, we can divide the applications to ones that

need predictable or guaranteed service and to ones that are settled for the best-effort service. From that, it is possible to categorize the predictable or guaranteed service applications to three types based on the service and performance requirements; mission-critical, rate-critical and real-time/interactive. The application types are defined by their requirements and service metrics.

Mission-critical applications require predictable, guaranteed and/or high–performance RMA. Rate-critical applications require predictable, guaranteed and/or high–performance capacity. Rate-critical applications require the thresholds and limits for the guarantee of minimum, peak and/or sustained capacity.

Real-time or interactive applications require predictable, guaranteed and/or high–performance delay. Often some applications are wrongly referred as real-time applications, so it is necessary to add more categories to describe them; real-time and non-real-time, which can be furthermore divided to the asynchronous and interactive applications.

True real-time applications have synchronous timing between the source and destination, with time boundaries set by timers. Timers keep the source and destination synchronous by dropping the information coming outside the time window. So in the real-time applications the information carried in the network adapts the time marginal, not vice versa. Real-time applications have a strong impact on the network architecture.

Most applications are non-real-time, where the end-to-end delay time requirements vary and the destination will wait a reasonable time for receiving of the information. The wait timers are set by the applications, devices and/or protocols used. Therefore the delay time adapts to the network conditions. Non-real-time applications include the asynchronous and the interactive applications, so they cover the majority of the applications. Asynchronous applications are in the opposite end of the performance requirements compared to the real-time applications. Asynchronous applications are not dependent on strict timing, or at least the timing is so loose, that it is outside the application’s session. Interactive applications are expecting some timing between the source and destination when the application is active, but the timing does not have to be strictly synchronous as in the real-time applications.

It is also practical to categorize the applications by their locations, along with the application types and groups. That way the traffic flows and requirements can be determined more precisely, when the locations are grouped by building, floor, user and/or user group. (McCabe, 2007: 66-76)

3.3.3 Device Requirements

As mentioned above, the device requirements are based on the user and application requirements. The device requirements build up from device types, performance characteristics and device locations, as can be seen in Figure 10. The devices can be grouped into three major categories; generic computing devices, servers and specialized devices.

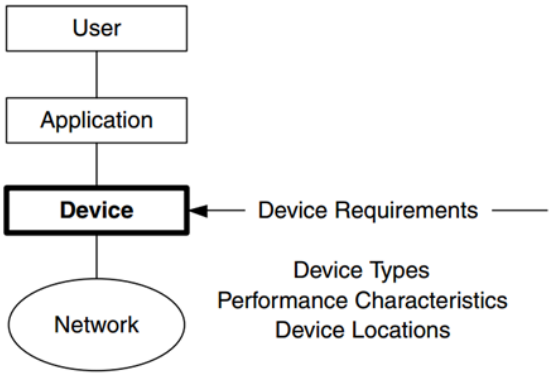

Figure 10. Types of Device Requirements (McCabe: 77)

Generic computing devices are devices, which resemble a normal computer like desktop and laptop computers. They act as an interface between the applications and the network. Their descriptions should have the device type along with the network controller interface type, processor, operating system and the most used applications. Device performance characteristics are important to describe, since often the device performance problems are misinterpreted as network problems.

Servers are important devices when considering network traffic flows, since usually they are designed to offer high-performance, predictable service to a large amount of users. Specialized devices are serving a specific purpose for their users. Specialized devices are often used for gathering, producing, and/or processing information to be sent to the users and they usually do not have direct access to the user applications.

They also tend to be location dependent, because of their nature of being the source or processing facility of the information.

Device locations, current and expected, need to be known, so the network traffic flows can be determined. The relationships between the users, applications and networks can be determined with the help of the locations of generic computing devices, servers and specialized devices. It is crucial to update the device locations when there have been changes, because the traffic flows might change dramatically. (McCabe, 2007: 76-83)

3.3.4 Network Requirements

In most of the cases, there are existing networks that are to be upgraded, instead of creating a completely new network from the scratch. So when creating the network requirements, if any existing networks would be co-operating with the new one, the existing network's characteristics, services and requirements should be considered. Figure 11 shows different kinds of network requirements; where from the most significant requirements are explained here:

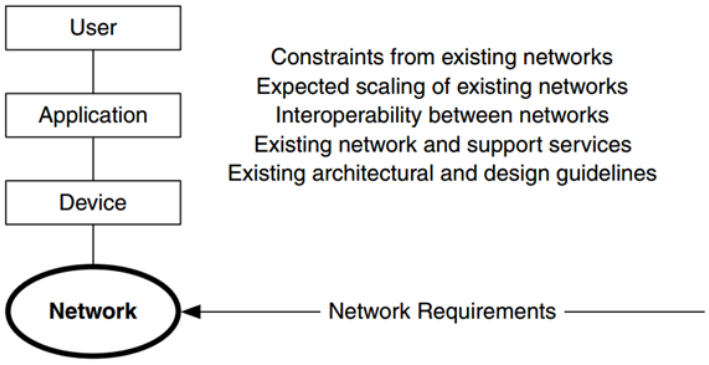

Figure 11. Types of network requirements (McCabe, 2007: 83)

Location dependencies are varied by the amount of changes made to the existing network. If a lot of changes are made to the system, it is more likely that the places and concentration of devices are about to change.

Performance constraints are defined by the existing network's performance characteristics, which will have an impact on the overall performance when the designed network will be built.
Also other network requirements, like the user, application and device requirements of the existing network must be considered and the network analysis should be done side by side with the new network's equivalents. (McCabe, 2007: 83-85)

3.3.5 Network Security and Management Requirements

In the analysis process the network management should be considered mainly from the view of monitoring. There are two types of monitoring; monitoring for event notification and monitoring for metrics. The monitoring in practice is collecting values from the various devices of the system, processing that data and showing the needed data to network operators.

Monitoring for event notifications is done by taking a frequent snapshot from the network to gain a better understanding of the network. Monitoring for metrics and network planning are long term processes where the monitored data is collected to create a set of characteristics to be used in the network performance analysis and for the management of the network. It is possible to collect one set of characteristics for all devices or an individual set for the each type of the network devices. (McCabe, 2007: 85-86)

The security requirements are created by determining the security risks for the both new and existing networks. The information about the current security situation and the requirements for new security features are obtained through security analysis. Table 1 shows an example of the risk assessment matrix, which can be done as part of the security analysis process.

Table 1. Example of the risk assessment (McCabe, 2007: 87)

Effect/ Probability	User Devices	Servers	Network Elements	Software	Services	Data
Unauthorized Access	B/A	B/B	C/B	A/B	B/C	A/B
Unauthorized Disclosure	B/C	B/B	C/C	A/B	B/C	A/B
Denial of Service	B/B	B/B	B/B	B/B	B/B	D/D
Theft	A/D	B/D	B/D	A/B	C/C	A/B
Corruption	A/C	B/C	C/C	A/B	D/D	A/B
Viruses	B/B	B/B	B/B	B/B	B/C	D/D
Physical Damage	A/D	B/C	C/C	D/D	D/D	D/D

Effect:
A: Destructive C: Disruptive
B: Disabling D: No Impact

Probability:
A: Certain C: Likely
B: Unlikely D: Impossible

The risk assessment matrix can be used to list the potential security problems, along with the system components to be protected and the likely-hood and severity of possible attack. The security requirements together with the risk assessment results are used to create a security plan and define security policies of the network. (McCabe, 2007: 86-87)

3.3.6 Financial and Supplemental Requirements

Also some other requirements, such as the financial and supplemental performance requirements and other possible requirements, apply to all of the system components. Financial requirements impact the whole system, since it is concerning the funding of the project. The funding usually includes the both types of expenditures; one-time and recurring costs.

One-time costs are the costs that are directly related to the planning and construction of the network. Such costs are generated from network architecture, design, purchasing, deployment, integration and testing along with all of the hardware and software components together with service provider service installations.

Recurring costs are the costs, which are expected to be paid on a periodic basis. Such costs are concerning the recurring tasks and components, which are expected to be

replaced or upgraded; such as network operation, administration and maintenance and provisioning and service provider charges.

Also so called supplemental requirements apply; they are defined by the following three characteristics of performance; first one is the operational suitability, which measures how well the customer operators can configure, monitor and adjust the network design. Second one, supportability, measures how well the customer can keep the system performing as designed over the life time of the system. Third, confidence, measures the network's ability to deliver data with a required throughput without errors or loss. Those three constrains should be identified, documented and validated with the customer at the analysis phase. It will clarify facts like the trade-offs between cost and performance, total ownership cost and operation limits to the customer. (McCabe, 2007: 88-90)

3.3.7 The Requirement Specification and Map

The requirement specification document has a prioritized list of the gathered requirements, which will be used in the architecture and design process together with the requirements map, which indicates the location dependencies between the devices and applications.

When creating the list of requirements, there will be various kinds of sources to gather the information from; the users, management, administration, staff and existing documentation about the network, devices and applications. All that data must be processed differently, depending on the source. Some of the data can be used as it is, but many must be derived or estimated from the source.

The requirement specification list specifies all the requirements with their priority levels, gathering sources and derivation methods; describing the reasons why some specific requirements were defined as the core requirements, network features, possible future requirements, rejected requirements or other informational requirements. Table 2 shows an example of a template suitable for the requirement specifications.

Table 2. Template for the Requirement Specification (McCabe, 2007: 91)

Requirements Specification							
ID/Name	Date	Type	Description	Gathered/ Derived	Locations	Status	Priority

When the requirements specification is ready, the requirements map can be created. The locations of the requirements are placed on a simple layout of the building to visualize the requirements in practice. (McCabe, 2007: 90-94)

3.4 Gathering, Developing and Analysing the Network Requirements

The process model in Figure 12 shows the steps to gather, analyse and develop the network requirements. The process consist of gathering, listing and managing of the requirements, developing measurable variables for the network to create the service metrics, describing the behaviour of users and applications, develop the performance requirements with their thresholds and limits, and map those requirements to the requirements map.

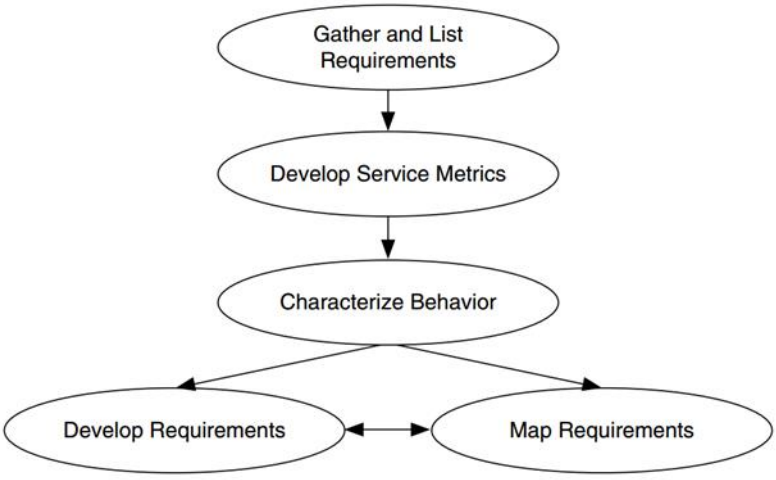

Figure 12. The requirements analysis process (McCabe, 2007: 100)

The network requirements analysis process starts from the defining the initial conditions to identify the borders of the project. The initial conditions include the current state of the existing the network (if such exists), type scope and the goal of the network project. (McCabe, 2007: 99-104)

3.4.1 Service Metrics

The service metrics measures the performance thresholds and limits and the performance characteristics in the system. The performance thresholds and limits are required to be defined in order to be used in the definition of the low and high performance levels of the network. Performance characteristics are used for identifying and defining the predictable and guaranteed performance levels.

One commonly used purpose for the service metrics is to separate responsibilities in the system, like between the end-to-end provider, WAN service provider and other inter-mediate providers. The service metrics are also useful with the network problem tracking and isolation.

There are their own service metrics for each of the performance characteristics. The service metrics for RMA are: Reliability, which is described with the mean time between failures (MTBF) and the mean time between mission-critical failures (MTBCF); Maintainability, described with the mean-time to repair (MTTR); and Availability, described with the MTBF, MTBCF and MTTR. Also uptime and downtime as percent of total time can be used. Service metrics for the capacity includes the data rates, data sizes and the service metrics for the delay, which includes the following; end-to-end or round-trip delay, latency and delay variation.

Service metrics are configurable and measurable quantities or they are derived from the measurable quantities and can be described in the terms of variables. Some variables used in the network devices can be for example the following ones; bytes in/out (per interface), IP packets in/out (per interface), Dropped ICMP (Internet Control Message Protocol) messages/unit time (per interface) and Service-level agreement (SLA) metrics (per user), which include the capacity limit, burst tolerance, delay and downtime.

The common network tools such as Ping (for round-trip delay and packet loss), Traceroute (for combined round-trip delay and per-link capacity with path traces) or such can be used for the measuring of the service metrics in addition to management protocols like SNMP (Simple Network Management protocol). Figure 13 illustrates an example of Ping being used for the availability monitoring and measuring delay and packet loss.

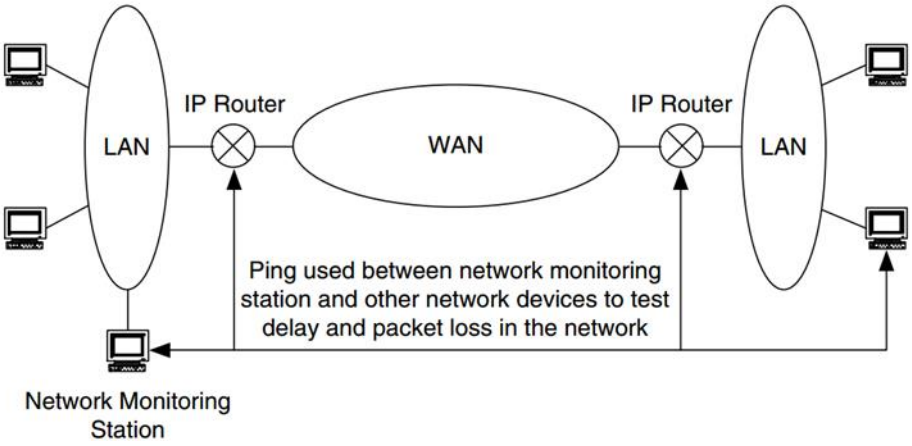

Figure 13 Using the Ping and IP packet loss as service metrics for RMA (McCabe, 2007: 112)

Even though Ping gives only an approximate information about the roundtrip delay times and packet loss, it is simple and can be used as a warning system for cases of network issues. It can also be used as a service metrics measurement method as example in Table 3 shows.

Table 3. An example of service metrics (McCabe, 2007: 112)

Service Metric		Where Metric Will be Measured in System	Measurement Method
1	LAN Delay	Between NM Device and Each Router in LAN	Ping
2	Wan Delay 1	Between NM Device and Local Router Interface to WAN	Ping
3	WAN Delay 2	Between NM Device and Remote Router Interface to WAN	Ping
4	LAN Packet Loss	At Each Local Router	SNMP

When the service metrics are developed, it is also good to determine, where each of the metrics are measured in the system and the available measurement methods. (McCabe, 2007: 109-113)

3.4.2 Characterizing Behaviour

To understand how users and applications are using the network, a behaviour characterization is needed. It is used to make the estimation of the network performance re-

quirements easier. The behaviour is represented either by the simple estimates of users session durations, the amount of active sessions and data sizes or by the complex and detailed models of the application and user behaviour.

The characterization begins from the creation a simple usage pattern from the user work times and durations, the total amount of users per each application, the frequency how often the user is expected to have that application session open (expressed by the amount of sessions per user per day), how long an average single application session will last (expressed in minutes) and estimate of the expected amount of users simultaneously having that application session open. By estimating the frequency and the duration of application sessions, along with the number of simultaneous sessions, it is possible to apply a modifier to the performance requirements for each application which are significant enough to be characterized.

When the application session frequencies, lengths and amounts have been determined, the next step is to determine the behaviour of those sessions. The characterization of application behaviour includes the data sizes of the application processes, which are passing through the network, the frequency and duration of data stream when it is passing through the network, along with the traffic flow characteristics and multicasting requirements of that application.

As with the user behaviour characterization, here also it is possible to apply either a simple estimate or a complex model to represent application behaviour. The most simple application behaviour model is to assume one application session to be active at any given time. Other way to characterize the application sessions is to apply ready models to the usage patterns and/or application behaviour; this is very effective on the applications which are well known.

As mentioned, even though the characterization could be done for all applications and users, it is not recommended; since only the most important applications and users are meaningful from the point of view of the network design and architecture. (McCabe, 2007: 113-116)

3.4.3 Developing RMA Requirements

Now the performance requirements should be developed and quantified if possible. To quantify the RMA requirements, two types of thresholds are needed to be discussed;

general thresholds and environment specific thresholds. The general thresholds are simple rules, which are based on the experience and can be applied to almost all networks. Environment specific thresholds are usually only for the specific network and cannot be used elsewhere. The environment specific thresholds are useful in determining the low and high performance levels for the network.

As already mentioned, reliability is a statistical indicator of the network failure frequency, which represents the unplanned outages of the network service. In simple systems the reliability can be measured with the mean time between failures (MTBF), which considers every failure to be equal and does not value one failure higher than the other. More accurate and suitable for the more complex networks or networks with limited resources is to measure the mean time between a mission-critical failures (MTCBF), which gives more accurate information, how many of those failures were causing mission-critical loss of the network service. Both measurement types are usually expressed in hours. MTBF is calculated by inverting the failure rate, which is estimated from the test results or analysed in terms of failures per hours of operation. The criticality of MTBCF is included so that the calculation is done only on the mission-critical components of the network. The calculation can be done per network component, where the failure rates (per hour) are added together and then inverted.

Maintainability is also a statistical measure, which represents the time to restore the system back to fully operational state after a failure. This is expressed with mean time to repair (MTTR). The restoration of services to operative status builds up from these stages; detection and isolation of the failure to a specific replaceable component, time required to deliver the new replacement component to the site where the failure has occurred, time to replace the component, and then test and restore the service.

As explained in Chapter 3.2.3., availability (A) is defined by the relationship between the reliability and maintainability, in other words the frequency of (mission-critical) failures and the restoration time of the service.

Availability takes only the unplanned outages and maintenance to the calculation, since planned maintenance is done at times when the components are not needed for performing the mission. The network availability analysis results in valuable information about the frequently failing components and ability to replace them preventively and schedule preventive maintenance. Also reducing the MTTR with hot spares at the site

or with a redundancy system can improve availability. Availability is also measured with the uptime, downtime and the error and loss rates.

Availability is often measured with the uptime and downtime in terms of percentage. Uptime is the time which the system and its components are available to the users (or devices or applications). Uptime is not only about the connectivity, it is about the user's ability to use the application through the network. The time user can connect but cannot use the application due to high loss rates or low capacity is also considered to be downtime. Few commonly used percentages of availability can be seen in Table 4.

Table 4. Uptime measured over different periods of time (McCabe, 2007: 119)

% Uptime	Amount of Allowed Downtime in Hours (h), Minutes (m), or Seconds (s) per Time Period			
	Yearly	Monthly	Weekly	Daily
99%	87.6 h	7.3 h	1.68 h	14.4 m
99.9%	8.76 h	44 m	10 m	1.4 m
99.99%	53 m	4.4 m	1 m	8.6 s
99.999%	5.3 m	26.3 s	6 s	0.86 s

Many networks operate at the 99.99% uptime level, and that can be used as general threshold for non-mission-critical networks. So the requirements with uptime requirements lower than 99.99% can be counted as low performance requirements and those having greater uptime requirements as high performance requirements. By adding the frequency of the uptime measurement to the requirements, the uptime is more binding and accurate. Without the measurement frequency, the 99.99% uptime can have 53 minutes of continuous downtime once a year, but if measured weekly it can be only 1 minute per week. This makes a great difference in practical availability.

Uptime can be measured for the whole network or the measurements can be done separately for some parts of the network, for example the network between servers and/or specialized devices might need higher uptime requirements than the network of the general users. Uptime can be measured between the user devices or between the network devices. Uptime can be measured in the terms of lacking connectivity or in the

terms of loss rate. For many applications, a two percent packet loss is enough to cause loss of the application session.

3.4.4 Developing Delay Requirements

The application delay requirements are measured in the network in terms of the end-to-end delay, round-trip delay and delay variation. The following delay limits and thresholds can be used to comprise the low and high performance delay requirements of the network; interaction delay (INTD), human response time (HRT) and network propagation delay.
Interaction delay (INTD) is an estimate of the time that the user is believed to be willing to wait for the system to respond during an interactive session. The normal and useful tolerance time range is between ten to thirty seconds, but can vary. INTD is used to characterize loosely interactive applications, where the user expects to have some waiting time.

The estimate of the time threshold, when the user starts to notice the delay in the system is called human response time (HRT). The time what the system can have delay, without user noticing it, is approximately one hundred milliseconds. In highly interactive applications the system delay can’t be more than HRT.

Network propagation delay is an estimate of the time that the signal takes, when it travels through the physical medium or link. It is dependent on the distance and technology used in the medium or link. The propagation delay provides the lower limits for the end-to-end and round-trip delays of the network and system. It can be used as a lower delay limit for applications.

Any of these delay limits and thresholds can be used as the delay limit for all services, or specially INTD as the limit for interactive services. But the network propagation delay always acts as the lower limit for the delay. An example of the delay estimates for user requirements can be seen in Figure 14.

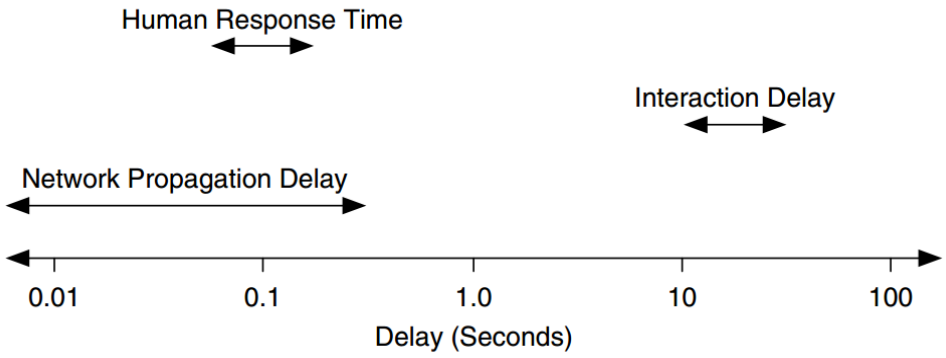

Figure 14. Delay estimates for the user requirements. (McCabe, 2007: 127)

The end-to-end delay and round-trip delay together are the sum of delays in different system components, which is caused by the propagation, queuing, transmission, input/output (I/O) device operations, switching and processing and which is usually measured with the Ping application. The thresholds and limits defined by the HRT, INTD and the network propagation are based on the different combinations of the following properties; physical limits of the network, device hardware and software performance, network protocol performance, application behaviour on the specific delay thresholds and the user interaction with the system at specific delay thresholds. When developing delay requirements, a limiting factor should be determined from the delay limits and thresholds. That limiting factor is the delay bottleneck for the whole system. Usually the limiting factor can be removed or reduced and it most probably will reveal a new limiting factor. That is why the process should be repeated as long as a limiting factor that cannot be removed or reduced can be found or the system delay is at acceptable level.

The delay variation together with the end-to-end or round-trip delay is used to de-scribe the overall delay performance requirement for the applications which are sensitive for variation in data arrival times. A usually used delay variation amount is 1 to 2 percent of the end-to-end delay or round-trip delay. (McCabe, 2007: 125-130)

3.4.5 Developing Capacity Requirements

In capacity requirements development, the focus is on the applications with large capacity requirements and on the applications which require certain range of capacities or

certain value of capacity, which is commonly expressed with the peak data rate (PDR), minimum data rate (MDR) and sustained data rate (SDR) or combination of those.

The estimation of the data rate is based on the information about the application's transmission characteristics. For some well-known applications the estimation is relatively easy, for example in telnet, the data rate is almost always small and in FTP, where the data rate is bigger. The data sizes and estimated or measured completion times (see INTD) are the tools to estimate the application data rates. An example of that can be seen in Table 5.

Table 5. Completion times and data sizes for the selected applications (McCabe, 2007: 132)

Application	Average Completion Time (Seconds)	Average Data Size (Bytes)
Distributed Computing (Batch Mode)	10^3	10^7
Web Transactions	10	10^4
Database Entries/Queries	2–5	10^3
Payroll Entries	10	10^2
Teleconference	10^3	10^5

If the transmission characteristics, such as the frequency of transmissions, sizes of the datasets to be transferred and duration of the transmissions, are well known, which is the case in transaction based applications (e.g. credit card processing) the upper and lower limits for the application's data rate or for the application's average data rate can be more easily estimated. (McCabe, 2007: 130-133)

3.4.6 Developing Supplemental Performance Requirements: Operational Suitability

The first of the three supplemental performance requirements is about how well the network can be configured, monitored and adjusted by the network personnel; if the network needs a lot of manual labour and it does not have good management tools, the lack of human resources on the network operations will wear out and frustrate the network operations personnel. So especially when the new network design is done due to increased performance requirements, the human resources should also be increased or even replaced with more skilled ones to keep the system performance at desirable

level. That is why this process must be well planned, documented and executed before the network reaches its initial operational capacity. The documentation must include, as precisely as possible, the requirements and constrains for the network and how they fit in to the design. These things will have a great impact on the level of automation in the design and to the skill level needed to operate it. This will clarify what changes there are needed to be done to the design to find balance between the amount of network personnel and the amount of automation or outsourcing.

Thus it is important to consider these following things during requirements analysis process: How the operators will monitor system performance to detect faults, failures or outages before the users do, or will the users detect and report the problems? How the users will interact with operators and system, and how they will report the problems, and how are the responses given and the reports tracked? When does an operations problem escalate maintenance action and how the ignorance of this is avoided? How will the operations personnel monitor system capacity and alert the management for the need of the capacity expansion? (McCabe, 2007: 134-139)

3.4.7 Developing Supplemental Performance Requirements: Supportability

The second supplemental performance requirement, supportability, is concerning the fact that the network must maintain the performance level that it has achieved when it was delivered to the end of its lifecycle. The five drivers of supportability are; RMA characteristics, workforce (including training and staffing levels), system procedures and technical documentation, standard and special tools, and spare and repair parts. Network needs two types of maintenance after it has been deployed; preventive and corrective.

The defining of the RMA requirements begins by making the descriptions for the types of mission scenarios, where the network will be used. Those descriptions answer for the following questions; when the network is used, how it will be used, what is important for the mission success, and how important it is for the users, along with the information how often the network will be used and the mission priority level.

Also in supportability, the usual problem is that the focus is in retaining the existing workforce and/or budget. The retraining or replacement of workforce may be needed, if new technologies unfamiliar to existing workforce will be used. It is also possible to

outsource the maintenance actions where existing workforce lacks the skills. Also the total outsourcing of the maintenance would come in place, if the network is only supporting the main line of business and there is no interest to develop highly skilled and expensive personnel to run it. When developing workforce requirements, it is good to have a clear picture on what kind of constrains there is to retaining the existing workforce. The skills, skill depth and the formal training of existing workforce should be surveyed and documented, which then can be used as a baseline for the new maintenance workforce.

The network needs three types of documentation to support it: technical documentation, maintenance procedures and casualty procedures. Technical documentation describes the system components with their characteristics and parts, and how they integrate to the system. It helps to recover quickly from the component failures, since it makes ordering of new components easier and faster. The maintenance procedures describe the periodic preventive component and system level maintenance actions and their schedules along with the needed pre-actions and testing. Casualty procedures describe the planned actions needed, when a fault occurs in the system, to restore the service as quickly as possible. Those procedures are divided into the immediate actions and supplemental actions. They describe the actions to restore part of the service to a safe state until the fault is isolated and repaired and the system is back in a fully restored state.

Proper tools, common and special test equipment, are needed for the support of the network. The tool requirements form up from the constraints set by the tools of the current system and if new tools will be acquired, those should be acquired together with the components to ensure that the tools are compatible for the planned network. Requirements should include the special test equipment, monitoring and diagnostic software with the descriptions how they improve the response time in fault or failure situations or in the performance issues. Also the ability to monitor, reconfigure or reset the components via out-of-band connections, like cellular connections, should be mentioned.

The repair and spare parts requirements are only qualitative constraints set by the owner of the system. They can, for example, describe the storage area and space for the spare parts, value of the spare parts and special privileges concerning spending of money in emergency situations. (McCabe, 2007: 137-143)

3.4.8 Developing Supplemental Performance Requirements: Confidence

The third supplemental performance requirement is confidence, which measures the network's ability to deliver data at the required throughput without error or loss. The confidence is estimated by using the error and loss rates, which are usually used only at the device level, but also some general performance estimates can be derived from them. The error and loss rates can be estimated from the following measurements measured from the system; bit error rates per link or circuit, packet loss rates between the network-layer routers and the end-to-end error/loss rates between the computers or applications.

To determine the thresholds for the error or loss rates, it is important to know the used applications. In different applications the guarantee for transmission varies between network layers, some use the transmission control protocol (TCP) at network layer or if the user datagram protocol (UDP) is used, the data-link or physical layer might provide the guarantee, and if not none of them is used, the application can have its own system to guarantee transmission in the application layer. The loss is usually measured at the network, data-link and physical layers and is indicated as a percentage of traffic in the network, as can be seen in Table 6.

Table 6. An example of the loss threshold (McCabe, 2007: 144)

Packet Loss Rate (% of Total Network Traffic)	**Maximum Total Time (Per Week)**
2% or Greater	Up to 1 Minute
Less than 2%	Remainder of Week

The network tool Ping can be used for the loss rate measuring at a general level, and if more accurate measurements are needed to be done, the SNMP polling of the router statistics or the remote monitoring variables are needed. One good to know issue with Ping is that all devices do not allow Ping's ICMP (Internet Control Message Protocol) packets with default configuration, so some consideration should be done when analysing the results. Ping is good to be used as a first threshold trigger, and when it goes off, it causes more accurate measurements to be started. (McCabe, 2007: 143-145)

3.4.9 Environment-specific Thresholds and Limits

The general thresholds and limits are the common estimates for the low- and high performance requirements for the network. Those can be used in cases where some inaccurate information about the performance thresholds is given, but no accurate information about the applications, users and/or devices is available. That information can be used to develop the thresholds and limits specific to that network.

The environmental-specific thresholds and limits are unique to each network; they take into account also the mission of the users of the network along with the unique local features and requirements of applications, users and devices. Thus the thresholds which distinguish between low and high performance are also unique to each environment. As it is with both the general thresholds and the environmental-specific thresholds, the reason to develop such is to find the applications or devices which have high-performance requirements. If crucial high-performance applications and devices are found, the network will most probably concentrate on the supporting those applications and devices along with their users.

To develop the environmental-specific thresholds and limits, a comparison between performance requirements of the applications is needed. The performance characteristics, meaning RMA, delay and capacity, of the applications are plotted and the plots are used to compare the relative performance requirements to create the thresholds or limits for that characteristic. An example of capacity requirements plot can be seen in Figure 15, where a group of applications have capacity requirements under 2 Mbit/s and two applications have their clearly higher requirements for capacity. The high performance threshold could be placed so that those two are grouped together to make them the high-performance applications, or so that only the application with the highest capacity requirements are considered as the high-performance application, depending on the needs.

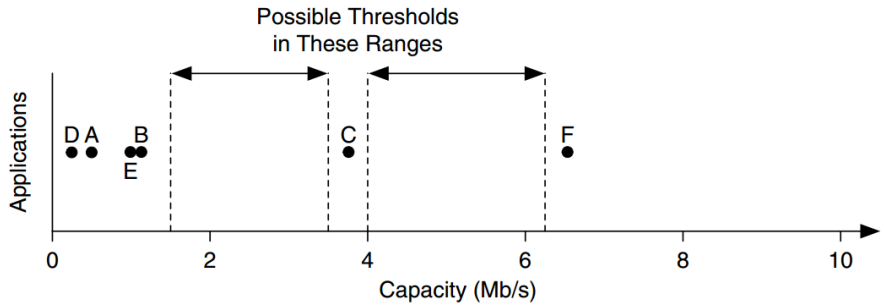

Figure 15. A plot of capacity requirements with possible thresholds (McCabe, 2007: 146)

The applications can also be spread all over the range of values, when the threshold between high and low performance might be difficult to determine. In those situations, a discussion with the management and users of the network might help. (McCabe, 2007: 145-147)

3.4.10 Requirements for Predictable and Guaranteed Performance

While determining the performance requirements, thresholds and limits, the requirements for the predictable and guaranteed services (if such exist) must be taken into count. Those are more dependent on the predictability and reliability and also have more strict performance requirements than the best effort services. The guaranteed services also have accountability as its property. The need for the implementation and maintenance resources (financial, manpower, intellectual and time) will increase gradually when moving from the best effort services towards the predictable and guaranteed services.

If some applications and/or users are more important for the mission of the organization, the traffic flows of those components will require more support, thus the requirements for those components should be predictable. The predictable performance determining is based on the following steps: The first is to determine if the application is mission-critical, rate-critical, real-time or interactive. The second step is to determine is there any environment-specific performance thresholds or limits. The third one is to apply the general thresholds and limits, if needed. And the last one is to discuss with the users or management to agree on the predictable requirements.

The degree of needed support for the performance requirements in the network defines is the requirement guaranteed or not. If the network has guaranteed performance re-

quirements, a service-level agreement (SLA) must be made between the users of the application and the network service provider. The SLA must include the information about the types of the guaranteed performance requirements, when and where they apply, how will they be measured and verified, and what happens when the requirement is not met. The guaranteed requirements must be counted end-to-end between the source and destination.

Guaranteed requirements are identified as the predictable requirements; applications which are mission-critical, rate-critical, real-time or interactive may have guaranteed requirements or when the high performance requirements will be identified. (McCabe, 2007: 147-149)

3.4.11 Developing Requirements Map and Specification

The requirements map collects and combines the information about the locations, devices and where the applications are used and binds that information with geographically described environment, which can be building, campus area, wide area or something between or an even combination of them.

From that map the correlation which applications are used in which parts of the network and how the traffic flows might form out within the application area, between the devices and between the applications. An example of the requirements map can be seen in Figure 16, where the campus area includes the specialized devices, servers and the user groups and the application usage. Even though showing a single user with his/her computer is not giving any valuable information, the amount of users give important information about the traffic flows.

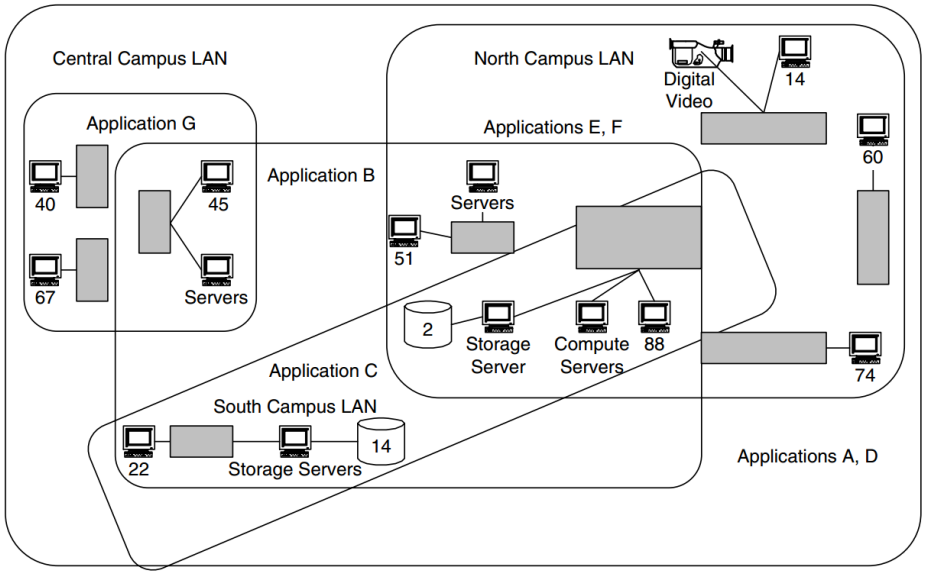

Figure 16. An example of campus area requirements map (McCabe, 2007: 151)

The requirement specification builds up from two parts; the initial conditions and the listing of requirements. As can be seen in Table 7, the initial conditions have the type, scope and goal of the network project determined along with the driving forces, which can be political, financial and/or administrative. The determination of the need of high or low performance can also be in the initial conditions. It also should have brief evaluation of the current situation and problems with the possible resource and schedule estimate.

Table 7. A template for the initial conditions (McCabe, 2007: 152)

Requirements Specification	
Section 1: Initial Conditions	
Project Type	Upgrade of building network
Project Scope	Single building, two floors, approximately 150 users
Project Goals	Improve performance to all users, particularly some mission-critical applications, and increase security to Internet
Other Conditions	Financial TBD
Problem Evaluation and Definition	Application performance has been a recurring problem, so management wants to upgrade network and has suggested upgrading interfaces to Fast Ethernet. Some users have GigE interfaces on their workstations.

As can be seen in Table 8, the listing of the requirements contains the information about the collection date, requirement type, description and the source where it is derived or gathered from along with the location, status and priority.

Table 8. An example of requirements specification (McCabe, 2007: 152)

Requirements Specification							
Section 2: Listing of Requirements							
ID/ Name	**Date**	**Type**	**Description**	**Gathered/ Derived**	**Locations**	**Status**	**Priority**
1	14Jan01	User	User distribution is: 60 engineers, 15 HR and Finance, 30 Manufacturing, 10 Management, 30 Sales/Marketing, 5 Other	Gathered from Management	TBD	Info	TBD
2	14Jan01	Network	Each area of the building must support Fast Ethernet connections to the backbone	Gathered from Management	TBD	TBD	TBD
3	14Jan01	Application	Database, Visualization, Manufacturing, and Payroll applications are considered mission-critical for this company. More information needed.	Gathered from Management	TBD	TBD	TBD
4	14Jan01	Application	Payroll application (PAY1) requires 100% uptime (while in operation) between Finance and outside payroll company	Gathered from Management	TBD	TBD	TBD
5	14Jan01	Network	Company must be kept secure from Internet attacks	Gathered from Management	TBD	TBD	TBD

The information builds up along with project. The first requirements are probably gotten from the initial meetings, and it can be collected with a questionnaire and meetings. (McCabe, 2007: 149-154)

3.5 Traffic Flow Analysis

The traffic flow analysis is based on the information collected in the requirements phase. The collected requirements along with the user and device locations from the requirements map are used for estimating the data flows of the network. (McCabe, 2007: 161)

3.5.1 Flows

Traffic flow (or data flow) can be described as an end to end connection between the source and the destination application, device and/or user. The flows may be bidirectional and both directions can be described as a single flow with common characteristics or as a two separate flows each having its own characteristics and requirements. In Table 9 are shown the commonly used flow characteristics.

Table 9. Common flow characteristics (McCabe, 2007: 163)

Flow Characteristics	
Performance Requirements	Capacity (e.g., Bandwidth)
	Delay (e.g., Latency)
	Reliability (e.g., Availability)
	Quality of Service Levels
Importance/ Priority Levels	Business/Enterprise/Provider
	Political
Other	Directionality
	Common Sets of Users, Applications, Devices
	Scheduling (e.g., Time-of-Day)
	Protocols Used
	Addresses/Ports
	Security/Privacy Requirements

A single session of an application creates an individual flow with a set of requirements. The individual flow can be a part of a composite flow, which consists from multiple individual flows or applications and their requirements that share a common link, path or network. If Individual flow has guaranteed flows it cannot be part of a composite flow; those flows should always be kept individual.

As seen in Figure 17, the flows can be represented as two sided bidirectional arrows or single sided unidirectional individual arrows.

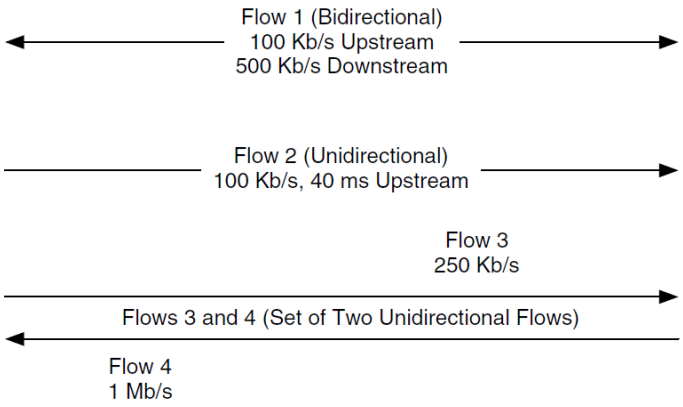

Figure 17. Flows are represented with uni- or bidirectional and separate arrows. (McCabe, 2007: 164)

The one sided arrow represents unidirectional flow with its requirements to that flow direction. If the arrow is two sided, it has the same requirements on both flow direc-

tions. The individual flows can also be combined in to a single arrow to represent the composite flow as in figure 18.

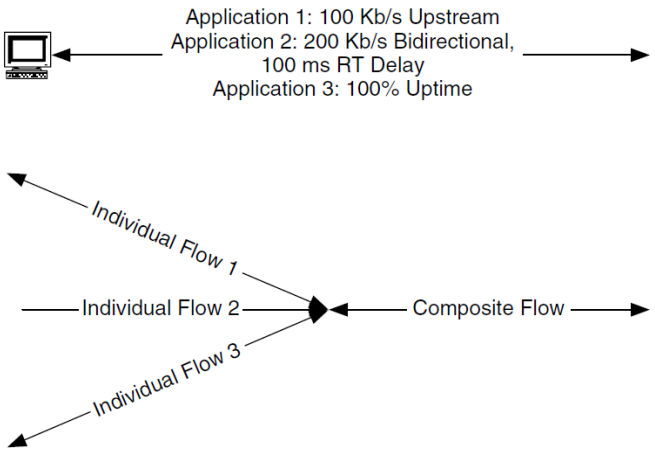

Figure 18. Examples of composite flows. (McCabe, 2007: 165)

The composite flows have the combined flow requirements of the combined individual flows which share a common link, path, or network. (McCabe, 2007: 162-165)

3.5.2 Identifying and Developing Flows

The requirements specification information can be used to identify and create the traffic flows, since it has the information about the users, applications and their behaviour, devices, locations and performance requirements. Flows should not be constrained by the existing network, topologies or technologies, because the flows are the driving force behind good design and architecture decisions. The flow determination is based on the requirements and locations of applications and devices that generate or terminate the traffic flows. The process can be seen in Figure 19.

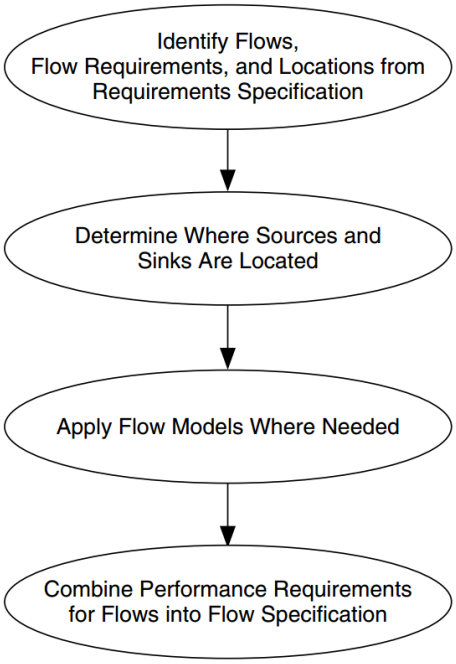

Figure 19. The process for identifying and developing data flows. (McCabe, 2007: 168)

The flow identification and development begins from the identification of flow sources and sinks. Those are the devices (and applications), which are believed to generate (source) or terminate (sink) the traffic flows. The requirements from the requirement specification, together with the information where and how each device and application is used, are used in the determination of the flow source devices and flow sink devices. When each flow with their sink and source along their locations is clear, they will be combined with the performance requirements to create a flow specification.

The identification of flows from the application point of view has few common approaches, which can be applied. The identification process might need one or more approaches to be used.

The first one is to focus on the particular application, application group, device or function. In this approach the benefit on the time used on the identifying the flows of chosen applications is maximized. As can be seen in Figure 20, the locations of users, applications and devices are used to make a map, which together with the behaviour of users and application is used to estimate or determine the flow occurrence between the networks, device groups or devices.

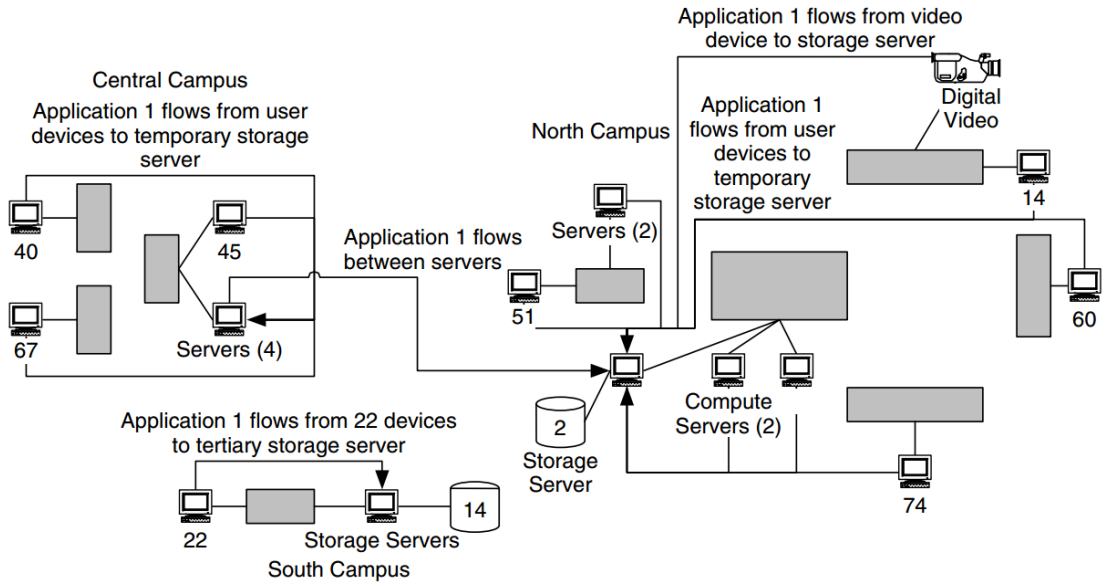

Figure 20. Flows estimated between the devices and application 1. (McCabe, 2007: 172)

Figure 21 is an example of single application focused approach, where the flows F1, F2 and F3 are representing the single-session requirement for the Application 1 for each building, and the flow F4 the performance requirement for the server-server flow between the Central and North Campuses.

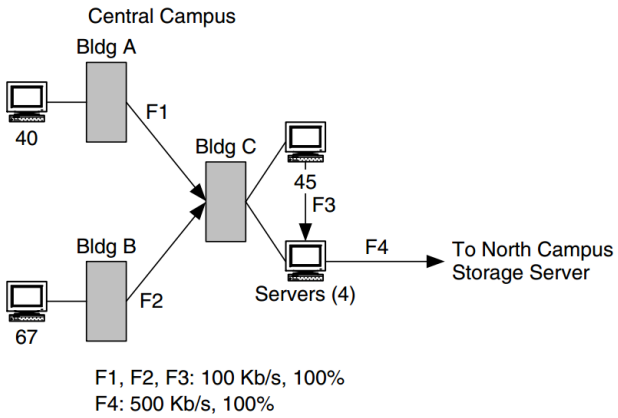

Figure 21. Performance information added to the Central Campus flows for the application 1. (McCabe, 2007: 173)

The second approach is to develop a common profile or for selected applications to be applied to across user population. This is recommended to be used when many different applications or application groups share same performance requirements. It can also be used when a group of users (or all users) shares the same performance re-

quirements for the set of common applications. The use of profiles will simplify the flow map and same information will not be documented multiple times. Figure 22 is an example where performance profile P1 is applied across the users using the Application 1. As can be seen at the top part of the figure, P1 have the following performance requirements: capacity=100 kbit/s and reliability=100%. Six of the flows have now P1 profile, flow F4 have the requirements mentioned in Figure 21, flow F5 combines the requirements of 51 users to the two servers in the same building, flow F6 does the same to the requirements of the digital video camera and 14 users in that building. Flow F7 also combines the performance requirements of 88 users and 2 servers.

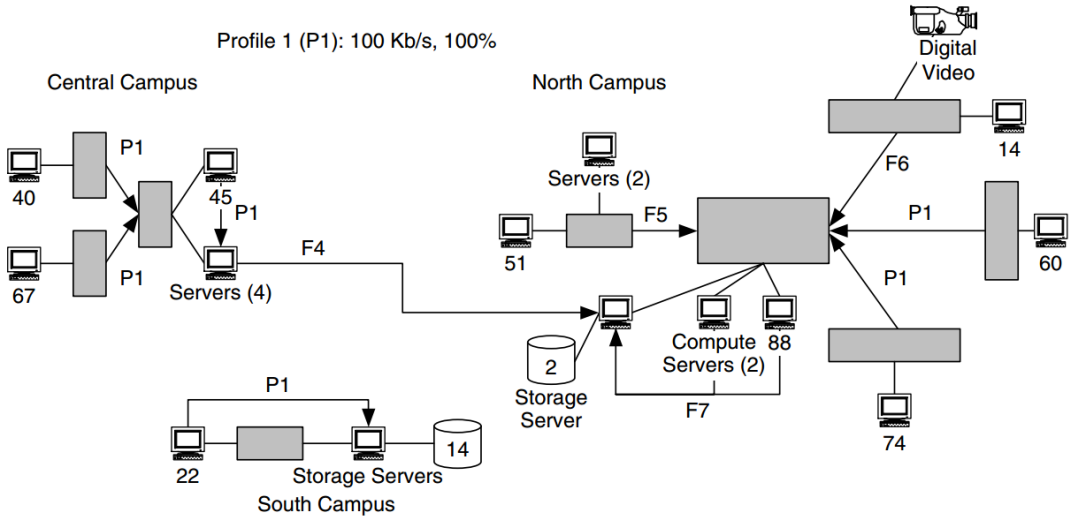

Figure 22. The performance profile P1 applied to multiple flows with the same performance characteristics. (McCabe, 2007: 173)

Third approach is to choose top N applications (for example top 5 applications) to be applied across entire network. It is a combination of the previous two approaches, where more than one application or application group is chosen, and the result will be similar to last approach's profile. This approach will help determine the most important requirements for the network. Those applications will act as performance drivers for the network, which most likely ensures that other applications will also meet their performance requirements.

The usage of different approaches in different parts of networks and with different scale is encouraged. Usually the top N applications are used network widely, when profiles and focusing on particular application are tied to specific locations. (McCabe, 2007: 161-174)

An example of the flow requirements can be seen in Table 10. The example of the flow requirements list describes the flows and the flow specific performance requirements.

Table 10. Example of the performance requirements for the flows (McCabe, 2007: 204)

Flow ID	Performance Requirements	
	Capacity (Mb/s)	Delay (ms)
F1: Flow Type 1		
Synchronization Files	320	100
Update Files	1600	1000
Final Files	160	10^5
Result for Flow Type 1	1600	100
F1: Flow Type 2		
Update Files	80	10^4
Final Files	160	10^5
Result for Flow Type 2	160	10^4
Result for F1	**1760**	**100**
Result for F2	**1760**	**100**
Result for F3	**1760**	**100**
F4: Flow Type 1	1600	100
F4: Flow Type 2		
Update Files	320	10^4
Final Files	640	10^5
Result for Flow Type 2	640	10^4
Result for F4	**2240**	**100**
Result for F5	**16**	**10^3**
Result for F6	**80**	**10^2**
Result for F7	**16**	**10^3**

This flow requirements list is used for describing the flows with the gathered performance requirements. These descriptions can be used for the definition of the flows on the flow map. (McCabe, 2007: 204)

3.5.3 Data Sources and Sinks

The determination of the data sources and sinks will help identify the flow direction. The traffic flow begins from the source, and ends to the sink. As almost all devices are generating and accepting data, they are considered to be both, the source and the sink. But some devices are mostly sources, like servers, computing clusters or other devices producing high amounts of data, as well as some devices are mostly sinks, such as data storages and archival devices along with other devices which use large amounts of data. The data sinks are represented with asterisks and sources with dots as seen in Figure 23.

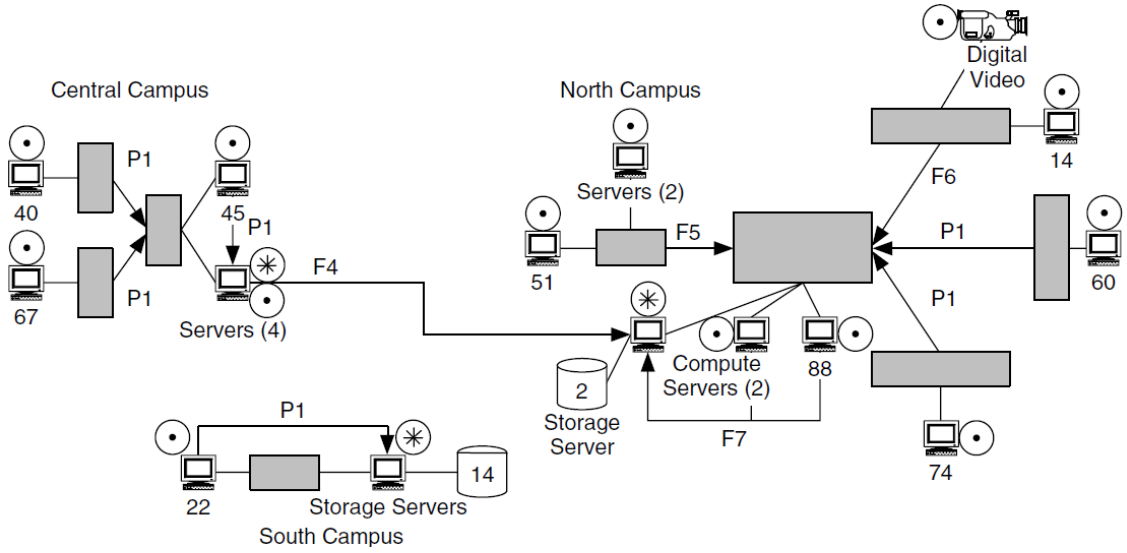

Figure 23. An example of a flow map with sources and sinks marked (McCabe, 2007: 177)

After the flow sinks and sources are defined, it is possible to determine the arrow (flow) direction on the flow map. The flow direction is usually towards the flow sink and away from the flow source. (McCabe, 2007: 175-180)

3.5.4 Flow Models

One way to describe the flows is to use well-known flow models. Flow models represent groups of flows, which have specific and consistent behaviour characteristics. This kind of flows inside flow models applies to a single application. The primary characteristics of the flow models are directionality, hierarchy and diversity. Directionality is describing the flow's property to have (or not to have) more requirements to the other

direction than the other. Hierarchy describes the segmentation of the network and the hierarchical model. Diversity describes the interconnections between hierarchical levels in the networks to balance the load or to make redundancy connections. Common flow models are peer-to-peer, client-server and hierarchical client-server models. (McCabe, 2007: 21,180-181)

In peer-to-peer flow model, the user and the applications have quite consistent flow behaviour throughout the network. As the model type already reveals, the two interacting users and/or applications are peers, working at the same level at the hierarchy, both being equal sink and source to each other and all flows being equally critical (or non-critical). Thus all the flows in this model are identical as can be seen in Figure 23 and can be described with single profile. The peer-to-peer model may apply also to flows' of user groups which have identical needs to access each other services.

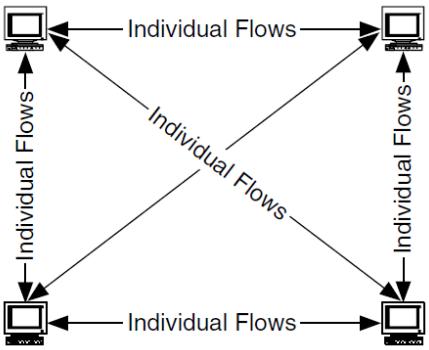

Figure 23 Peer-to-peer flow model (McCabe, 2007: 181)

The client-server flow model, seen in Figure 24, is the most common flow model used in the networks. It has high directionality and hierarchy and the flows are bidirectional and asymmetric; the client requests are relatively small compared to responses from server. In client-server flow model, the server acts as a data source and clients are acting as data sinks. The flows from the server to the client are the critical flows in this model.

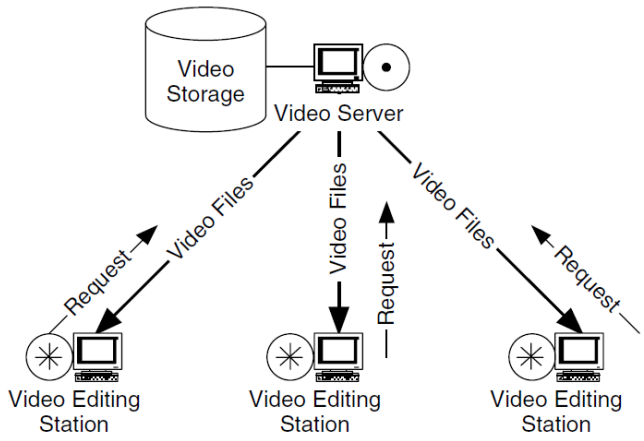

Figure 24. Client-server flow model (McCabe, 2007: 184)

The hierarchical client-server flow model is coming more and more common, as it is used in the Web service networks. It has the same characteristics as the client-server model, but between the servers there are multiple layers, or tiers as can be seen in Figure 25.

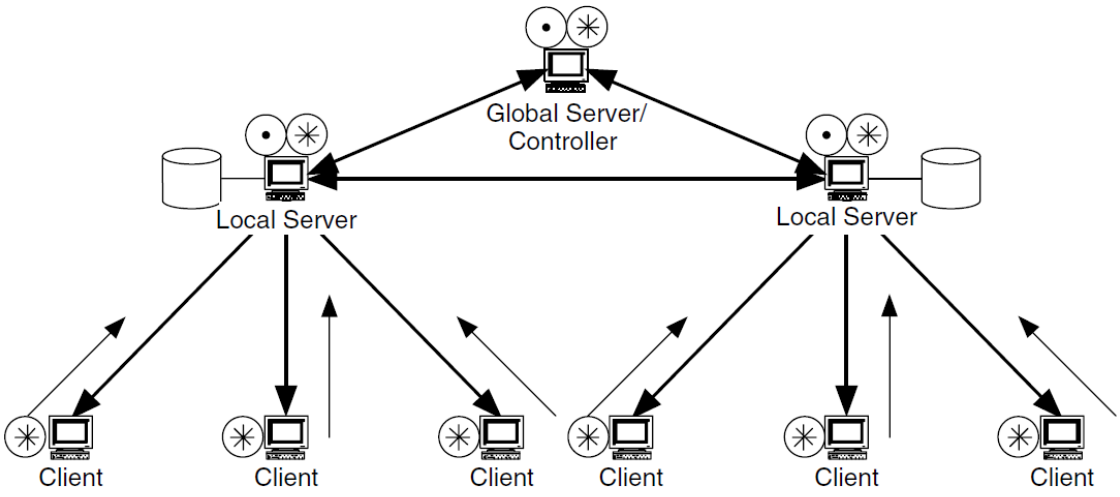

Figure 25. Hierarchical client-server flow model (McCabe, 2007: 186)

In this model also the server-to-server flows and the server to management device (server-to-manager) flows are possible. Due to added layers, a server may act as a source, sink or even as both at the same time. (McCabe, 2007: 15, 181-187)

3.5.5 Flow Prioritization

Flow prioritization is useful when considering the importance of each flow. The flows can be prioritized based on the common flow characteristics mentioned in Chapter 3.4.1, and the drivers for the prioritization can be business objectives, political objectives, performance requirements of the flow, security requirements of the flow, or the flow's number of users, applications and/or devices. The purpose of the prioritization is to determine which flow gets the resources first, or which flow gets the most resources.

Most common resource is funding, which is divided among the flows based on the priority; the highest priority gets most of the funding and lowest the least. The parameters used in the prioritization can be number of users per flow or/and the flow performance characteristics. (McCabe, 2007: 191-193)

3.5.6 Flow Specification

When the flows are identified, defined and described, the results will be combined to create a flow specification, or flowspec. The flows of the network along their performance requirements and priority levels are listed in the flow specification. It also describes the flows with best-effort, guaranteed and predictable requirements, including the mission-critical, rate-critical, real-time, interactive, and high and low performance flows. The flow specification can also be used for combining the performance requirements, if there are multiple application requirements in one composite flow or if all the flows in a section of a path are needed to be combined.

As can be seen in Table 11, there are three possible types of flow specifications to describe the individual or composite flows; one-part, two-part or multi-part. The level of detail in each type is dependent on the requirements of the flows; are they having best-effort, guaranteed and/or predictable requirements. A one-part flowspec only has the flows with best-effort requirements, where the performance requirements are described by the capacity. A two-part flowspec has flows with the predictable requirements and it may include also the best-effort requirements flows as well. Those flows' performance requirements are described by the capacity, reliability and delay. A multi-part flowspec has flows that have guaranteed requirements and it may have predictable and or best-

effort requirements flows also. These performance requirements of the flow are also described by the capacity, reliability and delay.

Table 11. Types of flow specifications with descriptions

Flow Specification Type	Types of Flows	Performance Description
One-Part	Best-Effort Individual and Composite	Capacity Only
Two-Part	Best-Effort and Stochastic, Individual and Composite	Reliability, Capacity, and Delay
Multi-part	Best-Effort, Stochastic, and Guaranteed, Individual and Composite	Reliability, Capacity, and Delay

The one-part flowspec is good for describing simple networks, or when the amount of information about the network flows is insufficient. The two-part flowspec is good when it is needed for the balance between the amount of details and the ease of development. The multi-part flowspec can be used when the network is more complex and has guaranteed flows.

A flowspec algorithm is a mechanism used to combine the performance requirements of multiple applications to create a composite flow or to combine multiple flows to create a flow for a section of a path. The outcome will have the optimal performance (capacity, delay, RMA) for that flow or group of flows. The flowspec algorithm is bound by these rules:

1. Best-effort flow calculation includes only the flows with the capacity requirements.
2. All performance requirements available are used in the calculations of predictable requirements flows. Each characteristic (capacity, delay, RMA) of performance requirements are combined to maximize the performance of each flow.
3. Guaranteed requirements flows are created by creating the individual flows for the each individual performance requirement.

As the one-part flowspec has only best-effort flows, the calculation is done by adding together the capacity requirements of each flow. The result is the total best-effort capacity (C_{BE}) as seen in Figure 25.

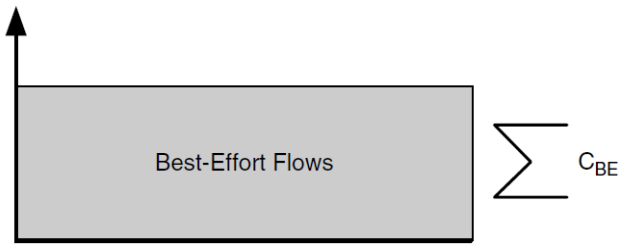

Figure 25. One-part flow specification (McCabe, 2007: 196)

The two-part flowspec is built on the top of the one-part by adding the predictable capacities, and delay and RMA on it. The predictable capacities are calculated the same way as the best-effort capacities. To achieve the best performance for both delay and RMA, the delay in predictable flow requirements is the minimum delay of all flows and the RMA is the maximum RMA of all flows. The result is total predictable capacity (C_P), predictable delay (D_P) and predictable RMA (R_P) as seen in Figure 26.

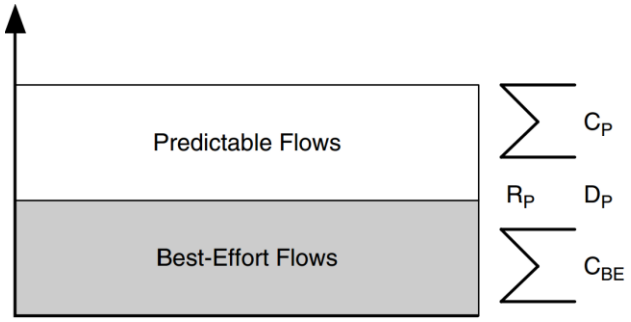

Figure 26 . Two-part flow specification (McCabe, 2007: 196)

The multi-part flowspec builds on top of the two-part flowspec, but adds the guaranteed requirements. Each individual guaranteed requirements set, which consists from guaranteed capacity (C_i), guaranteed delay (D_i) and guaranteed RMA (R_i), are added to flow-spec as seen in Figure 27.

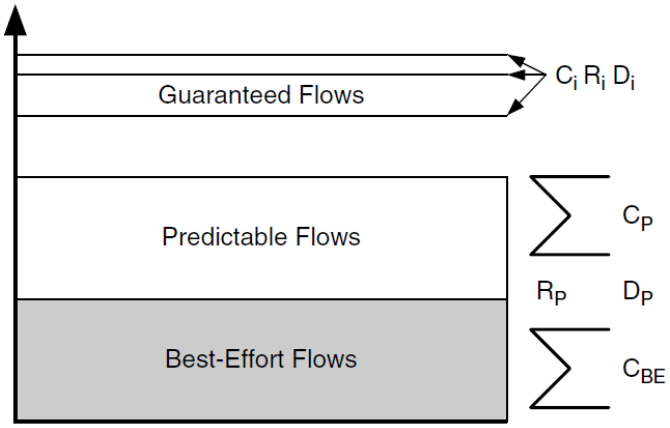

Figure 27. Multi-part flow specification (McCabe, 2007: 197)

Each of the guaranteed flow requirements sets are listed individually. This is done because it ensures that those guaranteed flows are really guaranteed throughout the network. (McCabe, 2007: 193-197)

4 Introduction of Present Network Infrastructure

The FinnRef network consists of 13 stations which are connected through a VPN (Virtual Private Network) to the Finnish Geodetic Institute's network. The control of the stations is done from the FGI's FinnRef server, which also downloads the GPS data from the receivers of the stations. The downloading along with the data handling is done using the several scripts that are modified for the use of FinnRef.

The FinnRef network structure is very simple but in the other hand, it has a wide variety of different kind of devices and long geographical distances between sites which are bringing more complexity to the network. The analysis process of such contradictive network will generate valuable information and eventually more reliable network. In this chapter the data connections and the stations with their equipment are investigated, also the usage of the network is reviewed.

4.1 System Premises

The FinnRef network, at the initial point of this thesis, had 13 stations all around Finland, from the south parts of Åland to the northern parts of Lapland, as can be seen in Figure 28. The FinnRef network consists of various types of buildings. Some of them are bigger research premises, which have a lot of other research equipment in the same building, while some are small huts in a rural area, including only the GNSS devices and possibly some other research devices, for example a seismometer. Most of the smaller buildings are FGI's own, but majority of the larger research premises are owned by the local research institute.

The FinnRef system's servers are at Finnish Geodetic Institute's (FGI) offices, which are located in Southern Finland in the municipality of Kirkkonummi. There are about 80 people working and doing scientific research at the offices. At the FGI offices, the power supply of the servers and core network equipment is redundant; they have an Uninterruptible Power Supply (UPS) to keep them powered if the main power has an interruption.

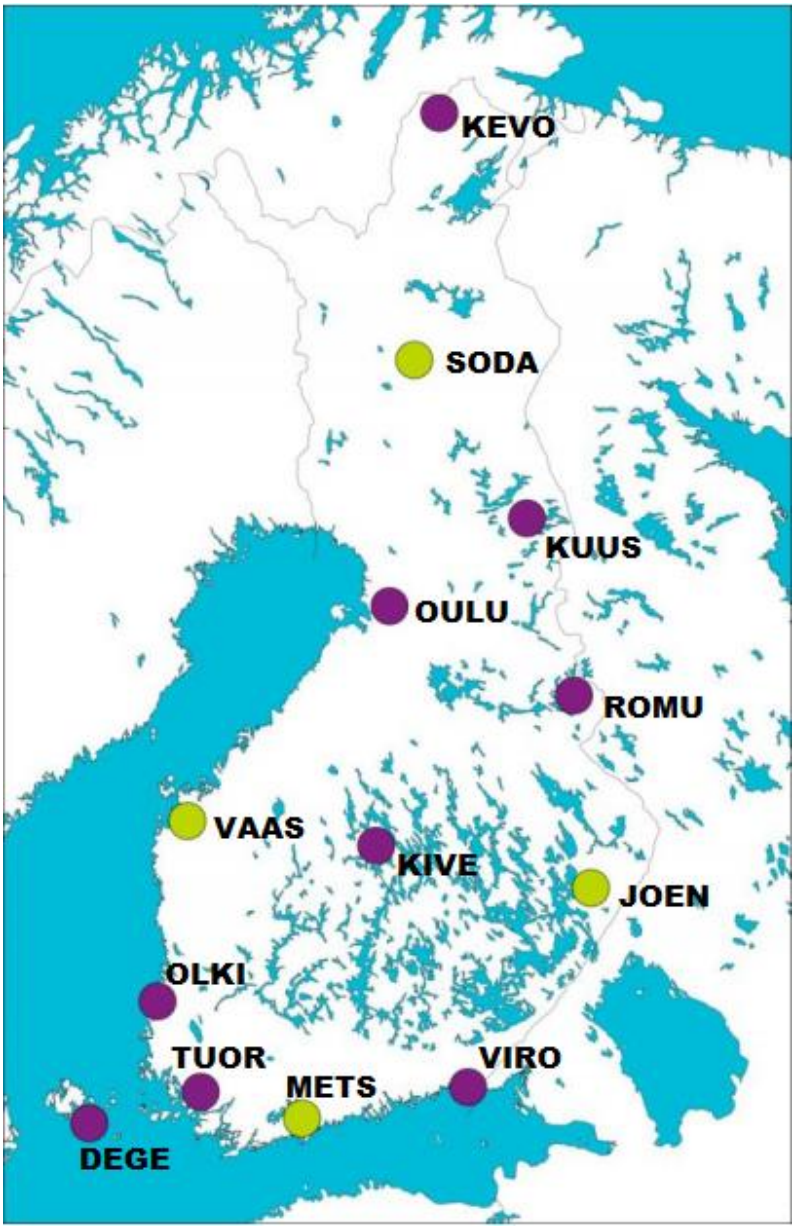

Figure 28. Map of the initial FinnRef network (Koivula, 2013)

From the usage point of the view, the network can be divided in to three parts; four of the stations are also part of the European Euref Permanent Network (EPN), one of which is also a part of the International GNSS Service (IGS) network. The EPN stations are Sodankylä (abbreviation: SODA), Vaasa (VAAS), Joensuu (JOEN) and Metsähovi (METS). Metsähovi's station is the one which is also in the IGS network.

Other stations of the network are Degerby (DEGE), Virolahti (VIRO), Tuorla (TUOR), Olkiluoto (OLKI), Kivetty (KIVE), Romuvaara (ROMU), Oulu (OULU), Kuusamo (KUUS) and Kevo (KEVO).

4.2 Data Transmitted in System

Two types of GPS data are transmitted in the network; Real-time and non-real-time GPS data. Both of them have their own specific form and characteristics, even though the data carried is similar.

The FinnRef stations are collecting hourly raw data files, which are converted to RINEX files in the FGI, with the following settings; it records data in 30 seconds intervals (sampling rate) and the elevation cut off angle is set to 5 degrees, which means satellites under 5 degree angle from antenna are not counted in. The variables which are collected are both L1 and L2 frequencies' phase observations, C/A-code, P-code of L1 and L2 (which are called P1 and P2 observations) and the Dopplers of both L1 and L2 frequencies (which are called D1 and D2 observations).

The real-time streams are streamed in the RTCM format from the stations to FGI's real-time server. The real-time server transforms the RTCM data to the NTRIP format and forwards it to EPN caster server.

4.3 System Devices

The FinnRef network consists of various types of devices, oldest of which are from the early nineties and some of them are the latest state of the art technology. The devices are located at the stations and at the FGI offices, which includes servers, network and GPS receiver equipment.

4.3.1 GPS Receiver

Ashtech Z12 (see Figure 29) is used as the GPS receiver at the stations. The receiver was released in the early nineties, and among geodesists it is said to be one of the best receivers ever made, due to its "Z-tracking" technology, which has the highest signal-to-noise ratio of all the codeless L2 tracking techniques (Rizos, 1999). The receiver is powered by two 12 volt batteries, which have an upkeep battery charger connected to them. This system keeps the receiver running for 1-2 days when the main power is lost.

Figure 29. Astech Z12 GPS receiver (Ashtech Inc., n.d.)

The receiver is connected to an external antenna, which is “choke ring” –type GPS antenna and equipped with preamplifier. The antenna mast height varies depending on the station, in KIVE, OLKI and ROMU the mast is 1.85 meter high and the most of remaining stations have 2.5 meters high antenna mast, but few exceptions apply: The KEVO station has five meter high antenna mast, OULU mast eight meters high and METS station twenty meters high. In Figure 30 there is the JOEN station antenna mast and station cottage.

Figure 30. JOEN station showing the antenna mast and the station cottage (Koivula, 2014b)

From the data communications point of view, the receiver has four serial ports, although two of those are virtual: When Y-shaped three way cable is plugged to port one, the port splits in to A and C ports, and when to port two, it splits to B and D ports. The hourly GPS data goes through port A in every station, and the real time data (RTCM) goes through C or D, depending on the station.

4.3.2 Network Devices

Each station has three network devices; the first device after the GPS receiver is Moxa Nport 5210, seen in Figure 31, which converts the serial data from the receiver to the Ethernet network suitable form. The serial settings in both devices, receiver and N-port, must be the same. The Nport also needs the server IP address and other network settings to establish the connection to the server in FGI.

Figure 31. Moxa Nport 5210 (Moxa Inc., 2012a)

The second device is a router manufactured by Zyxel acting as firewall and VPN endpoint. At the FGI's end, there is a separate router/firewall also manufactured by Zyxel handling the VPN (Virtual Private Network) connections to the stations. At the FGI offices there are a few network switches on the way from the FGI firewall to the server. Third device is the Internet service provider equipment, which varies depending on the service provider and the connection type.

4.3.3 Servers

The hourly data from stations is downloaded by the Linux server at FGI. It has Nport drivers installed and a virtual serial port for each of the Nport devices. Also the management and controlling of the firewalls and Nports is done through this server by using telnet or SSH (Secure Shell).

The download of the GPS files is executed hourly by using a script (series of commands) made for that purpose. It first reads the configuration file, which is unique for every station. Then the download scripts, which are based on the Unavco's Remote33 script compilation (UNAVCO, 2011), use the Zmodem script to download data from the receiver. Zmodem is used to directly control the receiver through a serial connection, it uses Ashtech serial port commands $PASHQ (for queries) and $PASHS (for settings). In this phase, also all of the data directories are created, and the log files are opened for logging. The file processing process flow can be seen in Figure 32.

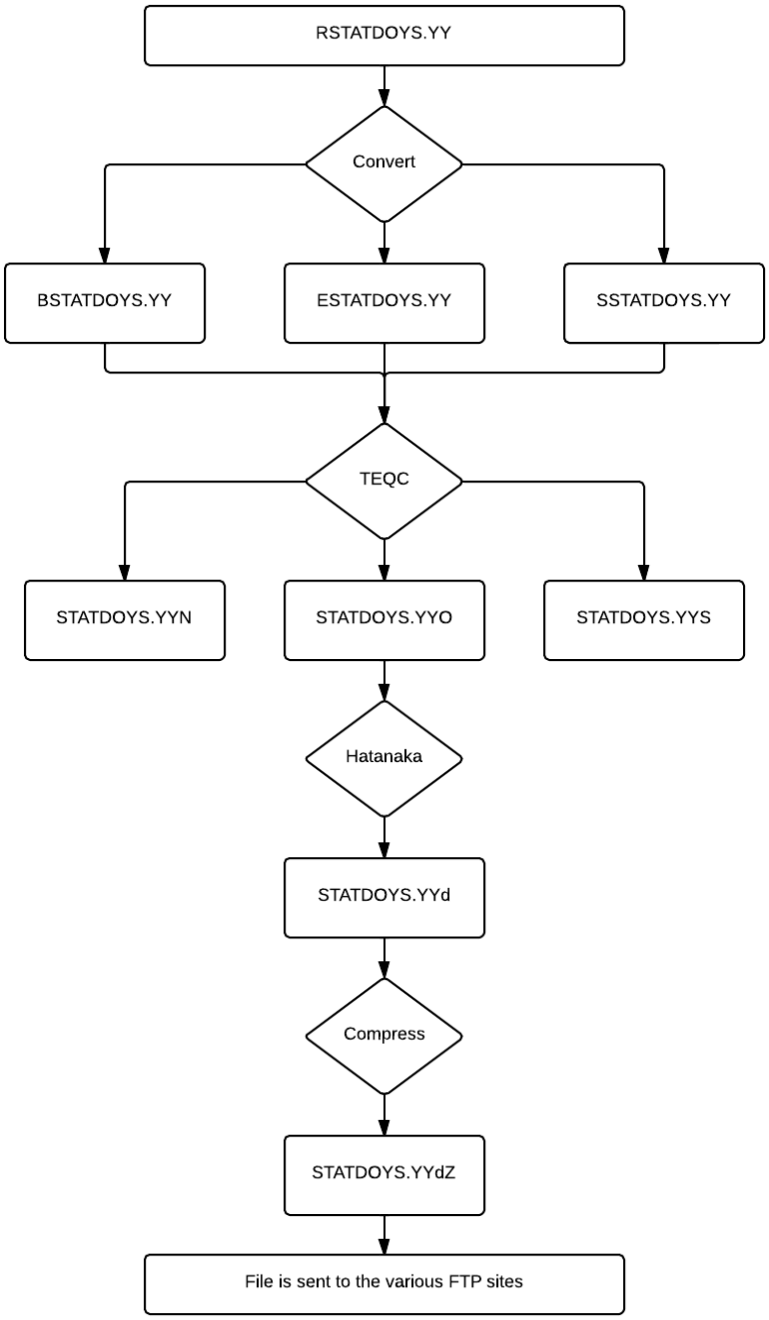

Figure 32. FinnRef server data handling process flow

When the download has finished, software called Convert is used to convert the downloaded Ashtech R-file to B, E, S and possible D-file (weather data file). The R-file is named so, that first letter is R, then station (STAT), the day of the year (DOY) and hour letter depending on the recording hour; letter A has data from 0:00 to 01:00 in UTC (Coordinated Universal Time) time and by following that logic; the letter X has 23:00 to 24:00 UTC. Two last digits are indicating the year.

Program called TEQC (Translation, Editing and Quality Check) is used to check the file quality and to create hourly RINEX-files from B, E, S and D –files of that hour. The created files are; observation file (O), navigation file (N) and summary file (S).

Then the RINEX observation file is compressed with a Hatanaka compression, which adds the letter d to the filename. The Hatanaka file is compressed more with the Linux software Compress, which adds the letter Z to the filename. Then the file is copied to the sending folders and is sent to the FGI's fileserver with FTP (File Transmission Protocol). If needed, the file is also sent to IGS and EPN. At the end of every day, a daily RINEX file is made from the sessions of the day (there should be 24 session files, from letter A to letter X and at least in B, E and S format.

A separate server connects the NTRIP client to the RTCM data stream from station. Then it streams the real-time data forward in NTRIP format to the EPN NTRIP caster in case of EPN stations.

4.4 System Network

The Network consists of various connection types and service providers. The connection type to the stations differs a bit depending on station. Most of the connections are ADSL (Asynchronous Subscriber Line) connections, but some exceptions exist. Since the data connection is done over Internet, it has been secured with Virtual Private Network (VPN), which creates a virtual tunnel between locations.

Remote locations like ROMU, KUUS and KIVE stations are equipped with a wireless technique, which are; WiMAX (Worldwide Interoperability for Microwave Access) connection (at ROMU & KUUS stations) and @450 (KIVE). SODA, TUOR and KEVO stations are using the lines which are provided by the owners of those research premises. Part of the SODA station's connection has been implemented with the Long Range Wi-Fi (Wireless Fidelity).

The rest of the stations are using ADSL connections, except JOEN and METS stations, which have an optical fiber connection. The paid capacity of WiMAX, @450 and ADSL connections varies from 512 kbit/s to 2 Mbit/s download and from 512 kbit/s to 1 Mbit/s upload. JOEN station fiber has 25 Mbit/s download and 10 Mbit/s upload capacity and METS station has 100 Mbit/s on the both directions. The connections through research

premises are not charged by the capacity and since it has been sufficient the capacity amount charged has never been discussed.

As mentioned, the server is connected via virtual serial port to the receivers in the stations. The serial data is packed to Ethernet frame in Nport, so the transmission of data is possible between FGI and stations. Data transmission happens over Internet through VPN connection, as seen in Figure 33.

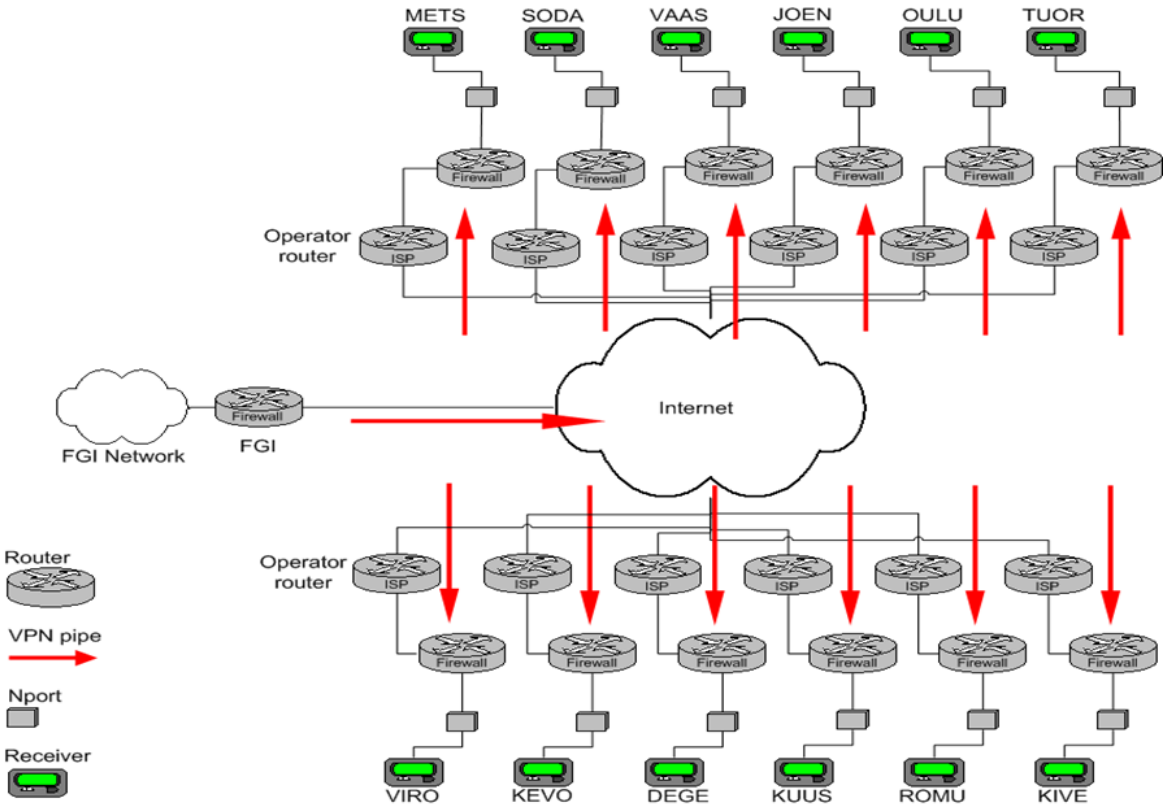

Figure 33. The FinnRef network and illustration of the VPN structure

In this case the Internet Protocol Security (IP-sec) VPN connection is used. IP-sec uses cryptographic tunnelling system, where the connection is first created with Internet Key Exchange (IKE) protocol to recognize each other, after that, the ESP (Encapsulating Security Payload) protocol is used to encrypt and protect the data which is transported in the created connection. Even though the IP-sec increases security, it has known drawbacks caused by the processing time it adds, which causes delay and delay variation to connections. Also the overhead data causes some issues to the connections. (Parmar and Meniya, 2013)

5 Problems of FinnRef Network

Here the problems of the network are discussed, the problems must be processed from the physical level to the application level to find and solve the problems.

The problems in the network have culminated to the loss of GPS data more frequently every year. The various types of data connections together with old GPS equipment have caused the challenges to keep the GPS data consistent. Also the server software and the FinnRef system complexity are making the management and upkeep of the system more and more challenging every year, when the usage of the network has grown and the requirements for the data consistency and quality have increased.

5.1 Problems in Premises

The premises have two major issues considering the overall reliability; the electricity problems and support availability. Storms might cut the power out for long periods of time, which causes that the electricity company is needed to fix the power lines. Storms may also cause that a lightning breaks devices in station and a person is needed to go to the site and change the equipment. Other potential risks are the problems in heating at the winter time and physical intrusion to the station.

Although some of the stations are located in research building, most of them are in the wooden or fiber plastic cottages, which cause some challenges in heating at cold winters. The temperature is tried to be kept near 10 C^{o}, but many times it has lowered to near zero or even few degrees below. The cottage structure causes also a potential risk of an equipment theft or vandalism.

Uninterrupted availability of electricity in certain remote stations has been a big problem. Lately snowy winters and stormy autumns have cut the electricity on those stations several times per year. For example, in KIVE station the remote location at the end of long power line is the major reason for the frequent cut outs and long repair times.

Another very usual problem with electricity in stations has been the lightning storms, which have caused loss of electricity and device breakups several times per summer.

As has been noted, the remote location of most of the stations is causing the outages of stations to extend from days to weeks, especially in very remote places like KIVE and ROMU, the fixing of the problem might take weeks. In most of the cases someone outside FGI must go to the site and see what the problem is and then if needed, ask FGI to send the spare devices. After receiving the devices, the person must travel to the site again to replace it. This is very expensive at the stations where there is no other local contact than the Internet service provider.

5.2 Problems in Devices

All the network devices in the FinnRef system tend to jam occasionally and only fix for that have been that someone goes to the site and makes a hard reboot (cycles the device power off and on again). Some smaller problems are caused by the Zyxel firewalls; the VPN connection jams and prevents the data downloading, but in that case, the firewall is still reachable through the public address to make a soft reboot (choose reboot from browser management menu). The Nports sometimes have problems with jammed server connection, which prevents the data to be downloaded from station. Soft reboot through telnet will clear the problem.

As said, lightning breaks the devices quite often, so it can be counted as a device problem. Also the old FGI firewall used to have stability problems when running all the VPN pipes and the whole institutes Internet traffic at the same time.

The problem with the Ashtech Z12 receivers is that it was designed in early nineties; first problem is it has a low amount of memory, which causes data loss after approximately two days if the connection is lost. Second problem is that it only has serial ports for communications, and even though the serial connection is converted to Ethernet connection, the serial port data needs stable and robust connection to work properly. Some of the problems have a workaround in the virtual serial port driver in the server, but if the connection is not stable enough, the download from station will not work.

5.3 Server, Network and Resource Problems

The server software consists of many different Perl-language scripts, which are used for downloading the GPS data from the stations and for sending forward the GPS data. The scripts have a lot of cross references and parallel actions. This issue along with the fact that the scripts have been written without proper commenting have caused that the interaction of the scripts is not entirely clear. Making changes to the scripts is difficult and troubleshooting is hard.

There are quite a many different network issues, from hardware problems to data reliability issues. First challenge to be mentioned is with the IP-sec VPN, which have caused some problems. One particular issue was the case where the previous FGI firewall did not manage to run 13 VPN tunnels and whole FGI's other network traffic which caused it to crash frequently. Another more common issue has been the IP-sec causing more delay and delay variation to the network and it seems to have caused download interruptions and data corruption on sites where those are values are higher than usual already. Also the conversion from serial data to Ethernet data and vice versa causes delay to the data, which causes some issues to data downloads.

The various types of connections have also generated problems. Especially the most exotic connections, like the Long Range WiFi in SODA have caused challenges together with IP-sec VPN. Also the KIVE's @450 connection has time to time some connectivity issues.

The greatest issue in the network hardware is the ADSL modems, which are prone to get damaged due to lightning storms, even though they are more reliable in other situations than wireless connections. Although some exceptions apply to the good overall reliability, one good example of that is the ADSL connection of VIRO, which has a weak signal quality. This was noticed when a lightning protection was added between the ADSL modem and the phone line. It caused the connection to become so unstable, that it could not be reliably used. When the protection was removed, it worked normally again. This was tested with various lightning protectors and always the result was same.

The human resources have been also a problem for a few years. FinnRef network has been run by two persons, the primary operator and the backup operator. When primary

operator was available, the backup operator did not do operator duties. Most of the information about the system and its special features was only on the primary operator. The backup operator was skilled enough to run the system in its normal state and can deal with the common and recurring problems. But if the system ran to a more special problem and the primary operator was unavailable, the system was unavailable until the primary operator was available again. Also the fact that the both operators had other responsibilities and projects caused conflicts between the projects and their managers. And for the same reason, the time available for the operator duties was limited, which caused the documentation and monitoring to be at poor level. The improvement and maintenance planning has also been at a low level.

6 FinnRef System Requirements

Here the requirements for the network are reviewed, including the constraints generated by the old network along with the requirements generated by the possible new services and features brought by the FinnRef system renewal. An example of that is the new real-time location service which will be offered without a charge. Even though the system is renewed, the old system must run few more years, to get enough parallel data to remove the differences in data caused by the properties of the two different receivers.

6.1 User Requirements

The user requirements were collected with a questionnaire from the FGI's network administration and FinnRef operator, and also from the department of Geodesy and Geodynamics researchers, who are using the GNSS data collected by the system.

Users stated the major problems to be the RMA issues and slowness of processes regarding to that. Also the lack of knowledge is a problem in many different areas, mostly concerning the following things; the GNSS data availability in the server, the amount of data missing per station per day and the possibility to retrieve missing data later. Also in cases where manual processing of the data is needed, the lack of knowledge concerns the amount of time to get the data ready for usage. And if the data is missing and it is needed, the time has been too long to get the data ready for usage.

The fixing of the problems at the stations has occasionally taken too long. It has been culminated especially at the METS station (which is part of two international networks) problems, where a lot of data has been lost, or the poor quality of the data, caused by broken hardware, has caused a loss of reputation.

From operation point of view, in the problem solving situations, usually there is not enough information available fast enough to support with determining the problem. The amount of information is low due to insufficient logging of changes and errors.

The FinnRef network operation should have a team where each member has their own area of responsibility, rather than single person operating the whole system. The team should have experts from the areas of data communication, server operation in the Linux and Windows environments and from the area of usage and calculation of the GNSS data. The team should also have a manager, which together with the team would make well-made and scheduled project plans to improve the system constantly. Those projects should be documented and regular meetings should be held, where the project targets should be evaluated. Most crucial is to have enough personnel and contacts in problem situations to get the situation normalized as fast as possible. This will be emphasized when the new FinnRef real-time location service will become more crucial.

6.2 Application Requirements

The application requirements are generated by the GPS data downloading and streaming applications, telemetric applications like Telnet and SSH, FTP transfers and web browser management of devices. Also as the data the new renovated system produces has been opened for free use, it might need new applications to be introduced to the system.

There are three types of applications used in the original FinnRef system. The first type of applications is the real-time applications; RTCM streams from the receivers to FGI and the NTRIP streams from FGI to outside (for example EPN). The requirement of the streams is that they should be consistent and no connection problems should happen, or the application using the stream loses it from that station and the service becomes (more) unreliable. In the original FinnRef system they were not used very much, but their role is much larger in the new location service offered by the FGI. Thus those new system's real-time streams might also become mission critical application streams.

The second one is the hourly GPS data download, which is mission critical and interactive application. The download will not be disturbed from the loss of connection or connection problems very easily, since the data download will be tried quite a many times at every hour, and if it still fails, the data will eventually be downloaded, when the connection has been restored. So it can be counted as an interactive application. The major issue in hourly data download is with the GPS receivers, since it has very has a small memory, which can hold approximately two days of data, and if the memory

comes full, it will stop storing the GPS data. The application is mission-critical, because without the GPS data downloaded from the station, the mission of FinnRef will not be fulfilled. It counts also as an interactive application.

The other application which is interactive is the FTP file transfers to EPN, IGS and the FGI archive server. The interactive FinnRef maintenance and management applications are representing the third application type; telemetry and command-and-control applications, which more precisely are SSH, Web and Telnet, which are used for the station devices management and control.

The location of the RTCM real-time streams and hourly data download applications are between the stations and the FGI server room. Thus the both of the GPS data applications are using the same connection; it has properties of both connections. The most meaningful are the real-time and mission-critical applications, which are requiring a predictable connection. The telemetry and command-and-control applications are used from the FinnRef operator computer to the stations, and the data from the FGI archive is used by the department of Geodesy and Geodynamics in FGI's second floor Geodesy department.

6.3 Device Requirements

The device constraints and requirements are generated by the old GPS receivers and FGI's head office network infrastructure. Also reliability of the devices is generating requirements.

The generic computing devices in the FinnRef system are the FinnRef operator desktop computer using Windows 7 as its operating system, and almost identical computer(s) in the department of Geodesy and Geodynamics, which are using the processed GPS data in research purposes.

All the servers are virtual; they are installed on the Citrix Xen server environment running on a Hewlett-Packard Blade rack server environment. Most of the servers, like the real-time and FinnRef servers are using virtualized Gentoo Linux, but the data storage/archive server is using Windows 2008 server.

The Ashtech GPS receivers are the only specialized devices in the FinnRef system. The GPS receivers are creating a bottleneck to the system, since they have small memory to store data in and the only connection type is serial connection, which is slow and incompatible with the Ethernet network without converters. When the new receiver was chosen, one criterion was that it must support straight Ethernet connections and have enough memory to tolerate longer network outages. But since the old receivers must be used few years, to get parallel data to remove the differences in data caused by the properties of the receivers, they define the base for the rest of the system to build on.

Since the over voltage problems are breaking up devices, the new devices should be equipped with better over voltage protection, or the devices should be protected with an external over voltage protection.

The receivers are located at the stations, the servers are in the FGI's server room, operator's desktop computer at IT staff premises and the department of Geodesy and Geodynamics is at second floor of FGI with their desktop computers.

6.4 Network and Security Requirements

The network requirements are generated by the future needs and the RMA issues together with the security threats. Reliability issues like the device breakups and the power and network outages are occurring quite often. The data loss they cause should be minimized or prevented if possible.

Maintainability must be increased, the need for station visits where device is just restarted should be prevented and the device breakups should be handled faster. The network needs to handle more users, due to new services in the system. The amount of users might increase radically in short period if applications which use the new data will come to wider use.

Since there are now new GNSS receivers in use, the data sizes has grown due to increased sampling rate and because the new receiver collects data from a several different GNSS systems at the same time. This fact must be taken into account in renewing of connections. Also the delays and other problems caused by the serial to Ethernet

conversion and IP-sec VPN packing and unpacking would be desirable to be minimized or removed.

Security requirements are defined with the help of the risk assessment matrix seen in Table 12. There can be seen that the greatest risks lie in unauthorized access, theft, denial of service, physical damage and corruption.

Table 12. FinnRef risk assessment

Effect/Probability	Receivers	FinnRef Server	Data Servers	Network Elements	Services	Data
Unauthorized Access	B/C	C/C	C/C	C/C	C/C	C/C
Unauthorized Disclosure	C/C	C/C	C/C	C/C	C/C	C/C
Denial of Service	D/D	B/C	B/C	B/B	B/B	D/D
Theft	A/C	A/D	A/D	A/C	D/D	D/C
Corruption	C/C	A/B	A/B	A/C	D/D	A/B
Viruses	D/D	B/C	B/C	B/C	D/D	B/C
Physical Damage	A/B	A/C	A/C	A/B	D/D	D/D

Effect:		Probability:	
A: Destructive	C: Distruptive	A: Certain	C: Unlikely
B: Disabling	D: No Impact	B: Likely	D: Impossible

As can be seen, the service disturbance caused by the security issues (physical and network) are possible, thus the system should be well protected from them.

6.5 Supplemental Requirements and Requirements Specification

The supplemental requirements are generated by the operation of the system and the system overall confidence. Also the requirement specification is described.

The operation and maintenance of the system should be well planned and done with the personnel and operators, which are trained to handle the system's problem scenarios. The maintenance actions and error situations should be accurately logged and the manuals and procedures kept up to date.

The system confidence can be derived partly from the hourly data; the amount of acquired GPS data epoch per year tells how well the network has succeeded in its primary mission. This does not tell the actual availability, since the network might have been down and the data was downloaded afterwards. But as mentioned, it gives a good picture about how well the primary mission of the system was fulfilled and the confidence level of the network. As can be seen in Table 13, the system's total average confidence to deliver GPS data epochs has been less than 90 percent for recent years, when it should be as near 100 percent as possible. Especially long blackouts on some of the stations have caused their individual percent to be undesirably low.

Table 13. The acquired GPS data epochs in the years 2011 and 2012 per stations

YEAR: 2011				**YEAR: 2012**			
	EPOCHS/ YEAR	**PERCENTS/ YEAR**	**OPERATION DAYS / YEAR**		**EPOCHS/ YEAR**	**PERCENTS/ YEAR**	**OPERATION DAYS / YEAR**
METS	944263	89,86 %	327,98	**METS**	88719	99,41 %	362,83
SODA	919919	87,54 %	319,53	**SODA**	718012	68,33 %	294,40
VAAS	918087	87,37 %	318,89	**VAAS**	988450	94,06 %	343,33
JOEN	947670	90,18 %	329,17	**JOEN**	1049617	99,88 %	364,58
KEVO	49082	80,80 %	294,92	**KEVO**	70402	78,88 %	287,92
KUUS	929465	88,45 %	322,84	**KUUS**	76771	86,02 %	313,97
KIVE	844154	80,33 %	293,21	**KIVE**	788213	75,01 %	273,78
DEGE	937572	89,22 %	325,66	**DEGE**	813788	77,44 %	282,66
OLKI	960429	91,40 %	333,6	**OLKI**	970898	92,39 %	337,23
OULU	839826	79,92 %	291,71	**OULU**	883325	84,06 %	306,82
TUOR	950646	90,47 %	330,2	**TUOR**	1029633	97,98 %	357,64
VIRO	914591	87,03 %	317,68	**VIRO**	1012268	96,33 %	351,60
ROMU	847299	80,63 %	294,3	**ROMU**	1045923	99,53 %	363,29
TOTAL AVG	***846385***	***86,40 %***	***315,36***	***TOTAL AVG***	***733540***	***88,41 %***	***322,70***
MAX:	1050835	100,00 %	365	MAX:	1050835	100,00 %	365

In Table 14 there are the FinnRef's initial conditions of the requirements specification, where the starting point for the project is defined. It describes the outlines of the project and defines the problems and goals.

Table 14. FinnRef requirements specification's initial conditions

Requirements specification	
Section 1: Initial conditions	
Project type	Upgrading of the FinnRef network
Project scope	FGI outbound connections and 13 GPS stations
Project goals	Improve the availability and integrity of the network
Other conditions	None
Problem evaluation and definition	Reliability issues have caused loss of GPS data, also the upcoming renovation of the system causes new challenges to network connections

And in Table 15 can be seen the FinnRef requirements specification. It briefly describes the requirements and their types acquiring method and the locations the requirement is concerning.

Table 15. FinnRef requirements specification

Requirements Specification				
ID/Name	Type	Description	Gathered/ Derived	Locations
1	user/device/network	Increase reliablity	Gathered from all	stations
2	user	Increase data availabiltity	Gathered from users	all
3	user	Human resource issues causing poor maintainability	Gathered from users & operator	FGI
4	application	Predictable connections from FGI to stations	Derived	stations/FGI
5	device	Constrains from old GPS receiver	Derived	stations
6	network	Better maintainability	Derived	stations
7	network/user	Possible increase in user amounts	Gathered from management	all
8	network	Increase capacity (more GNSS data)	Gathered from management	all
9	network	Increase physical security	Derived	all
10	other	Change/maintenance logging	Derived	all

In this case the requirement specification is not as comprehensive as the theory suggests, since the network is quite simple and the project scope was clear. Also the fact that there were enough resources to apply all the improvements at least on some level makes some parts of the suggested requirement specification unneeded.

7 FinnRef Data Flows

The data flows of the original FinnRef system are reviewed in this chapter, together with the developed service metrics and the user and application behaviour. This information is valuable from the viewpoint of network management.

7.1 Service Metrics

The service metrics were defined with the Ping application from a few different measurement points; straight from the server to the Nports of stations (FinnRef service metrics), from the monitoring computer to the public address of the stations (WAN & FGI LAN service metrics) and from the monitoring computer to the FGI router (FGI LAN service metrics). In Table 16 the results of the Ping measurements made from the FinnRef server to most of the stations' Nports can be seen. Some of the stations are missing from the results, since at the time of measurement those connections were already upgraded to the new MPLS connections and as mentioned; these service metrics are concerning only the original FinnRef system.

Table 16. Ping measurements between the FinnRef server and the selected stations' Nports

Station	packets sent	packets received	Delay Min (ms)	Delay Avg (ms)	Delay Max	Delay jitter (ms)	Availability (%)	time (min)
SODA	8905	8870	20.118	22.687	277.008	5.192	99,6070 %	149
VAAS	5203	5203	19.497	21.704	102.896	5.063	100,0000 %	87
JOEN	5619	5619	16.677	17.113	32.394	0.415	100,0000 %	94
OULU	5263	5263	26.520	28.984	160.090	4.895	100,0000 %	88
TUOR	4833	4833	26.782	27.697	51.742	1.016	100,0000 %	81
KEVO	7010	7009	35.598	37.325	133.775	6.941	99,9857 %	117
DEGE	3420	3420	15.594	17.306	58.561	4.611	100,0000 %	57
OLKI	5851	5851	36.345	37.744	90.903	2.506	100,0000 %	98
KUUS	3969	3951	34.800	51.932	126.398	8.092	99,5465 %	66
ROMU	7311	7289	44.688	74.473	134.726	11.841	99,6991 %	122

A monitoring software or script should be set up to Ping the stations and send alarms and warnings for the crossing of the service metrics limits and thresholds. The monitor software should also collect and save the daily Ping results as daily logs. The used limits and results should be near the ones seen in Table 17; from the server to Nport, the warning threshold for the packet loss should be in 1.2 percent and alarm in 3 per-

cent, and in the delay, the warning threshold for average delay should be somewhere between 50 millisecond to 75 milliseconds and the alarm limit in 150 milliseconds, and in the delay variation (in Ping it is called mdev), the warning threshold somewhere between 10 milliseconds to 30 milliseconds and the alarm limit in 50 to 75 milliseconds. Those thresholds and limits will become more accurate, when they have been in use for a while, and some problems have been solved using the Ping logs as part of the solving process. The service metrics should be used as support in the maintenance (warning thresholds and alarm limits) and also for the quality control of the service, which can be used in in the situations, where LAN service metrics shows no packet loss, but WAN & LAN service metrics are having packet loss.

Table 17. The service metrics for the original FinnRef system

FinnRef Service Metrics		
Type	Warning Threshold	Alarm limit
Packet loss	1.2%	3 %
Delay	50 - 75 ms	150 ms
Delay variation	10 - 30 ms	50 -75 ms

The service metrics might also need some tuning for the new FinnRef system, but as the new system is more sophisticated and has better error correcting, these service metrics are quite certain to fulfil its needs.

7.2 User and Application Behaviour in FinnRef

The data from the stations is downloaded at the beginning of every hour. Downloads are grouped to three groups of two to four stations, which are downloaded in intervals of three minutes. If the download process fails, it is retried four times with intervals of 10, 60, 200 and 600 seconds.

A single download takes one to three minutes to complete. That causes relatively high peak in the network traffic for the first five to fifteen minutes, even though the used capacity is not at a high level due to small data sizes. The monitoring and maintenance of the station devices is done from the FGI to the stations through the same connection.

7.3 Data Flows in FinnRef

The system has three types of streams; from the GNSS stations to the FGI FinnRef server (hourly & real-time data), which has predictable performance requirements due its mission criticality. The second flow is from the FinnRef operator's desktop computer to GNSS station devices (controlling devices), and the third is from the FGI to the data centres and the FGI archive (FTP).

GNSS receivers at the stations can be counted as data sources. Also the FinnRef server is data source as it produces RINEX data, but it is also data sink for the GNSS data from the stations. Other data sinks are the FGI's FTP server and other FTP sites where the data is transmitted from FinnRef server.

In Figure 34 the flows between the station and the FinnRef server can be seen. It can be described with single performance profile which is called P1.

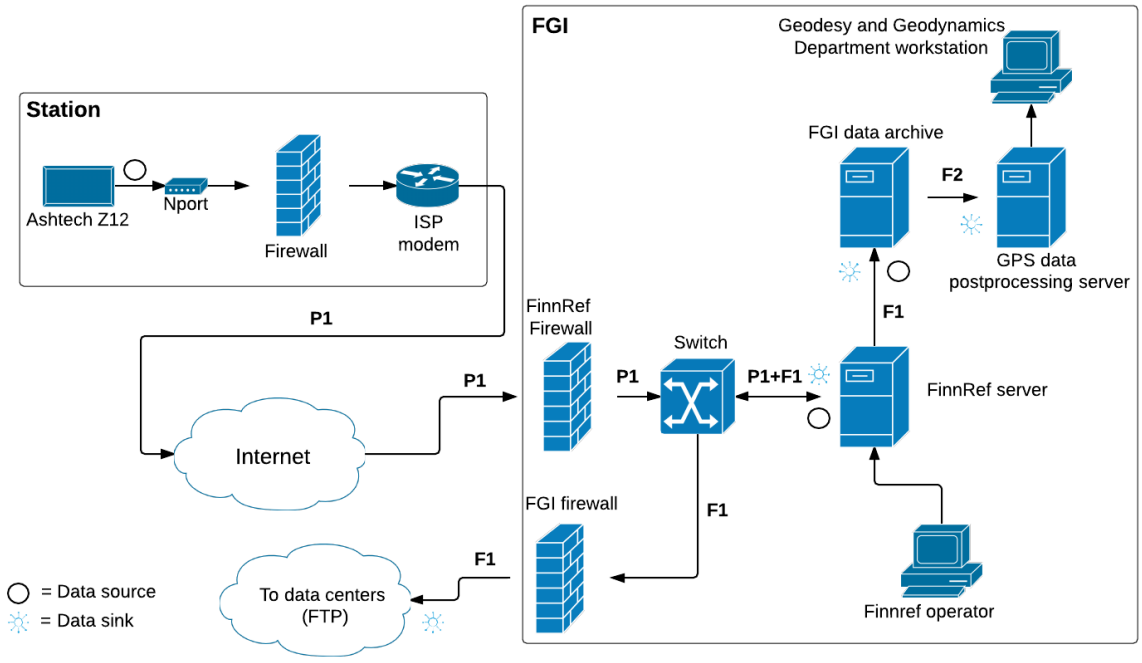

Figure 34. The FinnRef data flows from any one station's viewpoint.

The flow P1 has combined (two-part) flowspec, consisting the FinnRef hourly data and real-time application flows along with telemetry and command-and-control flows. The performance requirements for that flow are in line with the service metrics; the delay

shall not be over 150 milliseconds, the nominal capacity is at least 2 Mbit/s and the availability should be over 97%.

The flow F1 is a best effort FTP transfer flow. The flow F2 is the data flow between the GPS data processing unit and the FGI data archive; this flow has higher capacity requirements, since large amounts of GPS data is transferred from the archive to the processing server. The flow F2 should have an average capacity near 50 Mbit/s to keep the wait times tolerable.

7.4 Flow Prioritization

The most important data flow is from the GNSS stations to the FinnRef server which contains the hourly GNSS data. Secondarily important flows are the controlling flows to the GNSS station devices, and thirdly important are the flows of the hourly data from the FGI server to the data centres (only with the EPN/IGS stations). At the moment, the real-time data from GNSS stations are the least important, but the situation is very likely to change, when the usage of the new network's real-time services will increase and come more important.

8 The Renovation of the FinnRef System and Results of Thesis

This chapter discusses the results of the thesis and introduces the improvements already done for the FinnRef system and network. The overall renewal of the FinnRef included station reliability improvements, new stations and new GNSS receivers for all stations. Also a new server has been installed at the FGI with a new data processing centre. FinnRef improvement also concerned the core network; this chapter also discusses about the new network type. The network also got new stations to make the GNSS network denser. Old Ashtech receivers on the stations and the old server will be used for few years to get parallel data with the new ones, so the difference in the data caused by the hardware can be removed in the post calculation.

8.1 New Stations

New stations were established to make the GNSS network denser. FinnRef also begun to offer new open positioning service to offer real-time DGNSS (Differential GNSS) service, which is providing positioning corrections based on the error modelling of the code observation at the FinnRef stations. The system needs quite dense network, since the denser the network is, the more accurate positioning information it will offer, because the accuracy of the correction decreases when the distance to the nearest station increases. The new FinnRef station map can be seen in Figure 35; it shows the new stations along the old stations which have the new receivers. (FGI, n.d.; Koivula and Poutanen, 2014)

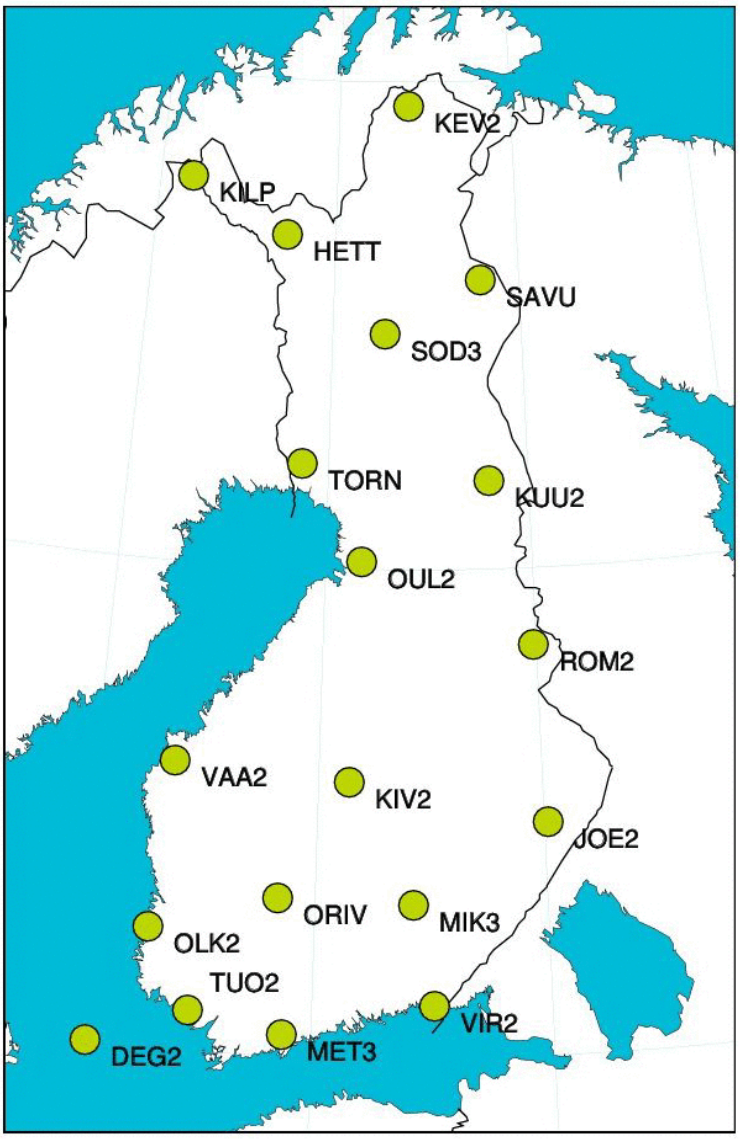

Figure 35. The new Finnref network, including the new stations and the old stations with new receiver. (FGI, n.d.)

Six new stations were built all around Finland; those new sites are Hetta (HETT), Kilpisjärvi (KILP), Mikkeli (MIK3), Orivesi (ORIV), Tornio (TORN) and Savukoski (SAVU). (FGI, n.d.)

8.2 New Receivers and Antennas

The Javad Delta 3GT receiver (see figure 36) is the new receiver used in the FinnRef stations. The receiver tracks all the current GNSS systems; GPS, Russian GLONASS, European Galileo and Chinese BDS signals, although the BDS option from the receiver

has not been opened yet, but will be, when need for it comes. Also the SBAS (Satellite Based Augmentation System) signals of EGNOS, WAAS and MSAT can be tracked. (Koivula & al., 2012)

Figure 36. Javad Delta 3GT (Koivula, 2013)

The receiver has an Ethernet port and 4096 MB of memory to hold data, if the data connections are not available. Since the new receivers use TCP/IP connections directly, the error correction of data is already built into the TCP protocol and the application relies on it. This overcomes the problems caused by low quality connections, from which the virtual serial data connections of the old receivers are suffering. Also the large memory overcomes the problem where the data connections are lost for long periods; the memory can hold approximately one year data, depending of course on the sampling interval and how many different GNSS system's data is collected.

In figure 37 the TORN station cottage can be seen together with its new antenna and its mast. The new antennas are also choke-ring type of antennas equipped with pre-amplifier, but they can receive all GNSS signals available at the moment.

Figure 37. TORN station in summer 2013, showing the antenna mast and the station cottage. (Finnish Geodetic Institute, n.d.)

The antennas were calibrated at GEO++ GmbH in Germany. Most of the new antennas are on top of a three meter antenna mast, but some stations have a six meter mast due to high obstacles nearby the station.

8.3 Station Improvements

Temperature problems now have a solution, which has been under testing in KIVE station; the equipment and a 500 watt heat radiator is surrounded by the insulation plates, which can be opened easily for the device maintenance. The solution seemed to be working, so it will be built to the other stations with similar problems for the next winter.

The local electricity grounding at the stations was improved to prevent the device breakages due voltage spikes caused by the lightning and other transient voltages. Also the incoming electricity was reinforced to sustain over current and voltage caused by the lightning. Installed protector is a two-step over voltage protection with a coarse and fine over voltage filtering and it is manufactured by Dehn and the model is Ventil. Coarse filtering removes the high voltage peaks, but little bit slower than the fine filter,

which takes the fast but lower voltage peaks. Both improvements were installed all of the stations where it was possible. These actions should protect the equipment from the lightning strikes quite well in the future.

Also an Uninterruptible Power Supply (UPS) was installed to keep the system running in a case where the main power is lost. The chosen UPS is manufactured by APC, and the model is Smart-UPS X 750VA, which keeps the system running for almost 24 hours if the main power is not available.

A remote boot system was installed to all of the stations; it was designed together with the Finnish company Ouman Oy, which was chosen as a vendor for the devices. The remote booting is based on the SMS (Short Message Service) messaging system, which is for the controlling the 12 volt and 230 volt electricity outlets at the stations. Each of the outlets can be controlled separately to force restart each device separately. The Ouman remote boot system has also a SMS alarming system for the main power loss and for the UPS power loss.

8.4 Core Network Renovation

The core network is based on the MPLS technique and is operated by Sonera and the product name of the connection is Sonera DataNet. It will be implemented over various connection types; fiber optics, ADSL and 3G. At the moment the DataNet connections are already installed to a few of the older stations and to all of the new stations; all the new stations and KIVE are using the 3G connections, since those places did not have the possibility for a wired connection (or it was too expensive to implement). METS station is using DataNet over fiber optic connection, because the area is in wider research use. These connections will be installed to the other older FinnRef stations in the near future.

The Sonera DataNet was chosen, because the MPLS connection has fewer routers between the FGI and stations than the Internet based network. As the MPLS data packet switching at routers is based on labels in the packet headers, it does not open the packet itself. This leads to decreased processing times at the routers, which will decrease the delays in the wired connections. It also increases security, as the data packets are not opened. As MPLS is also a closed network between routers, the data

will not go through the public Internet at any point. These things improve reliability and security greatly in the network. (Cisco Systems Inc., 2002)

Figure 38 illustrates the original proposed network configuration for the FinnRef renovation and a sketch of data flows.

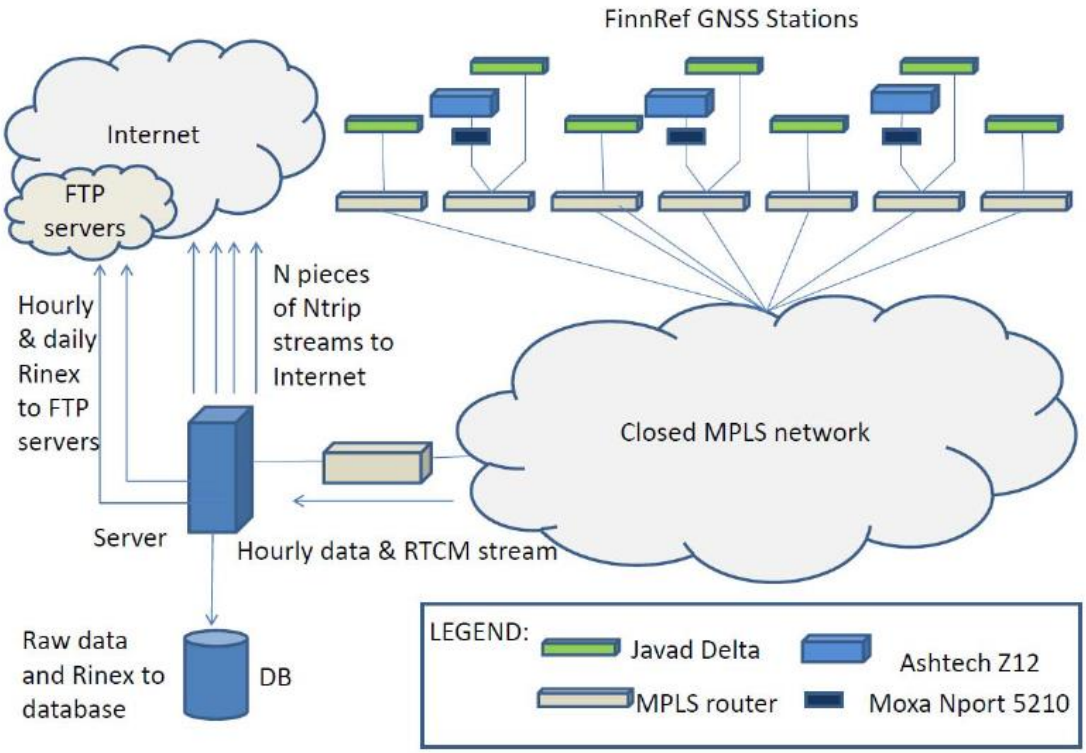

Figure 38. Proposed configuration of FinnRef by the end of 2013. (Koivula & al., 2012)

It defines the devices on the new and old stations and the devices at the FGI side, which include the MPLS router and the server along with the data archive.

8.5 New Server and Server Software

The new FinnRef server is a cluster server, which has a redundancy and/or test system running on the same cluster, but on different hardware. The server hardware is based on the Hewlett-Packard's (HP) BL460c Gen8 Blade hardware, which has two eight core processors running at a 2.6 GHz frequency with 32 GB memory per processor. Also one extra disk system and data recorder are installed for data backups. The virtual op-

erating system environment is based on Citrix Xen-server virtual environment, which has the production and test system are installed on. (Koivula, 2014a)
The disk system used in the server is HP 3PAR StoreServ 7400 2-N Storage Base system. The system has two types of HDD (Hard Disk Drive) disks; faster low capacity disks, and slower high capacity disks. The server software is running on disks with 300GB of capacity, and their operating speed is at 15 thousand revolutions per minute. The data from the stations is stored on the data disks which have the capacity of 3 TB and 7.2 thousand RPM operating speed. (Aarni, 2014)

The production system consists of three virtual Windows Server 2012's. One of those is used to acquire the real-time data and to make the real-time data calculation and error modelling for the location service. The second one acts as an interface to the users of the location service, it collects the data requests and shares the data forward. Third one is used for the research purposes and will host services which are generated by the research, and this server uses a duplicate of the first server's data. (Koivula, 2014a)

The test system is identical; it has also three virtual Windows Server 2012's running on it with same properties on each. It is used to test new features and parameters. It also can be used for training to prepare for the problem situations by creating expected scenarios. All modifications to the system are first tested on the test servers before they will be implemented to the production environment. (Koivula, 2014a)

The server software was provided by the German company Geo++ GmbH. The software is called GNSMART and as Geo++ it defines; it is based on n the procedures known as GNSS-SMART (Global Navigation Satellite System - State Monitoring And Representation Technique). The software monitors the FinnRef GNSS network state to calculate and remove the errors from the GNSS data caused by the error sources mentioned in Chapter 2.2 to provide the position to the user with a high accuracy, which is at centimetre level. The position can be provided in real-time and by post-processing. The system supports RTCM and RINEX formats. The GNSS network with more than five stations already provides some level redundancy in the GNSMART system, so the service availability and reliability along with the position accuracy are at high level. (Geo++ GbmH, n.d.)

8.6 FinnRef Services

Even though the renovation has completed and the new DGNSS location service has been opened, the system is still in its initial operational phase. The renovation was planned and executed having the full production use of the location service in mind, but the limited resources and the current low level of users has still kept the system mainly in research use.

The current maximum user amount was estimated to be approximately 200, which is set by the data disks of the current server disk system. If the discussions with other governmental institutes lead to wider usage of the location service, the system needs to be upgraded to support that. (Aarni, 2014)

8.7 Resources and Managing

More resources have been allocated for the project and the FinnRef operator has been changed, even though the responsibilities between the six members of the team are not yet clearly defined, some level definition already exists; the roles of project manager and operator are defined are quite clear.

At the moment the rest of the team are developing and testing the new system and some are doing some maintenance actions when needed. Clear short term and long term goals have been defined in weekly project meetings, where also the current situation is discussed.

8.8 Overall Improved Availability and Confidence

The confidence level has increased as the renovation has proceeded. As can be seen in Table 18; in the year 2013, the total average percentage of acquired GPS data epochs from old stations using Ashtech receiver has increased, even though the SODA and KIVE stations are preserving some of their issues. The situation on those stations should improve when the results from the new connections which are or will be installed will become visible in yearly acquired GPS data epochs percentages.

Table 18. Acquired GPS data epochs from the Ashtech receivers in year 2013 per stations

YEAR: 2013			
	EPOCHS/ YEAR	PERCENTS/ YEAR	OPERATION DAYS / YEAR
METS	1011140	96,22 %	351,21
SODA	836212	79,58 %	290,45
VAAS	1043771	99,33 %	362,55
JOEN	1048114	99,74 %	364,05
KEVO	1013469	96,44 %	352,02
KUUS	940988	89,55 %	326,85
KIVE	788213	75,01 %	273,78
DEGE	1040352	99,00 %	361,36
OLKI	1048620	99,79 %	364,23
OULU	925939	88,11 %	321,62
TUOR	1049486	99,87 %	364,53
VIRO	974487	92,73 %	338,48
ROMU	1032531	98,26 %	358,64
TOTAL AVG	***981025***	***93,36 %***	***340,75***
MAX:	1050835	100,00 %	365

As can be seen in Figure 39, the total average percentage of data acquired has increased year by year, so the overall reliability and confidence of the system has increased due to actions taken to improve it.

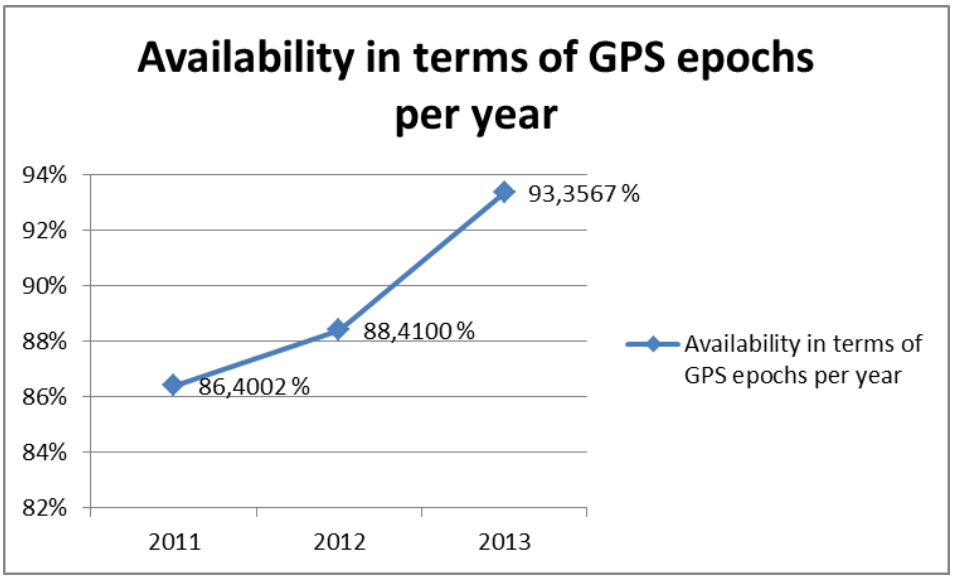

Figure 39. Availability of FinnRef in terms of acquired GPS data epochs between the years 2011 and 2013

The increase can be especially seen between the years 2012 and 2013, due to quite a many of the improvements mentioned earlier being installed during the last quarter of 2012 and the spring of 2013. FinnRef's reliability on year the 2014 should be increasing over the level it was at in 2013, as it will be the first whole year with all of the improvements installed.

9 Discussion and Conclusions

This thesis provides the answer to the research question “What can be done to improve reliability of FinnRef network?”, achieved through the network analysis. The used network analysis theory material showed, that the network renovation is more than just values and their simple improvements; if the network analysis is done well, it requires a lot of conversation with different kinds of people and studying of the whole system, including the users, managers, applications, work habits, devices and especially their requirements. This approach makes the solving of problems easier and it also will prevent many of the possible upcoming problems or at least the system is more prepared to handle them.

The network analysis theory part in this thesis is quite extensive, but it is meant to be informative enough, so it would help the further development of the FinnRef system and it could be used as a base for other network projects. To get the best results in the FinnRef network, this process should be repeated quite frequently in the upcoming years, since the whole system is still in initial operational phase, where the user amounts, connection types, data flows and such are most probably about to change. The iterative analysis process can be done by using only parts of this thesis as long as the big picture stays in mind.

The thesis process was slow and challenging due to constant changes in the system caused by the FinnRef system renovation and its some radical network requirement changes, like the introduction of the location service, which was not in the original renovation plan. That caused the thesis to grow together with the FinnRef renovation process, and many of the ideas for the thesis were put in to the practice almost right after inventing them. As the renovation project started together with the thesis, it made the suitable theory material finding very hard, since it needed to be flexible to be applied to a living project, and still wide enough to cover the whole system. The chosen approach was found to be good for this case, and as it can be executed on any network, the reuse value of this thesis is high.

The system and network analysis output includes the definition of the requirements, the original FinnRef system data flows, the service metrics, the set of recommendations (in Chapter 9.1) and more reliable network (as seen in Chapter 8.8). These combined form

the answer for the research question. The reliability and availability of the network will keep increasing when implementing the rest of the suggested improvements discussed in chapter 9.1 Even though the thesis process was difficult, the outcome should be very usable and hopefully the rest of the improvements will be implemented soon.

The thesis was reviewed by the FGI's IT Manager. He said that the thesis included quite a lot of important information which they did not previously have in a written form. He was also satisfied at the results, thanks to the increased reliability. (Aarni, 2014)

He noted that the theory part was a bit too extensive, even though it had a good point and a lot of useful information. His opinion was, that the set of recommendations was very good, and those improvements will be executed as soon as it is possible, even though some of them will need more resources, which a are bit tight at the moment. This situation needs to change, if governmental services start to use the FinnRef location services. In that situation additional governmental funding would be needed to preserve service quality. He concluded that the thesis was overall very good. (Aarni, 2014)

9.1 Set of Recommendations

Here the set of recommendations is discussed. It is defined by the requirements which were not fulfilled, at all or well enough, on the renovation process. The set of recommendations include the hardware, monitoring, resourcing and management improvements. The security issues have also been noted.

The connections should be improved further; even though the core network was changed and it increased reliability, it also caused some new challenges, since some of the existing connections were or will be changed to 3G connections and all of the new stations are using it. As can be seen in Table 19, that causes quite radical delay and delay variation increase, and the connection is more prone to weather changes than landline connection. The positive side in 3G is that since it does not have wired connection has better protection from over voltages caused lightning strikes. The best option would be to install fiber optic connections to every station, but it would be major financial issue, and it would need additional funding to be acquired. The additional funding might be possible, if the network would act in a bigger role in Finland’s core infrastructure in the future.

Table 19. Ping measurements between the old FinnRef server and the receivers of the new stations

Station	packets sent	packets received	Delay Min (ms)	Delay Avg (ms)	Delay Max (ms)	Delay jitter (ms)	Availability (%)	time (min)
HETT	34892	34887	43.117	70.730	9791.729	92.377	99,9857 %	582
KILP	34895	34888	46.064	72.465	10229.164	103.405	99,9799 %	583
MIKK	18539	18537	46.379	115.062	6328.149	201.979	99,9892 %	582
ORIV	24939	24744	32.961	64.210	2561.160	51.623	99,2181 %	416
SAVU	24921	24921	46.212	67.327	514.411	12.608	100,0000 %	416
TORN	18647	18647	53.253	73.126	525.578	13.063	100,0000 %	311

All ADSL connections, and also fiber connections where applicable, should be implemented with the redundancy connections, which would be using 3G connection to secure the data connection when the line problems occur.

The monitoring of the system should be improved. SNMP option should be enabled on all possible devices to gather the management and error data from the devices to get statistics from the maintenance actions and problems. With that data it is possible predict and point out the problems and plan improvements to increase the overall reliability and availability further. Also the monitoring computer with the thresholds and limits introduced in Chapter 7.1 should be installed. It collects the Ping results and would send alarm messages when alarm limit or warning threshold on the delay or packet loss would be exceeded. The results should be stored for few months, because those results would help in the problem solving situations.

The serial to Ethernet communication in the old stations should be improved; a new better protected model of used Moxa Nport is available, model 5210A, where the surge protection has been added to each port to protect it from transient voltages from lightings and electric network. The manufacturer also promises that the power consumption is 50% lower than in any other equivalent system in the markets. The manufacturer also states that MTBF has increased to 847,750 hours from the old model's 134,850 hours. (Moxa Inc., 2012b)

The lower power consumption will matter in a case of main power outage at the station and the system would be running on the UPS power. Every watt saved on the power consumption will increase the time system can survive without the main power. It also would have a major effect on the overall reliability, since quite often the lightings broke only the Nport device. Also the protection on the serial port will keep the data connec-

tion more robust and will protect the device from transient voltage peaks from the Ashtech side. The change should be easy, since the server drivers are same as in the current 5210 model.

The security of the stations should be increased with the alarm and camera system, which will help in preventing some of intrusions. Also possibility of contracting the regular security monitoring and visits should be considered.

The data disks for the server should be changed to Solid State Drives (SSD) if the amount of simultaneous users is going to exceed 150 As the system is using now traditional hard drive disks (HDD), the amount of Input/output (I/O) Operations Per Second (IOPS) the disk can handle is limited by the mechanical speed of the actuator and disk spindle. The type of task also has an effect to the IOPS, large sequential data is faster to read than small data pieces randomly. This creates a bottleneck into the system, as multiple users are simultaneously reading and writing data from and to the system, which causes congestion for the drives. (Hewlett-Packard Development Company, 2011)

This would not be the case with the SSD, which is based on NAND flash memory which has no mechanical limitations. It can handle random read operations over 100 times faster than the HDD drives, and the sequential read speeds are also tripled in high performance SSD drives. Drawback to the SSD is that their durability is bit lower due to the NAND memory limits. (Hewlett-Packard Development Company, 2011)

The maintenance processes should be developed further; especially in the situations where a part of the system is having a critical problem, which is causing major loss of data or poor data quality, the situation must be escalated and more resources allocated to solve the situation. This concerns especially EPN or IGS stations and the server side. The escalation process should be well defined and documented and the resources and responsibilities clearly allocated. Other recurring processes should also be documented, from the device changes and spare part ordering to handling of the spare parts and broken devices. Also all maintenance actions should be documented with the accurate times and reasons. When all recurring processes have a well-defined process description and the maintenance actions are logged, it is possible to predict the maintenance actions and thus it is possible to prevent some issues with pre-

maintenance actions, or at least, on the unpredictable situations, the resources would be usable immediately. This will increase the availability of the system.

As mentioned, the resources of the FinnRef operation should still be more precisely allocated, since the new network along with the new location service will need attention as well as the old FinnRef. This will be emphasized when the location service will get users, which are demanding some level of predictability from it. Since FinnRef operator acts a major role in the reliability of the FinnRef, it emphasizes that the load level of the FinnRef operator, or other equivalent team member, should be observed, and if it seems to be too high, immediate actions to distribute some of the functions to other team members must be taken to balance the situation. Then the reason for high load level should be located and eliminated as well as possible. This will improve the overall reliability and will keep the motivation level high in the whole team.

9.2 Future Plans

In the future, the FinnRef network will most probably have a major role in the field of accurate location data in Finland. As the renewal is now complete, and the system is in its initial operational state, along with the fact that the data streams have been released to the public use will lead to the situation where the data will be used by the location based software developed by various companies and also in some point in the future most of the governmental services should use it for various purposes. That will cause a highly increased demand for reliability and availability of the network, and possible Service Level Agreements to be made to ensure the uninterruptable data streams to the users and clients.

There has also been discussion about the founding of the lower level stations on top of governmental buildings all around Finland. These lower level stations' antenna position would not be as accurately defined as in higher level stations, but they would anyway increase the system reliability, availability and accuracy.

These facts together with the increased data amounts due to the addition of the new GNSS systems and new stations, together with the fact that the FinnRef data is open for free to use, will be one of the greatest challenges for the operating and maintaining the network and network service at the satisfying level.

One of the issues is, and will be, the remote locations of some of the stations. Those which will are using 3G connections now, most probably will need a connection upgrade if the amount of users/data will be greater than expected. If the 4G connections built in the future will cover those areas, the problem is solved, but if not, those sites will need big investments on the fiber or landline connections.

References

Ashtech Inc. (n.d.) Ashtech Z-12™ GPS Receiver: Full GPS Capability with Anti-spoofing Turned On, [Online]. Available: http://www.cgg.com/data//1/rec_docs/2478_z12.pdf [Accessed Jan 20, 2013]

Aarni, P. (2014) *Master Thesis comments.* [Interview]. May 2, 2014.

Chassagne, O. (2012) `One-Centimeter Accuracy with PPP´, Inside GNSS Magazine, vol. 7, no. 2, March/April, pp. 49-54, Available: http://www.insidegnss.com/node/2977 [Accessed Jan 22, 2013]

China Satellite Navigation Office (2013) *Report on the Development of BeiDou Navigation Satellite System (version 2.2)*, [Online]. Available: http://www.beidou.gov.cn/attach/2013/12/26/20131226fed336adf2184d52843d5bf81832e82c.pdf [Accessed Oct 12, 2013]

Cisco Systems Inc. (2002) *MPLS Concepts*, [Online]. Available: https://www.racf.bnl.gov/Facility/TechnologyMeeting/Archive/06-30-04-CISCO/CISCO-MPLS-Concept.pdf

European Space Agency (ESA) (2013) Galileo fixes Europe's position in history, March 12, [Online]. Available: http://www.esa.int/Our_Activities/Navigation/Galileo_fixes_Europe_s_position_in_history [Acessed March 8, 2014]

Finnish Geodetic Institute (2012) Annual Report 2010-2011, [Online]. Available: http://www.fgi.fi/fgi/sites/default/files/publications/annual_reports/FGI-annual-report-2012-Lowres.pdf [Accessed Mar 25, 2013]

Finnish Geodetic Institute (n.d.) *Positioning-service*, [Online]. Available: http://euref-fin.fgi.fi/fgi/en/positioning-service/ [Accessed Oct 7, 2013]

Geo++ GmbH (n.d.) *GNSMART*, [Online]. Available: http://www.geopp.de/index.php?bereich=0&kategorie=31&artikel=35&seite=1 [Accessed May 1, 2014]

Germany, Federal Agency for Cartography and Geodesy (BKG) (2013) *Networked Transport of RTCM via Internet Protocol*, Mar 01, [Online]. Available: http://igs.bkg.bund.de/ntrip/about [Accessed Oct 7, 2013]

Gurtner, W. Estey, L. (2007/2012) RINEX The Receiver Independent Exchange Format Version 2.11, [Online]. Available: http://igscb.jpl.nasa.gov/igscb/data/format/rinex211.txt [Accessed Nov 21, 2012]

Gurtner, W. Estey, L. (2009) RINEX *The Receiver Independent Exchange Format Version 3.0*, [Online]. Available: http://igscb.jpl.nasa.gov/igscb/data/format/rinex301.pdf [Accessed Nov 21, 2012]

Hatanaka, Y. (2008) `A Compression Format and Tools for GNSS Observation Data´, Bulletin of the GSI [Electronic], vol. 55, March, pp. 1-2, Available: http://www.gsi.go.jp/ENGLISH/Bulletin55.html [Accessed Jan 20, 2013]

Heo, Y. Yan, T. Lim, S. Rizos, C. (2009) `International Standard GNSS Real-Time Data Formats and Protocols´, In *International Global Navigation Satellite Systems Society IGNSS Symposium*, Gold Coast. Available: http://www.gmat.unsw.edu.au/snap/publications/heo_etal2009b.pdf

Hofmann-Wellenhof, B. Lichtenegger, H. Wasle, E. (2008) *Gnss: Global Navigation Satellite Systems: Gps, Glonass, Galileo*, and More. 449. Wien: Springer-Verlag.

Hewlett-Packard Development Company, L.P. (2011) *Server drive technology*, [Online] March. Available: http://h20566.www2.hp.com/portal/site/hpsc/template.BINARYPORTLET/public/kb/docDiplay/resource.process/?spf_p.tpst=kbDocDisplay_ws_BI&spf_p.rid_kbDocDisplay=docDiplayResURL&javax.portlet.begCacheTok=com.vignette.cachetoken&spf_p.rst_kbDocDiplay=wsrprsourceState%3DdocId%253Demr_nac010714964%257CdocLocale%253D&javax.portlet.endCacheTok=com.vignette.cachetoken [Accessed May 4, 2014]

Koivula, H. (2006) *Implementation and Prospects for Use of a High Precision Geodetic GPS Monitoring Network (FinnRef) Covering Finland*, Licentiate Thesis. Helsinki: Helsinki University of Science

Koivula, H. (2013) `Geodeettisen laitoksen GNSS-verkon uudistus ja julkinen data - paikannuksen uudet tuulet´, In Maanmittauspäivät, Kuopio. Available: http://www.kuopio.fi/c/document_library/get_file?uuid=fe3a3fe8-62d7-4705-9719-140452510eb5&groupId=12117

Koivula, H. (2014a) *Uuden FinnRefin tietoja*. [Email] Tenhunen, T. Feb 10.

Koivula, H. (2014b) *Kuvia asemilta*. [Email] Tenhunen, T. Apr 4.

Koivula, H. Kuokkanen, J. Marila, S. Tenhunen, T. Häkli, P. Kallio, U. Nyberg, S. Poutanen, M. (2012) `Finnish permanent GNSS network: From dual-frequency GPS to multi-satellite GNSS´, In *Ubiquitous Positioning, Indoor Navigation, and Location Based Service (UPINLBS)*, Helsinki. Available: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6409771

Koivula, H. Poutanen, M. (2014) `Geodeettinen laitos aloittaa ilmaisen paikannuspalvelun´, *Käytännön Maamies*, no. 1, January, pp. 46-49.

Küpper, A. (2005) *Location-Based Services*, West Sussex: John Wiley & Sons.

Liotine, M. (2003) *Mission-Critical Network Planning*, Norwood: Artech House.

Lucas, M. W. (2009) *Network Flow Analysis*, San Francisco: No Starch Press.

Martin, D.J. (2001) `GPS Basics: Static GPS´, *Professional Surveyor Magazine*, vol. 21, no. 11, December. Available: http://www.profsurv.com/magazine/article.aspx?i=828 [Accessed Jan 20, 2013]

McCabe, J.D. (2007) *Network Analysis, Architecture, and Design*, 3rd edition, Burlington: Morgan Kaufmann.

Moxa Inc. (2012a) *Nport 5200 series: 2-port RS-232/422/485 serial device servers*, [Online]. Available: http://www.moxa.com/doc/specs/NPort_5200_Series.pdf [Accessed March 25, 2014]

Moxa Inc. (2012b) Nport *5200A series: 2-port RS-232/422/485 serial device servers*, [Online]. Available: http://www.moxa.com/doc/specs/NPort_5200A_Series.pdf [Accessed March 8, 2014]

O'Driscoll, C. Petovello, M. (2010) `Generating Carrier Phase Measurements´, Inside GNSS Magazine, vol. 5, no. 5, pp. 18-22. Available: http://www.insidegnss.com/auto/julaug10-solutions.pdf

Poutanen, M. (1998) *GPS-paikanmääritys* [Position Determination with GPS], in Finnish, Helsinki: Tähtitieteellinen yhdistys URSA.

Parmar, M.S. Meniya A.D. (2013) `Imperatives and Issues of IPSEC Based VPN´, *International Journal of Science and Modern Engineering (IJISME)*, vol 1, no. 2, January, pp. 38-41, Available: http://www.ijisme.org/attachments/File/v1i2/B0129011213.pdf [accessed March 8, 2014]

Paros, J. Yilmaz, M. (2002) `Broadband Meteorological Sensors Co-located with GPS Receivers for Geophysical and Atmospheric Measurements´, In *Position Location and Navigation Symposium (PLANS)*, Palm Springs. Available: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=998900

Rao, K.R. Ramesh, D. Suryanarayana, K. (2006) `EMI Hardening of GPS Clock -A Case Study´, In *ElectroMagnetic Interference and Compatibility (INCEMIC)*, Bangalore. Available: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5419702

Rizos, C. (1999) *Principles and Practice of GPS Surveying*, [Online]. Available: http://www.gmat.unsw.edu.au/snap/gps/gps_survey/chap4/425.htm [Accessed Mar 26, 2014]

Royal Observatory of Belgium (ROB) /Euref (2012) Euref Permanent Network Documentation Formats, [Online]. Available: http://www.epncb.oma.be/_documentation/formats/ [Accessed Jan 19, 2013]

Sabatini, R. Palmerini, G.B. (2008) *Differential Global Positioning System (DGPS) for Flight Testing*, pp 1-1 – 1-3, [Online]. Available: http://ftp.rta.nato.int/public//PubFullText/RTO/AG/RTO-AG-160-V21///AG-160-V21-01.pdf [Accessed Mar 27, 2014]

UNAVCO (2011) *UNAVCO Knowledgebase: Remote33*, August 10, [Online]. Available: http://facility.unavco.org/kb/questions/374/Remote33 [Accessed Mar 26, 2014]

Ye, N. (2008) *Secure Computer and Network Systems: Modeling, Analysis and Design,* West Sussex: John Wiley & Sons.

FinnRef user questionnaire (original in Finnish)

1. Has the network/system reliability been at a satisfying level for the past five years?:

2. What kinds of problems have you encountered in the network during the last five years?:

3. How have these problems impacted your or your subordinate’s work?:

4. Has any security issues been noted in the network/system recently? If so, describe the problems briefly:

5. Do you have any wishes concerning the network features or performance?:

6. Do you have any propositions to improve the system/network reliability or other general comments?: