ADDIS ABABA UNIVERSITY
SCHOOL OF GRADUATE STUDIES
COLLEGE OF NATURAL SCIENCES
SCHOOL OF INFORMATION SCIENCE

PERFORMANCE ANALYSIS FOR WIDE AREA NETWORK OPTIMIZATION:

THE CASE OF ADDIS ABABA UNIVERSITY

BY:

TSEGAYE BERHANU NEGGA

ETHIOPIA, ADDIS ABABA

OCTOBER 2014

ADDIS ABABA UNIVERSITY
COLLEGE OF GRADUATE STUDIES
COLLEGE OF NATURAL SCIENCES
SCHOOL OF INFORMATION SCIENCE

PERFORMANCE ANALYSIS FOR WIDE AREA NETWORK OPTIMIZATION:

THE CASE OF ADDIS ABABA UNIVERSITY

A THESIS SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES OF ADDIS ABABA UNIVERSITY IN PARTIAL FULFILLMENT OF THE REQIREMENTS FOR THE DEGREE MASTER OF SCIENCE IN INFORMATION SCIENCE

BY:

TSEGAYE BERHANU NEGGA

ADVISOR:

WORKSHET LAMENEW

OCTOBER 2014

ADDIS ABABA UNIVERSITY
SCHOOL OF GRADUATE STUDIES
COLLEGE OF NATURAL SCIENCES
SCHOOL OF INFORMATION SCIENCE

PERFORMANCE ANALYSIS FOR WIDE AREA NETWORK OPTIMIZATION:
THE CASE OF ADDIS ABABA UNIVERSITY

BY

TSEGAYE BERHANU NEGGA

Name and Signature of Members of the Examining Board:

Name	Title	Signature	Date
1. Workshet Lamenew	Advisor	_____	October, 2014
2. Million Meshesha (PhD)	Examiner	_____	October, 2014
3. Micheal Melese	Chair Person	_____	October, 2014

TABLE OF CONTENTS

DECLARATION v

DEDICATION vi

ACKNOWLEDGEMENT vii

LIST OF TABLES viii

LIST OF FIGURES ix

LIST OF ABBREVIATIONS x

ABTRACT xi

CHAPTER ONE 1

INTRODUCTION 1

1.1. Background 1

1.2. ICT in AAU 4

1.3. Statement of the Problem 6

1.4. Objective of the Study 9

1.4.1. General Objective 9

1.4.2. Specific Objectives 9

1.5. Significance of the study 9

1.6. Scope and Limitation of the study 10

1.7. Methodology of the Study 11

1.7.1. Review of Related Literature 11

1.7.2. Data Collection Method 12

1.7.3. Data analysis 12

1.7.4. Network Traffic Analysis and Evaluation Toolset 12

1.7.5. Network Performance Analysis Procedures 13

1.7.6. Proposing WAN optimization Framework 14

1.8. Organization of the Thesis 14

CHAPTER TWO 15

REVIEW OF RELATED LITERATURE 15

2.1. Overview of Wide Area Networks 15

2.2. WAN OSI Architecture 16

2.2.1. WAN Physical Layer 17

2.2.2. WAN Data Link Layer 18
2.3. WAN Switching 19
2.3.1. Circuit Switching 19
2.3.2. Packet Switching 20
2.4. WAN Components 21
2.5. WAN applications and Application Protocols 22
2.6. Network Management Protocols 27
2.6.1. Internet Control Message Protocol (ICMP) 27
2.6.2. Simple Network Management Protocol (SNMP) 27
2.6.3. Management Information Base (MIB) 28
2.6.4. Windows Management Instrumentation (WMI) 28
2.7. WAN Performance of an Organization 28
2.8. WAN Performance Metrics 30
2.8.1. Availability 30
2.8.2. Bandwidth 31
2.8.3. Response Time 34
2.8.4. Latency and delay 34
2.8.5. Chatty protocols 37
2.8.6. Packet loss 37
2.8.7. Throughput 38
2.8.8. Congestion 38
2.9. Network Performance Assessment 39
2.10. Obstacles to Content Delivery Over a WAN 40
2.10.1. Network and Transport Barriers 40
2.10.2. Application and Protocol Barriers 41
2.10.3. Operating System and Hardware Barriers 41
2.11. The WAN optimization approaches 41
2.11.1. Improve bandwidth utilization 41
2.11.2. Reduce perceived latency 42
2.11.3. Improve lossy network performance 42
2.12. Network Traffic Analysis Tools 42
2.13. Review of Related Works 48
2.13.1. Global Related Works 48
2.13.2. Local Related Works 50

CHAPTER THREE 52
CASE DESCRIPTION AND ANALYSIS TOOL 52
3.1. Network Infrastructure of AAU (LAN/WAN) 52
3.1.1. AAU Network Architecture **53**
3.1.2. AAU Data center **54**
3.1.3. AAU WAN Applications and Services **55**
3.2. Performance of AAU-WAN 56
3.3. WAN Traffic Analysis Tool 59
3.4. Network Traffic data Collection from AAU WAN Link 60
3.5. Overview of Network Performance Monitor (NPM) 62
3.6. AAU WAN NODE 62
CHAPTER FOUR 65
RESULTS AND DISCUSSIONS 65
4.1. Performance Analysis of Network Latency at AAU WAN Link 65
4.1.1. Average Response Time 65
4.1.2. Percent Loss 67
4.1.3. Average Response Time and Packet Loss 68
4.1.4. Network Availability 69
4.1.5. Average Response Time and Availability 70
4.1.6. Average Response Time and Packet Loss 72
4.2. Summary of Findings for AAU Network 73
4.3. Proposed WAN Optimization Framework 74
4.3.1. Desirable Features of WAN optimization Framework 75
4.3.2. Components of WAN Optimization Framework 76
4.3.3. Validation of the WAN Optimization Framework 79
CHAPTER FIVE 81
CONCLUSION AND RECOMMENDATIONS 81
5.1. CONCLUSION 81
5.2. RECOMMENDATIONS 82
REFERENCES 84
APPENDIX 88

DECLARATION

I, the undersigned, declared that this thesis work is my original work, has not been presented for a degree in this or any other universities, and all sources of materials used for this thesis work have been duly acknowledged.

Name : *Tsegaye Berhanu Negga*

Signature: _____

Date : _____

This thesis has been submitted for examination with my approval as a university advisor.

Name : _____

Signature: _____

Date : _____

DEDICATION

IN MEMORY OF MY DAD WHO WAS A GOOD FATHER:

BERHANU NEGGA

ACKNOWLEDGEMENT

First and foremost, my greatest gratitude goes to the Almighty GOD who blessed me with His abundant love and care and gave me the strength to accomplish this thesis.

I would like to thank my advisor Workshet Lamenew for his guidance and advice until the end of this work. Throughout, he provided me all the supports I needed for this thesis work.

I would also like to thank Dr. Million Meshesha for his unreserved guidance and support. He has shown me the right direction to complete my research successfully. This thesis work would not have been possible without the endless supports and creative ideas from him.

I would also like to thank my friends Ermias Abebe, Amanuel Woldu, Tadele Gebremicheal, Meseret Ayano and Teklehaimanot Assefa for their support and care.

Finally, I would like to express my sincere gratitude to all my colleagues of the School of Information Science, Addis Ababa University for their unreserved support in many aspects whenever I required their professional assistance. I have to give a special thanks for my friend and colleague Ermias Abebe for his uncountable support.

LIST OF TABLES

Table 3.1 Node selected for minoring ……62

Table 4.1 Maximum and Minimum Response time……..66

Table 4.2 Packet Loss in Percentage…….67

Table 4.3 Average Response time and Packet loss……...68

Table 4.4 Network Availability in Percentage……..69

Table 4.5 Average Response time and Availability……..71

Table 4.6 The Minimum, Maximum and Average Response Time and packet Loss…...…72

Table 4.7 Performance Evaluation of the Framework by Domain Experts ……....…80

LIST OF FIGURES

Figure 2.1 WAN operations focus primarily on Layer 1 and Layer 2 (Cisco, 2008)……15

Figure 2.2 WAN Technologies and OSI Model (Cisco, 2008)……..16

Figure 2.3 Organization Network (WAN + LAN)…….28

Figure 2.4 Network management system composed of station and Agents……..42

Figure 3.1 An outline of how NPM monitors the network……47

Figure 3.1 AAU Network architecture……...54

Figure 3.2 AAU Data Center……..55

Figure 3.3 Procedural Architecture for WAN performance Analysis of AAU …….60

Figure 3.4 Overview of Network Performance Monitor ……...61

Figure 4.1 Average Response time…….66

Figure 4.2 Packet Loss in Percentage…….67

Figure 4.3 Average Response time and Packet loss……...68

Figure 4.4 Traffic Availability in Percentage…….69

Figure 4.5 Average Response time and Availability ……71

Figure 4.6 The Minimum, Maximum and Average Response Time and Packet loss rate….72

Figure 4.7 Building Blocks of the optimization Framework……..76

Figure 4.8 The system architecture of the WAN optimization Framework……..…….79

LIST OF ABBREVIATIONS

BDP	Bandwidth delay product
BRI	Basic Rate Interface
CIFS	Common Internet File System
CSU	Channel Service Unit
CW	Congestion Window
DSC	Digital Service Unit
DTE	Data Terminal Equipment
ER	Event Reporter
HDLC	High Level Data link Control
ICMP	Internet Control Message Protocol
IMAP	Internet Message Access Protocol
MAPI	Messaging Application Programming Interface
MIB	Management Information Base
MMS	Multimedia Server Protocol
NCR	Network condition recognizer
NFS	Network File System
NPM	Network Performance Monitor
OIDS	Object Identifiers
OSP	Optimization solution provider
POP	Post Office Protocol
PSTN	Public Switched Telephone Network
RTT	Round Trip Time
SMDS	Switched Multimedia Data Services
SMTP	Simple Mail Transfer Protocol
SNMP	Simple Network Management Protocol
VOIP	Voice over Internet Protocol
VPN	Virtual Private Network
WBD	WAN bottleneck determiner
WMI	Windows Management Instrumentation

ABTRACT

Wide Area Network (WAN) is one of the most important tool for organizations to run their day to day activities. Today, applications and services that are given using local area network (LAN) are also provided using WAN. However, there is more impact on the WAN performance. Besides, most protocols designed for LAN environments do not perform well over the WAN. The factors affecting the WAN are network availability, bandwidth, network latency, congestion and packet loss. WAN at AAU is no exception. This study is aimed to investigate the effects of different WAN factors using performance analysis tool, with the view to develop a WAN optimization framework that can improve the performance of the AAU network traffic flows over the WAN link.

To achieve the objective of the study, an experiment has been conducted using real-time cases that are taken from the AAU WAN environment. The experimentation is conducted through three major phases: network traffic data is collected using Network Performance Monitor (NPM) tool from the AAU WAN environment. The collected data is then analyzed and evaluated to investigate the network performance using metrics such as network availability, response time, and packet loss. Finally, based on the analysis result a WAN optimization framework is developed. In order to develop the WAN optimization framework the researcher followed steps such as Network Condition Recognizer, WAN Bottleneck Determiner, Optimization Solution Provider and Event Reporter.

The results of this study indicates that, high response time rate, high packet loss rate, and fluctuating network availability is exhibited in the AAU WAN environment. The WAN optimization framework is developed to solve the real time bottleneck status of the WAN link and apply optimization techniques accordingly. There is also a need to evaluate application performance of the network from the point view of users' experience. It is therefore recommended for further research.

Keyword: Network Performance Metrics, Network Traffic data, AAU WAN, Optimization framework.

CHAPTER ONE

INTRODUCTION

1.1. Background

As stated by Kansanen (2009), Network is one of the most important tools for organizations to run their operations. Many business critical applications and services rely on the organizations network. Nowadays there has been a switch from local area networking (LAN) to wide area networking, as a result of which a number of challenges related to network performance are observed, because most of the applications used today are designed to be used in LAN, the transition to WAN will have an impact on network performance. Also the amount of data transferred and services provided through networks is growing at the same time. Due to these reasons network performance has received a lot of attention.

As reported by Zhang (2012), today's IT organizations tend to deploy their infrastructures geographically over a wide area network (WAN) to increase productivity, support global collaboration and minimize costs, thus constituting to today's WAN-centered environments. As compared to a local area network (LAN), a WAN is a telecommunication network that covers a broad area; WAN may connect across metropolitan, regional, and/or national boundaries. Traditional LAN-oriented infrastructures are insufficient to support global collaboration with high application performance and low costs. Deploying applications over WANs inevitably incurs performance degradation owing to the intrinsic nature of WANs such as high latency and high packet loss rate (Zhang, 2012).

As the computer networking became more popular, every aspect of life has been shifted to network oriented technology, which results in increase computing power, sharing of resources and communication between users. The proliferation in the network technologies poses challenges to the network administrator on how to manage and control the emerged network. In computer networks, challenges may arise, which may disrupt the state of such network. Typically, computer network management challenges grow as the computer network expands. Thus, the need to manage the network arises upon the network growth (Yusuff, 2012).

Several ways to evaluate the performance of network systems have been developed by Saito (2002): statistical analysis, benchmarking and network monitoring. Statistical analysis is applied to activity logging in servers. This is a popular way to evaluate server performance. However, this analysis cannot provide any hints about network links or clients. Benchmarking is another performance measurement method. It can provide various indices of server performance. However, benchmarking requires a special environment, and the results are valid only for that environment. Network monitoring allows us to evaluate network usage at the data link level. However, it focuses on network equipment management in the network layer. Its results do not indicate the quality of application services. This is because the application level performance includes not only the characteristics of the data link, but also many other performance factors.

An active network may be simplistically viewed as a set of "Active nodes" that perform customized operations on the data flowing through them. Traditional data networks were designed with the aim of transferring bits from one end system to another. The transport mechanism achieved its objectives with minimal computation within the network. In contrast, active networks allow the network nodes to perform computation on the data passing through them. In fact, some implementations also allow their users to inject customized programs into the nodes of the network that may modify, store or redirect the user data flowing through the network (Prabhavalker, 2003).

According to Prabhavalker (2003), there is an increasing number of applications that require more support from the network nodes besides the storage and forwarding of bits that the nodes presently provide. These applications include group communication strategies, scalable network management, provisioning for quality of service, efficient routing protocols and congestion control mechanisms. Active networks provide a new networking platform that is flexible and extensible at runtime and supports the rapid evolution and deployment of networking technologies to suit current needs. They allow the network nodes to perform application specific computation on the data flowing through them. Although, with active networking the possibilities for refining current applications and introducing new ones are tremendous, it is important to demonstrate the performance benefits accrued from an active networking platform.

The ultimate goal of all network infrastructures is to deliver applications efficiently to users. Continuous performance improvements in PCs and LAN infrastructure have made it easy for IT to enable effective application delivery within a building or campus. But maintaining an acceptable level of application performance across the WAN proves a consistent challenge. To cope with poor performance across the WAN, enterprises have had to make a number of accommodations. IT organizations have been forced to proliferate data centers and server hardware around the world, install applications locally within branch offices, and endure the high cost of increased WAN bandwidth typically second only to staffing as IT's highest expense just because applications run slowly or simply cannot operate across the typical poor performing WAN link (Juniper, 2005).

Juniper (2005) states that the drop in bandwidth as traffic moves from the LAN to the WAN is obvious and well understood (Section 2.8.2). It is also known that, the effects of latency slows down the network performance even when ample bandwidth is available. According to Juniper (2005), "Network managers who do not spend time addressing the latency issue will not meet the service levels that global applications and business processes demand." Finally, application contention becomes far more prevalent on bandwidth-restricted WAN links, sometimes getting worse as a result of addressing the bandwidth limitation. To improve application delivery, IT must look for tools that increase available WAN bandwidth, accelerate applications despite the presence of latency, and resolve contention. In their search for these tools, IT is best served by looking for options that incorporate support for both general TCP acceleration as well as application-specific acceleration to improve the performance of key business applications and processes. Ultimately, IT needs a broad-based approach to WAN optimization that will accelerate the widest cross-section of their business applications, improve the flexibility of their delivery options, and provide the necessary monitoring and reporting to track application (Juniper, 2005).

According to Yusuff (2012), Performance management is the top level network management operation. It is responsible for monitoring, controlling and optimizing the overall network performance of network services. Performance management includes functions such as gathering statistical information, maintaining and examining logs of the system state histories and altering system modes of operation for the purpose of conducting performance management activities. As

stated by Cole and Ramaswamy (2000), the factors affecting network performance are mostly latency, congestion and packet loss.

Latency is the time that data travels in one way or as a round trip time (RTT) in the network. Congestion refers to a network state where a node or link carries so much data that it may deteriorate network service quality, resulting in queuing delay, frame or data packet loss and the blocking of new connections. In a congested network, response time slows with reduced network throughput.

Congestion occurs when bandwidth is insufficient and network data traffic exceeds capacity. Packet loss from congestion is partially countered by aggressive network protocol retransmission, which maintains a network congestion state after reducing the initial data load. This can create two stable states under the same data traffic load - one dealing with the initial load and the other maintaining reduced network throughput. Congestion has been described as a fundamental effect of limited network resources, especially router processing time and link throughput. Cumulative router processing time greatly impacts network congestion. Routers may actually discard data packets when they exceed its handling capability. When this occurs, additional data packets may be sent to make up for unreceived packets, which exacerbates the problem. Network congestion often leads to congestion collapse. Packet loss can be caused by a number of factors including signal degradation over the network medium due to multi-path fading, packet drop because of channel congestion, corrupted packets rejected in-transit, faulty networking hardware, faulty network drivers, or normal routing routines. Packet loss can also happen intentionally through network dissuasion technique for operational management purposes (Cole and Ramaswamy, 2000).

1.2. ICT in AAU

Addis Ababa University (AAU) is the first university in Ethiopia. As stated by AAU (2014) “On March 20, 1950, Emperor Haile Selassie I declared the foundation of the University College of Addis Ababa, which includes the faculties of Arts and Science”. It was renamed Haile Selassie I University in 1962 and then Addis Ababa University in 1975. At the starting time there were only 33 students enrolled. But nowadays the university enroll students in thousands. And, at

this time the government of Ethiopia is using it as a main resource to train qualified staff for other universities in different subject areas (AAU, 2014).

The University has made a remarkable contribution to the country through provision of trained manpower, research and community services. The service it has rendered in the training of high level skilled manpower and professionals in various key areas of development is unprecedented, and, yet, shines everyday as it stood the only University for decades. Its role and impact in the country's progress in various spheres of development is far-reaching. It currently runs 65 undergraduate and 220 graduate programs (out of which 69 are PhD programs) in 14 campuses which are distributed over major campuses and minor campuses, all within the capital, except the Akaki campus that is 45 km south of the capital.

All campuses under AAU, including the recently acquired Akaki campus have had inter-campus connection in the form of a hybrid wired and wireless connection for the last three years. There is a current effort underway, in collaboration with ETC, to convert the somewhat limited wireless connections to fiber. The main AAU campuses (Sidist Kilo, Amist Kilo, and Arat Kilo) serve as the core of the network with redundant high speed connectivity (AAU, 2013).

The ICT Development Office was established around the summer of 1996 through visionary leadership of few individuals who realized that the Addis Ababa University (AAU) would be wise to join the information age by adopting the technology that has been transforming the world. The newly formed office initiated a project named AAU-Net that has resulted in a wide area network (WAN) whose first phase of construction completed in November 2001. The network, which connects all the 14 widely distributed campuses of the university, has been growing since then. The services delivered through the infrastructure have also been increasing. The internal network (LAN/WAN) is connected to the internet via a 100Mbps link to the ETC exchange. The connection is managed internally through a gateway and protected from intrusion and virus and other attack by two firewalls.

The national attention given to the expansion and improvement of higher education as critical factors in the country's development has explicit and implied requirements for the use of ICT in realizing the objectives. AAU's role as a major contributor to these expansion and enhancement efforts, along with the imperatives contained in its own ambitious strategic plan, call for the speedy improvement of the efficiency and quality of its academic and administrative functions. This is hard, if not impossible, to accomplish without adequate ICT support. There are currently many initiatives that are underway to enhancing network performance, both at the ICT development office and various quarters around the University, to meet the growing demand for and address the ICT support needs of the university.

1.3. Statement of the Problem

Applications, deployed over a wide area network (WAN) suffer from performance degradation owing to unavoidable natural characteristics of WANs such as high latency, congestion and high packet loss rate. This is due to the fact that, most of the existing protocols are not designed for WAN environments; therefore, several application protocols do not perform well under the WAN condition (Sevcik and Wetzel, 2008).

The reports from Aberdeen Group (2007) shows that eighty-seven percent of organizations face difficulties in managing the performance of their WAN, as new technologies and applications are being rolled out over since 2005. New technology rollouts, convergence of voice and data networks, and multiple bandwidth intensive applications are driving organizations to add more bandwidth to support the role of their network in achieving strategic organizational goals. However, according to Riverbed (2013), adding more bandwidth is a short tactic for addressing poor application performance problem. Because, throwing more bandwidth may mask the worst symptom of a poorly performing network for a while, but it does nothing to treat the underlying problem itself.

Many kinds of applications on the Internet are widely used. So it is important that quality of these services, such as response time and throughput, be high. To achieve high quality the service in the network systems, there is a need to evaluate how the systems are working and operate so as to

perform users' requests and optimize performance. To manage network system performance, it is important to be aware of system usability factors, such as access delay, Congestion, packet loss processing time, and data transfer throughput (Saito, 2002).

Organizations rely heavily upon their network and applications to drive day-to-day operations and support employees and customers. With more remote and mobile users dependent on the wide area network (WAN) and more Web-based applications coming online every day, the amount of traffic crossing organizations' networks is growing exponentially. When bandwidth congestion and network issues are allowed to impede the performance of critical applications, productivity suffers and the entire organization may be put at risk. To address the above problems, organizations can overcome these challenges by implementing a Unified Performance Management (UPM) solutions that address the three key requirements of application performance management: visibility, control and optimization (Exinda, 2009).

To improve application delivery, IT must look for tools that increase available WAN bandwidth, accelerate applications despite the presence of latency, and resolve contention. In their search for these tools, IT is best served by looking for options that incorporate support for both general TCP acceleration as well as application-specific acceleration to improve the performance of key business applications and processes (Juniper, 2005).

As per the discussions with the AAU ICT office experts, different services and applications are consolidated in the AAU ICT Datacenter located at 6kilo. And, hence the WAN is being used intensively by different services and applications like Web Service using Joomla, e-mail Service using Google, e-Learning Service using Moodle, Domain Name System using BIND, Proxy Server, Library System, e-Granary/Internet in a BOX, Video Conferencing, DHCP Service, FTP Service, UCIS /University College Information System, Sun Ray Server / Thin Clients and CISCO Networking.

From the interview conducted with the ICT staff experts of AAU, they asserted that the network performance problem is affecting the operations of the University. Slow internet connection, unable to transfer file using FTP, message send failure and network unavailability are the main

network performance problems. Part of achieving the goal of high performance is active monitoring of networks to help in the identification and prevention of network problems. Despite the pioneering role, AAU has played in the deployment and use of ICT; however, AAU recently has a relatively old WAN infrastructures and it is still far from a point where it served adequate services provided for the user. At the same time, AAU's need for and dependency on effective ICT support is now greater than ever. Most of the time, some applications are given high priority than others by organizations considering their criticality for the business. To address the above problems, there is a need to evaluate how the AAU WAN is working and operate so as to perform users' requests and optimize performance.

Few studies have been conducted on the AAU webserver. Tadele (2011) studied web usage pattern discovery of Addis Ababa University (AAU) website using a server log file data. Awet (2011) has studied his research on exploring the navigational behavior of users of Addis Ababa University (AAU) official website.

On the other hand, as to the researcher knowledge, there is no attempt to study the network performance evaluation in AAU WAN environment. Therefore, this study attempts to investigate the network performance to identify the critical WAN factors that degrades the efficiency of the AAU network.

Hence, the research questions this study attempt to answer are the following.

- What is the performance level of the AAU network?
- What are the main bottlenecks affecting the performance of AAU network?
- What causes the bottlenecks?
- What is the suitable WAN optimization framework that can improve the performance problem of AAU WAN?

1.4. Objective of the Study

1.4.1. General Objective

The general objective of this study is to analyze the Performance of AAU WAN so as to develop WAN Optimization Framework that enables to enhance the performance efficiency of AAU network.

1.4.2. Specific Objectives

In order to achieve the above stated general objective, the researcher has formulated the following specific objectives.

- To review literatures so as to identify performance indicators and network evaluation mechanisms.
- To capture network traffic data from the AAU WAN link
- To determine the major performance network Bottlenecks that impacts the AAU WAN
- To analyze the impact level of these bottlenecks on the AAU network using a suitable Performance metrics.
- To propose WAN Optimization Framework that enables to enhance the network performance.

1.5. Significance of the study

The basic concept of "**you cannot improve what you cannot measure and understand**" must be applied for the optimization of the network latency of the organization. The goal of this paper is to enhance understanding of network latency sources and underlying trade-offs, to facilitate effective measurement and analysis, and to support development of a consistent network latency monitoring and improvement strategy. Moreover, this study is to provide a clear view on the performance problems of the AAU networks and to offer a possible solution for overcoming these problems. The AAU ICT experts can use of these findings to identify potential areas of concern or WAN bottlenecks and enhance their already deployed network application and services. Moreover, it will initiate other researches to undertake studies towards evaluating the performance of local and wide area networks in other institutions in general, and services and applications in particular.

The WAN optimization framework developed in this study can be transformed in to commercial application to be used by organization WANs when their network is in low resourced.

This study has both theoretical and practical contributions towards improving the performance of network traffic flow. The theoretical contribution of this study is researchers can make use of the defined metrics as well as the proposed framework for further studies. The practical contribution of this study for system and network administrators by using the Network Performance Monitor is to provides an ideal solution for the following needs: Isolating traffic bottlenecks within the network, Graphing real-time results, Identifying high traffic nodes, Building customized reports and Alerting on any network properties, Posting charts and reports on the web using the HTML Publish feature. Finally, this study suggests further research for enhancing the efficiency and effectiveness of the proposed WAN optimization framework

1.6. Scope and Limitation of the study

The scope of this research is limited to evaluate only the performance of the AAU Wired Wide Area Network (WWAN). This research studies the network performance of AAU, it does not include the application performance. The WAN factors selected for this experimentations are limited to conditions which is mostly shown in under resourced university networks. Besides, the Real-time traffic data is collected from a selected and managed device which is located at the datacenter of the main campus. The collected traffic data is about various features of the AAU WAN, including network latency in terms of response time, packet loss and network availability of the AAU WAN link. The real-time traffic flow data is collected from August 10, 2014 to October 10, 2014 using a SolarWinds Network Performance monitor toolset. Hence, the recent real-time traffic data of 30 days are selected as a representative (from August 14, 2014 to September 14, 2014) and are used in this study.

The limitation of this study is that, the study is conducted during the summer time when the AAU is partially closed during the month of August and September. At this time there are less number of network users because most of the students and some academic staffs are at leave. Therefore, it would have been good if this study is conducted during the AAU regular academic year in order

to investigate the real effects of network users on the AAU WAN performance, because the number of network users have an impact on the performance of the network. The other limitation of this study is on the usage of network traffic data. Data is collected only from the real-time traffic data (online traffic data) of the target organization (AAU) WAN because the offline traffic data (log file data) was not accessible and captured by the node which can indicates the long time performance of the AAU network. The other limitation of this study is the fluctuation of electricity during the capturing of network traffic data and is resulted in broken charts and lines. Finally, this research is aimed to develop only the WAN optimization framework for AAU WAN environment. Developing an algorithm and a model is left as a recommendation for further research areas.

The other limitation of this study is in measuring the performance of the AAU WAN. This study used three performance metrics such as response time, packet loss and network availability. Other performance metrics such as bandwidth utilization, throughput and traffic received and transferred are not used in this research because of the lack of time and budget. Besides, this study didn't cover the application performance and it only uses the real time network traffic data. These are the main challenges in this study and they are left for further research.

1.7. Methodology of the Study

In order to achieve the general objective of the current research, the researcher has employed an experimental research methodology described as follows:

1.7.1. Review of Related Literature

The researcher reviewed different related literatures (books, journal articles, conference papers, research reports and web documents) in order to have detail understanding of approaches and methods pertaining to the research under consideration. A literature review is conducted in order to assess the main issues and concepts in the field of networks, network performance problems, WAN performance metrics and their effects, WAN optimization solutions and network performance tools. Besides, local and global related researches to network performance analysis has been reviewed.

1.7.2. Data Collection Method

To identify, understand and analyze the business problems, primary data collection methods such as observation and interview and secondary data collection methods such as manual and document analysis have been used. The researcher took secondary data to have a better understanding about the target organization to know how the AAU network architecture looks like and what are the existing network infrastructures. The potential source of data used to undertake this research was mainly the real-time network traffic data of the AAU WAN link which is collected using a network performance monitoring tool (NPM). The real-time traffic data is collected from August 10, 2014 to October 10, 2014. To collect the data from the AAU WAN link, a Network performance monitor tool is installed and configured on the AAU router. Data is collected on the researcher's machine Using a web console techniques. The researcher collected 30 days of data and is taken as a representative for analysis purpose; that is from August 14, 2014 to September 14, 2014. In addition to this, the researcher has conducted an interview with the ICT office IT experts and an observation is done at the datacenter of AAU ICT office to get a thorough understanding of the business/domain.

1.7.3. Data analysis

Preliminary data analysis is done to understand the collected traffic data. This enables to determine the major performance bottlenecks that influence the AAU network performances and to identify applications of the organization under study. Experiments are conducted in the real environment of the AAU WAN and on the given configured WAN Link so that the traffic flow is analyzed and evaluated in terms of different performance metrics. The analysis and evaluation is done using Solarwinds Network performance monitor toolset by examining the traffic flows in the managed device, which helps to determine the network performance at different time intervals. Traffic flow analysis further helps to find a clear pattern in the performance behaviors of the network for different WAN scenarios such as Response time, packet loss and network availability.

1.7.4. Network Traffic Analysis and Evaluation Toolset

Network is an environment with millions of variables and traffic analysis tools are a way of getting an idea of how much and what king of traffic goes though the network. Network traffic analysis is useful for detecting performance problems in the network. In this study, Network performance

Monitor (NPM) v9 developed by Solarwinds, is used to capture, analyze and evaluate real-time traffic statistics data from the AAU WAN environment. Analysis of the collected traffic data is presented in the form of bar charts, line charts, step charts and tables. So, in this study, Solarwinds network performance monitor toolset is used for analyzing and evaluating the AAU wide area network performance because of the following advantages. Here are the top five reasons to use SolarWinds Engineer's Toolset:

- All the network tools you need in one complete package
- Monitoring tools include Real-Time Interface Monitor, SNMP Real-Time Graph, and more
- Diagnostic tools include Ping Sweep, DNS Analyzer, Trace Route, and more
- Network discovery tools include Port Scanner, Switch Port Mapper and more.
- Cisco management tools include Real-Time NetFlow Analyzer, Config Downloader

1.7.5. Network Performance Analysis Procedures

To evaluate the effect of different factors on application response for WAN end users, different performance metrics are used. Bai, Oladosu, & Williamson (2007) Uses five metrics: loss rate, response time, retransmission rate, end-to-end delay and effective throughput for detailed performance measurement. Metrics like reply rate, throughput, response time, and error rate statistics are also used for a high-level overview of the performance results.

To analyze the AAU network, the researcher follows some steps (mentioned in section 3.3). The first step is installing and configuring the Network performance monitor toolset on the AAU WAN link which is the Router device to ensure that it is exporting Network statistics data. Using a Web console technology on the local machine of the researcher, a network traffic data is captured from the managed device of AAU WAN Router from August 14, 2014 to October 14, 2014. After capturing the network traffic flow, data analysis is done in order to understand the performance of the network using different network performance metrics such as network availability, packet loss and response time. The analyzed data is interpreted and evaluated to have a clear understanding on how the AAU WAN is performing effectively and efficiency. Network traffic data analysis and interpretation is done using network performance monitor toolset. Finally, based on the findings of the experiment, a WAN optimization framework is developed to enhance the performance of the AAU network.

1.7.6. Proposing WAN optimization Framework

This study is aimed to adopt an approach which uses both the network performance and the interview results in proposing a WAN optimization Framework. Therefore, the results of the network performance as well as the conducted interview is used to propose a WAN optimization Framework which is to enhance the network performance of the AAU WAN. The proposed WAN optimization framework has been validated by five AAU ICT senior experts through consecutive discussions and a formal questionnaire (see section 4.3.3).

1.8. Organization of the Thesis

This thesis report consists of five chapters. The first chapter deals with the general overview of the study including background, AAU ICT, statement of the problem, objectives, scope and limitation and methodology of the research.

Chapter two discusses network performance concepts from the WAN perspective, Network management protocols, factors affecting network performance and optimization techniques are reviewed in detail. Related works with the current study are also discussed and finally tools for network performance analysis are investigated.

Chapter three discusses case description of the target organization, how the network performance monitor tool is used, and an architectural design on how the network performance evaluation is conducted on the target organization to evaluate the performance of AAU WAN.

Chapter four demonstrates the findings from the experimentation and discusses the result. This section is the core part of the thesis, since it presents and discusses the major analysis results of the experimentation at the managed Device “6-kilo_AR46-GW” in terms of Network Latency using response time, packet loss and network availability performance metrics. It also discusses the proposed WAN optimization framework. Finally, a validation of the proposed WAN optimization framework is discussed.

Chapter five presents conclusions of the research and implications for future work in the area of WAN performance improvement.

CHAPTER TWO

REVIEW OF RELATED LITERATURE

The researcher has reviewed different related literatures (books, journal articles, Conference proceeding papers, and the Internet) in order to have detailed understanding on the present research.

2.1. Overview of Wide Area Networks

Wide Area Network (WAN) is a computer network covering multiple distance areas and multiple smaller networks, such as local area networks (LANs) or Metro Area Networks (MANs). The world's most popular WAN is the internet. The key difference between WAN and a LAN/MAN is scalability, hence, WAN must be able to grow as needed to cover multiple cities, even countries and continents. WAN may be privately owned or rented from a service provider (Orike and Okwoli, 2011).

Wide Area Networks (WAN) have become more prevalent with the expansion of global organizations. WANs have provided more geographical flexibility, shared resources, and have even eased the workload in most organizations when performing at optimal levels. Historically, however, such IT configurations have not been found to provide a measurable level of productivity despite the rapid advances in computing technology. This shortfall has prompted decision makers to explore the Knowledge Value Added aspects of IT solutions to define the true return on investment associated with each adoption (Sanders, 2011).

WAN applications have become one of the most popular applications in current distributed computing (Tuosto, 2003). The Internet and World Wide Web is now the primary environment for designing, developing and distributing applications. Theoretical models for formally reasoning on WAN applications must consider many crucial aspects and their mutual relationships, e.g. mobility, network awareness, security and service level agreement (Teuosto, 2003).

2.2. WAN OSI Architecture

Open System Interconnection Reference model was developed by the International Organization for Standardization (ISO) in 1984, and it is now considered the primary architectural model for inter-networking communications. OSI is a reference model for describing how information is transferred between two applications in two different computers through a telecommunication network.

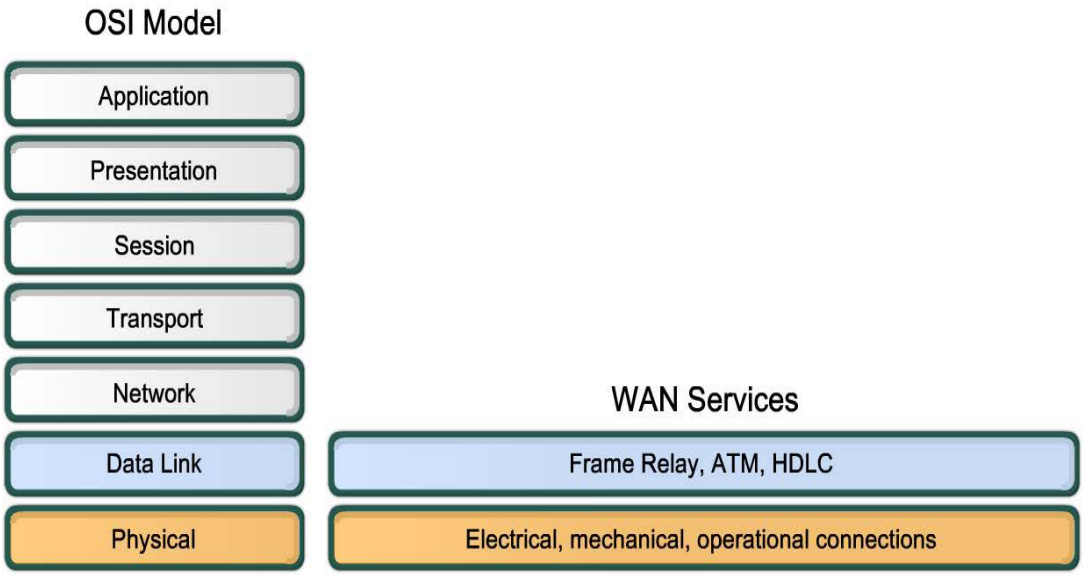

Figure 2.1: WAN operations focus primarily on Layer 1 and Layer 2 (Cisco, 2008).

This protocol stack helps to divide the tasks involved with moving information between networked computers into seven smaller, more manageable task groups. A task or group of tasks is then assigned to each of the seven OSI layers. These seven layers include: the application layer, the presentation layer, the session layer, the transport layer, the network layer, the data link layer, and the physical layer (Grevers & Christner, 2007). The upper layers of the OSI model deal with application issues and generally are implemented only in software. The highest layer, the application layer, is closest to the end user. Both users and application layer processes interact with software applications that contain a communications component. The lower layers of the OSI model handle data transport issues. The physical layer and the data link layer are implemented in hardware and software. The lowest layer, the physical layer, is closest to the physical network medium.

Information being transferred from a software application in one computer system to a software application in another must pass through the OSI layers. For example, if a software application in System A has information to transmit to a software application in System B, the application program in System A will pass its information to the application layer (Layer 7) of System A. The application layer then passes the information to the presentation layer (Layer 6), which relays the data to the session layer (Layer 5), and so on down to the physical layer (Layer 1). At the physical layer, the information is placed on the physical network medium and is sent across the medium to System B. The physical layer of System B removes the information from the physical medium, and then its physical layer passes the information up to the data link layer (Layer 2), which passes it to the network layer (Layer 3), and so on, until it reaches the application layer (Layer 7) of System B. Finally, the application layer of System B passes the information to the recipient application program to complete the communication process (Jenson, 2009).

Actual communication is made possible by using communication protocols. A protocol implements the functions of one or more of the OSI layers. A wide variety of communication protocols exist. Some of these protocols include LAN protocols, WAN protocols, network protocols, and routing protocols. LAN protocols operate at the physical and data link layers of the OSI model and define communication over the various LAN media. WAN protocols operate at the lowest three layers of the OSI model and define communication over the various wide-area media (Cisco, 2008). WAN technologies generally function at the lower two layers of the OSI reference model (Cisco, 2008): the physical layer and the data link layer. This does not mean that the other five layers of the OSI model are not found in a WAN. It simply means that the characteristics that separate a WAN from a LAN are typically found at the physical layer and the data link layer. In other words, the standards and protocols used in WANs at Layer 1 and Layer 2 are different from those used in LANs at the same layers. Figure 2.2: illustrates the relationship between the common WAN technologies and the OSI model of the internet (Cisco, 2008).

2.2.1. WAN Physical Layer

The WAN physical layer (OSI Layer 1) describes the interface between the data terminal equipment (DTE) and the data circuit-terminating equipment (DCE). Generally, the DCE is the

service provider and the DTE is the attached device. In this model, the services offered to the DTE are made available through a modem or a channel service unit/digital service unit (CSU/DSU) (cisco, 2008).

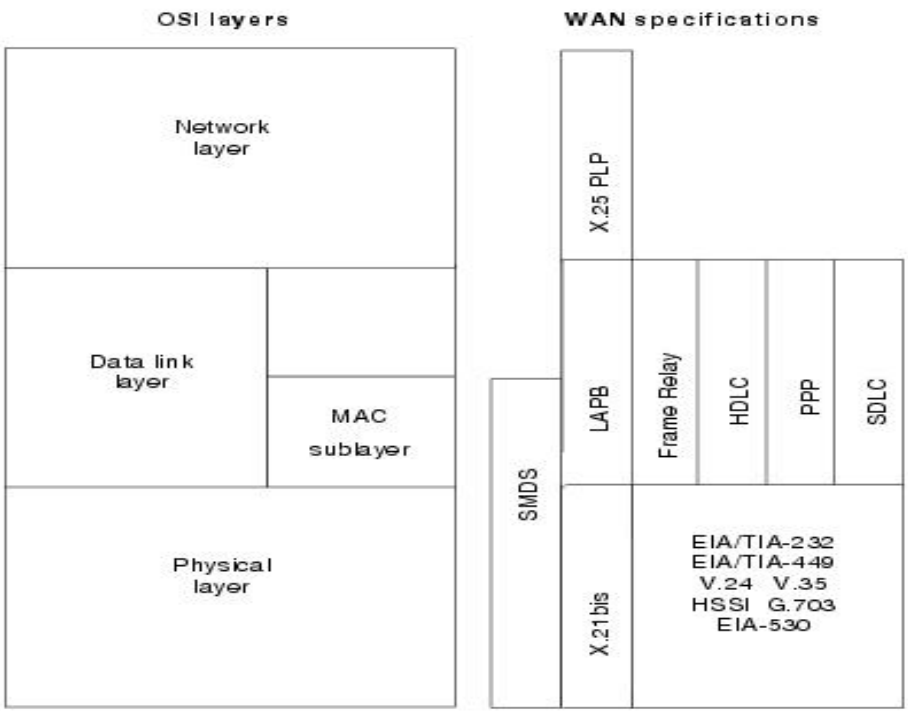

Figure 2.2: WAN Technologies and OSI Model (Cisco, 2008)

2.2.2. WAN Data Link Layer

WANs require Data Link layer protocols to establish the link across the communication line from the sending to the receiving device. The data link layer (OSI Layer 2) protocols define how data is encapsulated for transmission toward a remote location and the mechanisms for transferring the resulting frames. A variety of technologies are used, such as ISDN (Integrated Services Digital Network), Frame Relay and Asynchronous Transfer Mode (ATM). Some of these protocols use the same basic framing mechanism, High-Level Data Link Control (HDLC), an ISO standard, or one of its subsets or variants.

Data from the Network layer is passed to the Data Link layer for delivery on a physical link, which is normally point-to-point on a WAN connection. The Data Link layer builds a frame around the

Network layer data so that the necessary checks and controls can be applied. Each WAN connection type uses a Layer 2 protocol to encapsulate a packet while it is crossing the WAN link. To ensure that the correct encapsulation protocol is used, the Layer 2 encapsulation type used for each router serial interface must be configured (cisco, 2008).

Frame Encapsulation Formats: The frame always starts and ends with an 8-bit flag field. The bit pattern is 01111110. The address field is not needed for WAN links, which are almost always point-to-point. The address field is still present and may be 1 or 2 bytes long. The control field is protocol dependent, but usually indicates whether the content of the data is control information or network layer data. The control field is normally 1 byte.

2.3. WAN Switching

A set of switches and routers are interconnected to form a WAN using packet switching and circuit switching technologies (Cisco, 2008).

2.3.1. Circuit Switching

Circuit switching is a one type of WAN network that creates straight physical link between sender and receiver. The system selects which connection to follow the transaction depends on the connection. Circuit switching method is suitable for analog based communication. Disadvantage of this switching is waste of bandwidth since a connection is booked for sender and receiver (Kokulan, 2010).

Circuit switching allows data connections to be established when needed and then terminated when communication is complete. A good example of circuit switching is Integrated Services Digital Network (ISDN). When a router has data for a remote site, the switched circuit is initiated with the circuit number of the remote network. Virtual private network (VPN) is a technology widely used in a public switched network to provide private and secured WAN for an organization. VPN uses encryption and other techniques to make it appear that the organization has a dedicated network, while making use of the shared infrastructure of the WAN. With this capability, files can be shared from the head office of an organization to the branch offices online and in real-time (Orike, Okwoli, 2011).

PSTN (Public Switched Telephone Network) and ISDN (Integrated Services Digital Network) are two types of circuit switching technology that may be used to implement a WAN in an enterprise setting. In the case of ISDN circuits, the device actually places a call to the telephone number of the remote ISDN circuit. When the two networks are connected and authenticated, they can transfer data. When the data transmission is complete, the call can be terminated (Cisco, 2008).

2.3.2. Packet Switching

Packet switching allows users to share common carrier resources so that the carrier can make more efficient use of its infrastructure. In a packet switching set-up, networks have connections into the carrier's network, and many customers share the carrier's network. The carrier can then create virtual circuits between customers' sites by which packets of data are delivered from one to the other through the network (Orike, Okwoli, 2011).

In the packet switch network transmission of messages or data's are divided into small parts called packets. The system is then routing these packets to destination address via network connection. These packets are carried by different paths in the network and packets are arrived in different order. The end system arrange all the packets in proper order. Packet switching splits traffic data into packets that are routed over a shared network. Packet switching networks do not require a circuit to be established, and they allow many pairs of nodes to communicate over the same channel. The switches in a packet-switched network determine which link the packet must be sent on next from the addressing information in each packet. There are two approaches to this link determination, connectionless or connection-oriented (Kokulan, 2010).

Frame relay is a type of packet switching wide area network technology that connects several company LANs. Compared with other technology, it is inexpensive, so, it is suitable for small companies. Frame relay is connection oriented network. The main advantage of frame relay is inexpensive and works at high speed ("44.376 Mbps"). It is working at physical layer and data link layer of OSI reference model. The disadvantage of this technology is that there is no setting for error correction and flow control (Cisco, 2008).

2.4. WAN Components

WANs use numerous types of devices that are specific to WAN environments: WAN switches, access servers, modems, channel service unit/digital service unit (CSU/DSU), and ISDN terminal adapters. Other devices found in WAN environments that are used in WAN implementations include routers, ATM switches, and multiplexers (Kokulan, 2010). The above WAN devices are discussed as follows:

WAN Switch: A WAN switch is a multiport internetworking device used in carrier networks. These devices typically switch such traffic as Frame Relay, X.25, and SMDS, and operate at the data link layer of the OSI reference model. Two routers at remote ends of a WAN can be connected by WAN switches.

Access Server: An access server acts as a concentration point for dial-in and dial-out connections. An access server concentrates Dial-Out connections into a WAN illustrates an access server concentrating dial-out connections into a WAN.

Modem: A modem is a device that interprets digital and analog signals, enabling data to be transmitted over voice grade telephone lines. At the source, digital signals are converted to a form suitable for transmission over analog communication facilities. At the destination, these analog signals are returned to their digital form. A modem connection through a WAN handles Analog and digital signals illustrates a simple modem-to-modem connection through a WAN.

CSU/DSU: A channel service unit/digital service unit (CSU/DSU) is a digital-interface device used to connect a router to a digital circuit. The CSU/DSU also provides signal timing for communication between these devices. The CSU/DSU stands between the Switch and the Terminal illustrates the placement of the CSU/DSU in a WAN implementation.

ISDN Terminal Adapter: An ISDN terminal adapter is a device used to connect ISDN Basic Rate Interface (BRI) connections to other interfaces, such as EIA/TIA-232 on a router. A terminal adapter is essentially an ISDN modem, although it is called a terminal adapter because it does not actually convert analog to digital signals. The Terminal Adapter Connects the ISDN Terminal

Adapter to Other Interfaces illustrates the placement of the terminal adapter in an ISDN environment.

Router: Router is one of the important devices in the network. It is used to connect more than one network together such as two WANs or two LANs and it allow data transmission between the networks. Routers work at transport and network layer of the OSI reference model. Routers maintain traffic control in the network. A router typically connects to several links; this is accomplished through the incoming and outgoing ports of the router. These ports are connected to a switching fabric, which is a combination of hardware and software that moves data from an input port to an output port. The main task of the router is to forward traffic, and it is necessary that this is performed efficiently to avoid congestion. In order for routers to forward packets onto appropriate links, they make use of routing tables where it is specified which link to use for a specific packet in order to reach the destination. It is important that these tables are continuously updated with valid information to avoid incorrect decisions. The time needed for routers to search through the tables and to find the required information is often referred to as the processing time (Bergfeldt, 2010).

2.5. WAN applications and Application Protocols

A protocol is a common language by which computers communicate. It is a set of rules or standards used by computers to convey, transfer and share information across a network. These rules can be implemented at the hardware or software level, or using a combination of the two. Protocols are not applications themselves, they are used various applications since the protocols are designed for a particular purpose, the purpose drives the behavior more than the implementation does (Cisco, 2008).

Internet Protocol (IP)

Main function of internet protocol (IP) is transmitting message from one end to another end. When the messages sent via this protocol, they are divided into many parts, each part containing IP address of end system and the small segments sent through different ways are received in different order. This protocol is implemented in network layer of OSI reference model and internet layer of

TCP/IP model; so internet protocol (IP) supports data transmission to upper level layers in OSI and TCP/IP reference model. This protocol is connectionless protocol and we cannot make sure about delivery of message. Nowadays two different version of internet protocols (IP) are available in the networking, such as internet protocol version4 (IPV4) and internet protocol version6 (IPV6) (Kokulan, 2010).

Transmission Control Protocol (TCP)

Transmission control protocol (TCP) applied in transport layer of OSI reference model and host to host transport layer of TCP/IP model for message transmission. This protocol is a connection-oriented and reliable transmission protocol. Connection oriented means source (sender) set up connection with destination (receiver) node before sending data. Error detection, error correction and retransmission of data is possible in TCP. Message transmission via this protocol is very slow because of overhead (20bytes) for error correction/detection and acknowledgement. There are sending and receiving buffers that are used for storage, flow control, error detection and correction. Using TCP messages can be transferred in both direction. It provides full duplex transaction and also multiplexing and de-multiplexing are possible in TCP (Kokulan, 2010).

TCP is the de facto standard designed to provide reliable end-to-end delivery of data packet in the wired networks. Normally, TCP is an independent protocol that is not related to the underlying network technology. However, some assumptions of TCP, such as consideration of only static node, packet losses due to congestion or buffer overflows are inspired from the features of wired networks. In order to apply TCP to an ad hoc environment, TCP has to overcome many problems, such as packet losses due to congestion, high bit errors, node mobility, longer delay and so on (Oo, Othman, 2011).

User datagram protocol (UDP)

User datagram protocol (UDP) installed in transport layer of OSI reference model and host to host transport layer of TCP/IP model for message transmission between the layers. This protocol is a connectionless and unreliable transmission protocol. Message transmission via UDP is very fast

because there is small header (8bytes). This protocol containing many drawbacks such as no error control and no flow control. Receiver can't identify whether this message is original or duplicate when message transfer via UDP. Encapsulation and decapitation of messages are possible in UDP when messages transfer from source to destination (Kokulan, 2010).

Hypertext Transfer Protocol (HTTP)

Web applications, running over HTTP (Hypertext transfer protocol), are similarly subject to "Ping-Pong" behavior. In the case of HTTP, the applications can transfer data at the full TCP window size, but the protocol retrieves the individual objects on each page one at a time. The TCP window scale option is an option to increase the receive window size allowed in Transmission Control Protocol above its former maximum value of 65,535 bytes. Since most web pages have a few dozen objects per page, it can take dozens of Round Trip Times to retrieve all the objects associated with a URL. Web caching often cannot accelerate this process since, in most cases, the protocol will still check with the web server to confirm the freshness of an object before sending it to the client. As a result, the latency impact remains even when caching reduces the WAN bandwidth (cisco, 2008).

Web browsing is an HTTP-based application that is characterized by short-lived TCP transfers. The performance of such an application strongly depends on packet loss, hence we chose to present the results we obtained using it. The traffic of interest (HTTP) competes with the background traffic to occupy queue space which induces loss and for being serviced which induces delay (Ivanovici, Beuran and Davies, 2005).

Windows File Access

Microsoft file services rely on the Common Internet File System (CIFS) protocol, which behaves like Exchange in its use of small data blocks for file reads and writes. Again, the "ping-pong" behavior as each transmission requires an acknowledgement delays users in branch offices trying to open, read, or write to files stored on centralized servers (Juniper, 2005).

Simple Mail Transfer Protocol

Networking was created to make communication and file sharing more efficient. The most widely used electronic communication application is e-mail. E-mail is made possible by the Simple Mail Transfer Protocol (SMTP). SMTP was designed to "transfer mail reliably and efficiently". FTP and HTTP also facilitate reliable and efficient communication, but they require both the sender and receiver to be connected at the same time. To facilitate communication with a host that is not on the network, SMTP is used (Jenson, 2013).

Electronic mail (e-mail) is one of the most popular network services nowadays. Most e-mail systems that send mail over the Internet use simple mail transfer protocol (SMTP) to send messages from one server to another. The messages can then be retrieved with an e-mail client using either Post Office Protocol (POP) or Internet message access protocol (IMAP). SMTP is also generally used to send messages from a mail client to a mail server in "host based" mail systems, where a simple mbox utility might be on the same system via Network File System (NFS) provided by Novell for access without POP or IMAP (Riabov, 2005).

Voice over IP (VoIP)

Voice over IP (VOIP) uses the Internet Protocol (IP) to transmit voice as packets over an IP network. So VOIP can be achieved on any data network that uses IP, like Internet, Intranets and Local Area Networks (LAN). Here the voice signal is digitized, compressed and converted to IP packets and then transmitted over the IP network. Signaling protocols are used to set up and tear down calls, carry information required to locate users and negotiate capabilities. One of the main motivations for Internet telephony is the very low cost involved (Arora, 2000).

File Transfer Protocol (FTP)

According to Cisco (2008), many network administrators consider the File Transfer Protocol (FTP) a "necessary evil." Legacy applications and logging hosts commonly use FTP for simple, authenticated file transfers. FTP is viewed as a simple solution but with a potential for a major impact. Everything from the NIC of the server to the WAN link that the traffic will traverse is

impacted by the manner in which FTP transfers content. FTP can disrupt the ability to pass other traffic at the same time an FTP transfer is taking place. By nature, FTP consumes as much bandwidth as possible during its transfer of content, based on what TCP is allowed to consume. FTP tends to use large data buffers, meaning it will try to leverage all of the buffer capacity that TCP can allocate to it (Cisco, 2008).

FTP is such a common protocol on the Internet and WAN networks. The objectives of FTP include are (Jenson, 2009): promote sharing of computer programs, data and/or files, encourage indirect or implicit use of remote computers, shield a user from variation in file storage systems among hosts, and transfer data reliably and efficiently. Precautions exist within many operating systems and third-party applications that allow the administrator to define an upper limit to any given FTP transfer, preventing congestion situations. FTP is not fault tolerant and, by nature, is very unforgiving to disruptions. In many cases, a network disruption requires that the file be retransmitted from the beginning. Some application vendors have written client and server programs that leverage FTP as a control and data transfer protocol that allow for continuation of a failed transfer, but these mechanisms are not built into FTP as a protocol itself (Jenson, 2009).

Secure Sockets Layer (SSL)

As stated by Deirks and Allen (1999), encrypting critical business processes has a significant role in organization networking. More and more applications are used through the web and require encryption. There are many possibilities for improving the security of these applications. Secure Sockets Layer (SSL) has become the most used. SSL can be used with protocols like HTTP, File Transfer Protocol (FTP) and Simple Mail Transfer Protocol (SMTP). SSL was developed first by Netscape. There are two main phases of SSL: handshake and data transmission. Handshake is done between client and server and the purpose of the handshake is to determine the secret-key parameters. These secret keys are then used during the data transfer to encrypt and decrypt the data sent over the network. The encrypted data cannot be seen by network devices like routers or accelerators. This makes analyzing or for example prioritization of the SSL encrypted data difficult. SSL also demands a lot of processing power, especially during the handshake phase, which also affects the performance. One of the downsides of SSL is that also spyware and peer-to-peer applications, like instant messaging, exploit the secure SSL tunnel.

2.6. Network Management Protocols

Network management protocols are used by the Network Performance Monitor to access managed devices. NPM uses the following most commonly used network management protocols, namely Internet Control Message Protocol, Simple Network Management Protocol, Management Information Base and Windows Management Instrumentation (Solarwinds, 2014).

2.6.1. Internet Control Message Protocol (ICMP)

NPM uses the Internet Control Message Protocol (ICMP) to poll for status using echo requests of managed devices. When NPM polls a managed device using ICMP, if the device is operationally up, it returns an echo reply that NPM uses to calculate a response time. NPM also records any reported packet drops. This information is used by NPM to monitor status and measure average response time and packet loss percentage for managed devices. NPM only uses ICMP to poll devices for status, average response time, and packet loss percentage.

2.6.2. Simple Network Management Protocol (SNMP)

For most network monitoring and management tasks, NPM uses the Simple Network Management Protocol (SNMP). SNMP-enabled network devices, including routers, switches, and PCs, host SNMP agents that maintain a virtual database of system status and performance information that is tied to specific Object Identifiers (OIDs). This virtual database is referred to as a Management Information Base (MIB), and NPM uses MIB OIDs as references to retrieve specific data about a selected, SNMP-enabled, managed device. Access to MIB data may be secured either with SNMP Community Strings, as provided with SNMPv1 and SNMPv2c, or with optional SNMP credentials, as provided with SNMPv3. To properly monitor devices on the network, SNMP must enable on all devices that are capable of SNMP communications for performance information. SNMP credentials secure access to SNMP-enabled managed devices. SNMPv1 and SNMPv2c credentials serve as a type of password that is authenticated by confirming a match between a clear text SNMP Community String provided by an SNMP request and the SNMP Community String stored as a MIB object on an SNMP-enabled, managed device.

2.6.3. Management Information Base (MIB)

A Management Information Base (MIB) is the formal description of a set of objects that can be managed using SNMP. MIB-I refers to the initial MIB definition, and MIB-II refers to the current definition. Each MIB object stores a value such as bandwidth utilization. During polling, SolarWinds NPM sends a SNMP GET request to each device to poll the specified MIB objects. Received responses are then recorded in the SolarWinds database for use in NPM, including within Web Console resources.

Network devices can support several different types of MIBs. While most devices support the standard MIB-II MIBs, they may also support any of a number of additional MIBs that the network administrator may want to monitor. Using a fully customizable Universal Device Poller, one can gather information from virtually any MIB on any network device to which a person has access.

A MIB is a database used to store management information in networks. MIBs are used by network Management systems to identify network data objects that are stored, retrieved and set by the system. A MIB uses a hierarchical tree structure to store an extensible collection of data. The MIB used for SNMP is standardized; however, it allows private organizations to insert custom objects into the structure (Moceri, 2014).

2.6.4. Windows Management Instrumentation (WMI)

Windows Management Instrumentation (WMI) is a proprietary technology used to poll performance and management information from Windows-based network devices, applications, and components. When used as an alternative to SNMP, WMI can provide much of the same monitoring and management data currently available with SNMP-based polling with the addition of Windows-specific communications and security features. Due to specific characteristics of WMI polling requests, polling a single WMI-enabled object uses approximately five times the resources required to poll the same or similar object with SNMP on the same polling frequency.

2.7. WAN Performance of an Organization

According to Kansanen (2009), the term performance in networks is used to describe the performance of applications and to how the end users experience it. He studied the network performance from the application performance point of view. The focus is on the network

performance in WAN, although many of the characteristics can be also found in LAN. He describes the different factors affecting to network performance. Cole and Ramaswamy (2012) state that the factors affecting network performance are bandwidth, latency, throughput, congestion and packet loss.

From the network perspective an organization have at least 3 different kinds of offices inside the network: branch offices, regional offices and datacenters. Many companies prefer to have the employees as close to the customer as possible and therefore the number of small branch offices within one organization can be high. The different offices are connected to each other with WAN links. Branch offices or remote locations are connected to a bigger regional office or a datacenter usually with low bandwidth high latency WAN links. Therefore, branch offices are the ones suffering most from a poor performance. Organization WAN consist of the carrier backbone and the point-to- point links between the carrier cloud and different office building LANs. All of these are separated from each other by routers. The point-to-point links can be Digital Subscriber Line (xDSL) or Multi-Protocol Label Switching (MPLS) links (Kansanen, 2009). Another important factor affecting organization network performance are the applications used through the network.

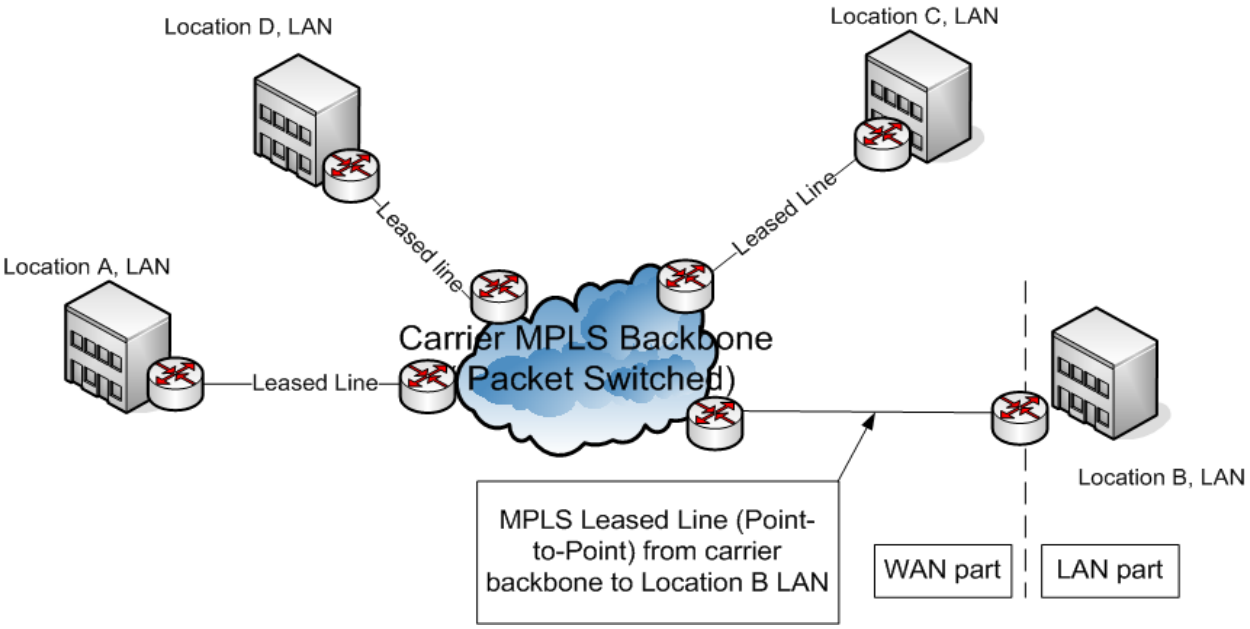

Figure 2.3: Organization Network (WAN + LAN) (Kansanen, 2009)

The above Figure 2.3 shows an example of how an organization network can be constructed from the WAN and the LAN part. WAN consists of the carrier network cloud and the point to point links between the different offices and the carrier cloud (Kansanen, 2009).

Grevers and Christner (2010) state that largest part of the network costs in a global organization comes from the servers, deployed infrastructure and the management of the servers in these branch offices. Only from the performance perspective it is ideal for the branch office workforce to have a local server infrastructure for applications. To reduce the costs and ease the management many companies have moved these local services into global datacenters. Due to this consolidation of services the branch office users have experienced lower application performance. The large amount of small offices and at the same time the centralization of servers into datacenters has made the network a critical point for the organizations.

2.8. WAN Performance Metrics

WAN performance refers to measuring of the efficiency and service quality of a wide area network infrastructure and the network traffic flows through them. The following metrics are suggested for measuring the efficiency network performance (Bai, Oladosu and William, 2007). The following metrics are suggested for measuring network performance.

2.8.1. Availability

The first step in measuring network performance is to determine if the network is even working. If traffic cannot traverse the network, there is a bigger problems than just network performance issues. The simplest test for network availability is the ping program. By attempting to ping remote servers from a client device on the network, you can easily determine the state of the network. The ping program sends an Internet Control Message Protocol (ICMP) echo request packet to the destination host. When the echo asks it the packet has been received, the remote host immediately returns an echo reply packet to the sending device. Receiving an echo reply packet from the remote host means that there is an available network path between the client and server devices. If no echo reply packet is received, there is a problem with either a network device or a link along the path (Kakay, 2006).

2.8.2. Bandwidth

In the context of data networks, the term bandwidth quantifies the data rate that a network link or a network path can transfer. In a packet network, the terms “bandwidth” or “throughput” often characterize the amount of data that the network can transfer per unit of time. Bandwidth corresponds to the amount of data that can be carried in a physical medium in a given time interval (usually in a second) (Raghavan, 2010).

Bandwidth refers to a measure of frequency ranges, typically used for digital communications. The band part of broadband is short for bandwidth, meaning that the device uses a relatively wide range of frequencies. Bandwidth indicates the theoretical maximum capacity of a connection, but as the theoretical bandwidth is approached, negative factors such as transmission delay can cause deterioration in quality. Bandwidth refers to the transmission capacity of a computer channel or communications line or bus, usually stated in bits per second (Hailay, 2011).

The most obvious restriction for application delivery over the WAN is the reduced bandwidth available on WAN links. As a percentage of LAN bandwidth, WAN speeds have actually increased over the past couple decades, but discrepancies of 100 to 200 fold are still routine. While LAN infrastructure has scaled to 10Gbps of bandwidth, a 155 Mbps OC-3 link is considered very high bandwidth among WAN connections. Typically, businesses rely on T-1/E-1 links running at 1.5 to 2 Mbps or T-3/E-3 connections running at 45 Mbps or 34 Mbps. The typical 100 Mbps fast Ethernet LAN provides more than 60 times the bandwidth of a T-1 link (Juniper, 2005).

As stated by Juniper (2005), businesses today run more applications across the WAN, and often the bandwidth requirement per application has increased. Web-enabling applications such as ERP systems can cause a transaction’s bandwidth to increase as much as 10 fold compared to the bandwidth required for that same transaction in client/server application architecture. To keep up with these increases, enterprises have had to constantly increase the size of their WAN links. But given the high cost of these recurring expenses, enterprises are understandably reluctant to increase them.

A compelling alternative is compression technology, which replaces repeated data sequences with short flags for transmission across the WAN link. Traditional compression techniques provide limited bandwidth increases and often introduce additional latency, clearly at odds with improving application performance. Network infrastructure staff today must look for next generation compression techniques that can dramatically reduce the transmitted traffic and do not slow application delivery (Juniper, 2005).

IT also needs to consider another element when investigating compression options, the kinds of data patterns that benefit from compression. Various compression algorithms and implementations benefit different data sequence types. Algorithms that run solely in memory and operate on shorter data patterns benefit short, chatty applications such as SQL and HTTP. Other approaches that add hard-disk storage to the solution can store longer data sequences and can store them for longer periods of time. This kind of sequence caching approach eliminates repeated data patters in larger files, such as a PowerPoint presentation, even if the file itself has changed and even when the last transmission occurred weeks earlier. Sequence caching is an ideal technology for collaboration projects that rely on file sharing and for storage applications. IT needs a combination approach of compression techniques to achieve the highest overall reduction across a broad range of application types (Juniper, 2005).

The difference in bandwidth creates a bottleneck between the LAN and the WAN. These problems of oversubscription and utilization are explained as follows (Kansanen, 2009).

Oversubscription

The performance related problem caused by different bandwidths used inside network is called oversubscription. When data is coming from a higher rate network to a lower rate network the lower rate part might get oversubscribed. Oversubscription causes queues in the devices, usually routers, handling the change of bandwidth and thus slows down the transmission (Cole and Ramaswamy, 2000).

As an example let us assume that the LAN part of the connection is 100 Mbps Fast Ethernet and the WAN part is constructed with a point-to-point T1 (1.544 Mbps) links. These two parts are connected with routers to each other. The first point of oversubscription can happen already inside the LAN when

the three client workstations (each having a100 Mbps link to the switch) are there joined to one circuit of only 100 Mbps. Connecting three lines into one in the LAN causes 3:1 oversubscription. When moving from LAN to WAN the difference in the network speed means an oversubscription of 67:1 (100 Mbps / 1.544 Mbps). From these two the oversubscription presented in LAN is not causing as much performance problems as the oversubscription when entering the WAN. The LAN oversubscription might not even be present at all times since the three workstations might not be in use at the same time (Cole and Ramaswamy, 2000).

Utilization

Utilization is a way to measure the performance of the network. Utilization percentage describes how much of the total network bandwidth, network capacity, is in use at a certain point of time. The best way to detect over utilization in the network is to monitor it constantly. Measurements can be taken for instance from the busiest hour of the day and from the busiest minutes of the day (Kansanen, 2009). Kansanen further states that overall utilization should be most of the time less than 35%. It is hard to define when the performance is good - these reference values might change based on the network and how it is used. One important factor in defining whether the network is over utilized is to get the opinions of end users. In general, the reference values are based on the examinations of the statistical properties of the traffic.

A wealth of literature and tools guide end users and network operators in the successful management of their Ethernet networks. One good tool is flow analysis applied in the service provider network or in the end user network to identify top talkers and see why they are utilizing so much bandwidth. It could be that malware has entered the computer of an end user, generating excessive spam email traffic which saps private and public networks alike. Or, peer to peer file exchange may be occurring in violation of copyright laws at the same time as absorbing great network capacity. These sorts of problems can make users think something is wrong with their Carrier Ethernet service, when in reality, the service is working fine and provides plenty of bandwidth for proper usages. Service Providers can gain access to special flow analysis software and systems available in their router/switch management systems that provide excellent insight into the exact real-time sources of loads on the network. End users and service providers alike are advised to consult the rich literature on these general network management subjects (MEF, 2010)

2.8.3. Response Time

Response time is the key performance measure in on-line transaction processing systems and other client-server architectures (Harrison, 2010). Not only it is important to achieve a low average response time and correspondingly high throughput, but response time should also be fairly consistent in order to provide a good quality service. Response time is the time elapsed between a client issuing a request to the server and the arrival at the client of the response (e.g. data); alternately, it is sometimes considered as the elapsed time between successive requests, including processing time and thinking time at the client side.

It is important to be able to estimate the probability distribution of response time and, indeed, 95th quintiles are specified in the TPC benchmarks. In order to calculate average response time for a given transaction path, we merely need to sum the average delays in each component in that path. In the simplest case this is just mc + mn1 + ms + mn2 where mc, mn1, ms, mn2 are the mean delays at the client, network (from client to server), server and network (from server to client) respectively (Harrison, 2010).

As stated by Kakay (2006), in large networks there are many factors that can affect the response times between a client and server. These factors can include: Overloaded network segments, Network error, Faulty network wiring, broadcast storms, Faulty network devices and overloaded network hosts.

2.8.4. Latency and delay

Latency is the time taken by a packet to be returned to the sender and approximately equals the Round Trip Time (RTT). In reality, the actual latency is affected by many factors such as the network path taken to the destination (which includes the physical media that is used – fiber versus satellite versus copper), and the number of hops or routers between the source and the destination, with each hop adding some latency to the packet. Latency is typically expressed in milliseconds (Raghavan, 2010). Latency is used for describing the time that data travels in the network. It can be expressed as one-way latency or as roundtrip latency. One-way latency, also known as delay, simply means the time it takes for data to travel from the transmitting node to the receiving node.

Roundtrip latency, also known as Round Trip Time (RTT), measures the time data travels from transmitting node to the receiver plus the time that it takes for the transmitting node to get a response (acknowledgement) from the receiving node. When studying application performance, the most commonly used form of latency is the RTT (Cole & Ramaswamy, 2000).

The impact of latency has historically been a little less obvious and less well understood than the bandwidth limitation on WANs. Latency, which Kansanen (2009) calls "the silent killer of application performance," refers to the round trip time (RTT) for a packet to traverse from a sender to a receiver. On WAN links that cross the United States, typical latency times are 75ms to 100ms. In global networks, that RTT routinely reaches 250ms or more. Latency on satellite links routinely reaches 320ms to 430ms. Different application types are impacted by differing amounts of latency. While e-mail can still perform reasonable well over links with high latency, terminal services will be significantly impacted on the same link, resulting in dropped sessions in many cases. Sometimes latency doesn't just degrade application performance—sometimes it limits the overall application throughput. Enterprises that have purchased sizable WAN links often assume they have protected against application performance problems because they have sample bandwidth. But latency can limit throughput regardless of bandwidth. Reducing latency itself is not possible; latency is simply a result of the physics of the speed of light over longer distances combined with store-and-forward hops across routers. What IT needs to consider, then, is how to reduce the impact that latency has on how enterprise applications behave. Applications based on TCP as the reliable transport protocol are especially susceptible to latency limitations.

Latency can be divided into smaller delay components that together generate the overall network latency. These components are propagation delay, serialization delay, processing delay and forwarding delay. It is important to make the distinction between the different types of delay, since some of them are fixed and some are at least partly controllable. Understanding what parts of the delay can be controlled becomes important when talking about improving the network performance (Cole & Ramaswamy, 2000).

Propagation Delay

Propagation delay is a form of delay that is caused by the distance between nodes and physics in terms of how fast data can be transferred in the network. Propagation delay is one of the fixed factors affecting the overall network latency. Propagation delay is measured as the time the data packet spends to go through the network. The speed packets can be transferred is called propagation velocity and it is normally around 2/3 of the speed or light (3x108m/s). Propagation delay becomes significant in long distances, which is usually the case in WANs. Propagation delay Tp is defined as (Grevers & Christner, 2007):

Tp = $\frac{d}{v}$ Where d is physical distance (m) and v is propagation velocity (~ 2x108).

Serialization Delay

Serialization delay is measured as the time it takes to move bits of a packet into the line. It consists of the size of the packet, network medium and speed of the interface. Usually serialization delay is more significant in lower-speed networks. For example, assume the line speed to be 256kbps and the packet size 100 bytes. This creates a serialization delay of 3.1ms per hop. As an example of a high-speed network in a 1Gbps link the same delay would be 100ms. Serialization delay T_s is defined as:

$T_s = \frac{s}{r}$, where s is size of packets (bits) and r is transmission rate (bps).

Processing Delay

Time that it requires for a network node like a router, switch to perform required actions on the packet is called processing delay. In a router processing delay means the comparison of a piece of data to the access list. The forwarding architecture has to also be counted in processing delay: the node can either wait until the entire packet is received before it makes any decisions what to do with it (store and forward) or the forwarding of the packet can start as soon as the header is received. For instance, in a router the processing delay can vary between less than 1ms to even 10ms when the router is congested. For example the processing delay is estimated to be 1ms for each hop in the network, making a total of 3ms end to end processing delay (Cole & Ramaswamy, 2000).

Forwarding Delay

In routers or switches the time spent on deciding where to forward the packet. For example in a router the packet could go through in 1ms and under a load it could take 3 to 5ms for the same job. The forwarding delay is estimated to be 1ms per hop creating a total of 3ms forwarding delay for the end to end connection (Cole & Ramaswamy, 2000; Grevers & Christner, 2007).

Applications waiting for these processes to complete cannot fully fill the available WAN bandwidth and are likely being slowed down by latency. The easiest way to estimate the impact of latency on a network's performance is to calculate the bandwidth delay product (BDP). The bandwidth-delay product is an equation that says the capacity of a link is equal to the bandwidth of that link multiplied by latency as measured in round trip time (RTT). BDP is a product of latency and bandwidth and is typically used to express the amount of data that can be held in a given physical medium. The BDP, whose unit is in (MK) represents the capacity of the physical medium. The goal of all WAN optimization products is to keep the WAN completely full at all times (Grevers & Christner, 2007).

2.8.5. Chatty protocols

Chatty protocols are inefficient protocols, which clients use to communicate with servers. The protocols cause many applications that run well on the LAN to fail to deliver adequate performance on the WAN. On a LAN where the RTT is negligible, the chatty nature of these protocols is not significant enough to cause performance problems. However, on a WAN with even a moderate RTT, the performance impact becomes significant enough to not only be noticeable, but in many cases renders the application unusable. The chattiness of a protocol is directly proportional to the RTT of the link (Raghavan, 2010).

2.8.6. Packet loss

Packet loss occurs when one or more packets fail to reach their destination, typically due to quality of the physical medium. Higher quality networks, which are significantly more expensive, have lower packet loss and perform better. However, as noted by Raghavan (2010) on public networks

(like the Internet), loss of 1% is typical, and results in retransmissions in reliable protocols like TCP. Retransmissions make the latency problem even worse for TCP based applications.

For wired Ethernet, the transmitting stations listen for incoming signals (collisions) and emit a jamming signal to notify all other stations if a collision is detected. This provides accurate and timely feedback to the CSMA protocol which triggers a backoff in order to resolve the concurrent access. A packet loss could be due to weak signal, that is, the signal at the receiver side may be insufficient given the data rate that the packet was modulated. Determining the cause of a packet loss (collision and weak signal) is significant as this dictates the corresponding action to be taken at the link layer, for collisions, the transmitting station would perform an exponential backoff, while for weak signal the rate-adaptation algorithm would be invoked. As a result on the specific reason for packet loss, different actions should be taken at the data link layer, each adjusting different transmission parameters for minimizing packet loss (Rayanchu et al, ND).

2.8.7. Throughput

Throughput is the rate of successful data transfer in the network. Grevers and Christner (2007) define throughput as a sum of three different parameters: network capacity, latency and packet loss. Capacity means the maximum amount of information that can be transferred between two network nodes. The throughput of a network is never more than the capacity of the slowest hop within that network (Grevers and Christner, 2007). For example, if the network connection inside a branch office is 1Gbps and it is connected through a 1.5Mbps WAN link to the datacenter router and again the datacenter devices are connected to each other with a 1Gbps connection, the throughput of the connection would never exceed the 1.5Mbps. For throughput, latency means the time it takes to transfer data between two nodes and also the distance of the nodes. Packet loss adds the element of lost data to the sum, how many packets are dropped for example, due to congestion. Besides to these three parameters, when talking about throughput for a network, the transport protocol itself can limit the throughput. In cases like this adding more capacity to the network would not necessarily improve the throughput at all.

2.8.8. Congestion

As stated by Bergfeldt (2010), a large amount of packets may suddenly arrive to a specific router, which could give rise to congestion (especially if the majority of the packets should be forwarded

through one particular outgoing port). If the arrival intensity of new packets goes beyond the router's transmission capacity regarding a targeted outgoing port, the router is overloaded. In order to take care of this and to avoid losing the incoming packets, routers usually have a buffer capacity which makes it feasible to temporarily store outgoing packets.

Congestion occurs when load on the network (the number of packets to be sent to the network) is greater than the capacity of the network (the number of packets that a network can handle).The collisions in the network cause the routers and switches to have queues (buffers that hold packets before and after processing). If either packet arrival rate or packet departure rate is higher than the packet processing rate the input/output queue becomes longer, thus leading to congestion. So collisions introduce congestion (Malhotra, Gupta & Bansal, 2011).

2.9. Network Performance Assessment

A key issue in application performance assessment is the understanding of the fact that network Environments perturb application behavior by delaying and dropping the application traffic. Networks are therefore degraded environments, and quality degradation in the network is reflected in the performance degradation at application level.

There are three steps to take in order to assess application performance (Kansanen, 2009): first observe the application behavior at the end-node level, then accurately measure the quality degradation experienced by the application traffic and finally, correlate the above two steps.

Scientific method requires the use of objective metrics to perform both the network and application level performance assessments. In case of network quality degradation there is already a series of widely used metrics (Kansanen, 2009): one-way delay, one-way packet loss and throughput. However, when application performance must be determined, each application class requires the definition of specific metrics that take into account the application nature. For example, for Voice over IP (VoIP) one can use the Perceptual Evaluation of Speech Quality score. In case of file transfer, useful metrics are transfer time performance and good throughput.

2.10. Obstacles to Content Delivery Over a WAN

The performance degradation occurs when applications are deployed over a WAN owing to its unavoidable intrinsic characteristics. The obstacles to content delivery over a WAN can be categorized into four classes: network and transport barriers, application and protocol barriers, operating system barriers, and hardware barriers (Zhang, 2012).

2.10.1. Network and Transport Barriers

Network characteristics, such as limited bandwidth, latency, packet loss rate and congestion, impact the application Performance.

1. **Limited Bandwidth***:* The available bandwidth is generally much higher in a LAN environment than that in a WAN environment, thus creating a bandwidth disparity between these two dramatically different networks. The limited bandwidth impacts the capability of an application to provide high throughput. Furthermore, oversubscription or aggregation is generally higher in a WAN than that in a LAN. Therefore, even though the clients and servers may connect to the edge routers with high-speed links, the overall application performance over a WAN is throttled by network oversubscription and bandwidth disparity because only a small number of requests can be received by the server, and the server can only transmit a small amount of data at a time in responding to the clients' requests. Protocol overhead, such as packet header and acknowledgement packets, consumes a noticeable amount of network capacity, hence further compromising the application performance.

2. **High Latency***:* The latency introduced by transmission distance, protocol translation, and congestion is high in the WAN environment, and high latency is the major cause for long application response time over a WAN.

3. **Congestion and High Packet Loss Rate***:* Congestion causes packet loss and retransmission, and leads to erratic behaviors of transport layer protocols that may seriously deteriorate the application performance.

2.10.2. Application and Protocol Barriers

The application performance is constantly impacted by the limitations and barriers of the protocols, which are not designed for WAN environments in general. Many protocols do not perform well under WAN conditions such as long transmission path, high network latency, network congestion, and limited available bandwidth. Several protocols such as CIFS and MAPI are chatty in nature, requiring hundreds of control messages for a relatively simple data transfer. Some other popular protocols, e.g., Hypertext Transfer Protocol (HTTP) and TCP, also experience low efficiency over a WAN (Zhang, 2012).

2.10.3. Operating System and Hardware Barriers

The hosts, including their operating systems, which host the applications; and hardware platforms, which host operating systems, also impact the application performance. Proper selection of the application hosts' hardware and operating system components, including central processing unit, cache capacity, disk storage, and file system, can improve the overall application performance. A poorly tuned application server will have a negative effect on the application's performance and functionality across the WAN.

2.11. The WAN optimization approaches

With the different techniques available for WAN optimization, it is important to understand that simply turning on all available optimizations in every situation does not provide the best performance gains. Several considerations must be taken into account when determining the best set of optimizations to use for any given situation. The available optimization techniques are a set of tools available to solve a problem; selecting the right set of tools to use for a given situation is critical for maximum performance (Raghavan, 2011). To overcome the problems of network performance over the WAN include the following:

2.11.1. Improve bandwidth utilization

Techniques like data de-duplication exploit the redundancy inherent in the data traversing the WAN by maintaining dictionaries on each device and sending references to the redundant data,

thereby significantly reducing the amount of data over the WAN. This reduces the congestion and increases the effective capacity of the WAN. Compression is also used to compress data over the network to yield further bandwidth savings. De-duplication in conjunction with compression is the technique used to maximize bandwidth over the WAN (Raghavan, 2011).

2.11.2. Reduce perceived latency

This is a protocol specific optimization that is used to provide LAN like response times as perceived by the client program making the request. The technique typically involves a deep understanding of the underlying protocol and using read-ahead and write-behind approaches to stage the data on the device that's closest to the client. Because the WAN optimization devices in this case are a man-in-the-middle, transparently intercepting the client requests, it is not suitable for all protocols (Raghavan, 2011).

2.11.3. Improve lossy network performance

Using this technique, the WAN optimization device takes advantage of the symmetric deployment to provide TCP optimizations that would not normally be available when communicating with a general purpose server's TCP stack. Since both ends of the WAN TCP connection are "owned" by the WAN optimization device, the TCP stack behavior can be made more aggressive when addressing packet loss conditions in the network (Raghavan, 2011).

2.12. Network Traffic Analysis Tools

Several commercially and freely available network traffic analysis tools exist currently. The widely used network flow monitoring tools are OPNET (Optimized Network Engineering Tools) modeler and SolarWinds network performance monitor toolset.

OPNET supports most of network protocols and standards both in the wired and wireless area networks. It is outstanding because of its concurrent running of different simulation scenarios. OPNET is a well-known commercial suit of simulator and has a trail version that can be used for this project. OPNET is capable to simulate various network environments from local area network to global satellite network (Sanders, 2011).

Network performance monitoring is a critical component of a complete network management tool and performance. With SolarWinds Toolset that receive several tools that provide the ability to monitor everything from the traffic on a single port to the response time and availability of every server in the network (Solarwinds, 2014).

Overview of Network Performance Monitor (NPM)

SolarWinds Network Performance Monitor (NPM) delivers comprehensive fault and network performance management that scales with rapid network growth and expands with the network monitoring needs, allows to collect and view availability and real-time and historical statistics directly from the web browser. While monitoring, collecting, and analyzing data from routers, switches, firewalls, servers, and any other SNMP-, ICMP-, or WMI-enabled devices, SolarWinds NPM successfully offers a simple-to-use, scalable network monitoring solution for IT professionals juggling any size network.

Users find that it does not take a team of consultants and months of unpleasant surprises to get SolarWinds NPM up and running because the NPM experience is far more intuitive than conventional, complex enterprise network management systems. Because it can take less than an hour to deploy and no consultants are needed, NPM provides quick and cost-effective visibility into the health of network devices, servers, and applications on the network, ensuring that the real-time information needed to keep the systems running at peak performance.

SolarWinds NPM monitors the following critical performance metrics for physical and virtual devices on the network (Solarwinds, 2014).

- Network availability
- Bandwidth capacity utilization
- Buffer usage and errors
- CPU and memory utilization
- Interface errors and discards
- Network latency
- Node, interface, and volume status
- Volume usage

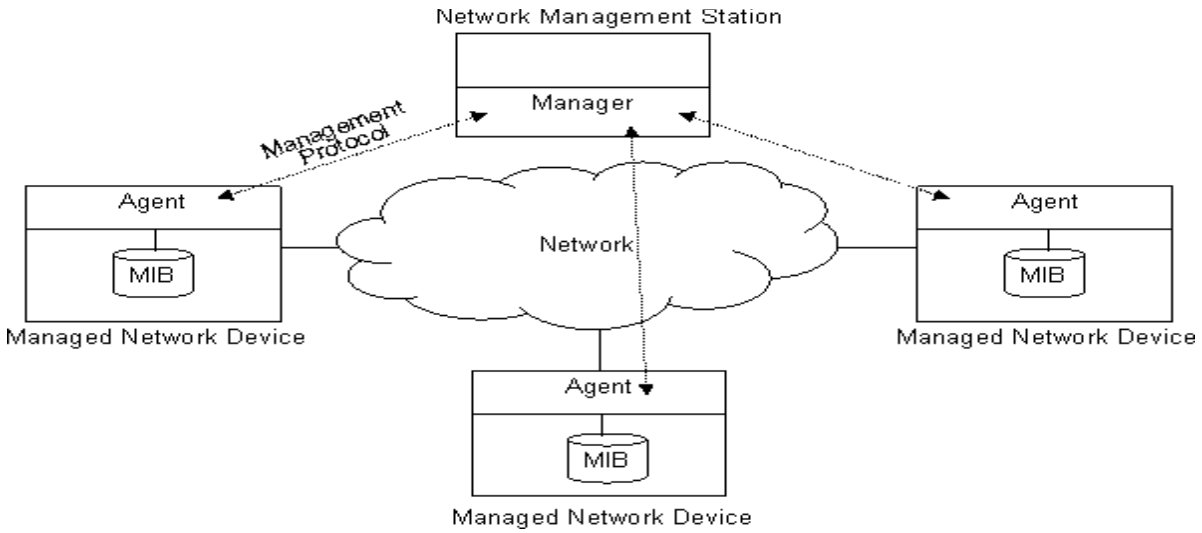

Figure 2.4: A network management system composed of station and agents (Solarwinds, 2014).

Solarwinds Network engineer's toolset is an integrated of SNMP and is designed to provide a variety of network management solutions ranging from individual monitoring tools to complete, full-featured monitoring platforms. Network Performance Monitor is the comprehensive monitoring solution built on SNMP. The network management application features a web interface with real-time monitoring of availability, bandwidth utilization, network latency and many other network performance metrics. The system automatically summarizes data and prioritizes events and alerts for easy monitoring and troubleshooting. Each event, statistic or alert also has a drill-down feature which provides all of the details on a given piece of information. This interface is also customizable with the ability to visually map network components and links, further easing the process of monitoring and finding errors (Solarwinds, 2014).

As discussed above, SNMP only provides the messaging format used in a network management system; individual devices must still be configured to work with a particular management system. So, NPM includes auto-discovery and auto-configuration features that simplify the process of adding network devices to be monitored. This is only a brief overview of NPM (Solarwinds, 2014). Network performance monitoring is an important part of network functionality.

Society has become dependent on networks and their ability to perform optimally is crucial. As a result, numerous network performance monitoring tools have emerged. Many of these are based

on the standardized management protocol SNMP and provide administrators a complete view of a network and it performance.

These monitoring capabilities, along with a fully customizable web-based interface, alerting, reporting engines, and flexible expansion capabilities, make SolarWinds Network Performance Monitor the easiest choice to make involving the network performance monitoring needs. Many other types of tools are also available. Ethereal provides a tool for capturing and analyzing individual packets off of a network. Web metrics GlobalWatch is an application performance monitoring tool that utilizes distributed agents to continuously monitor the performance of web applications. Network flow monitoring tools provide an increasingly popular means of monitoring network performance. Cisco's NetFlow architecture provides flow monitoring, aggregation, and analysis and has even gone to IETF as an Internet draft. As the trend towards higher performance networks continues, more comprehensive monitoring tools like NetFlow become as commonplace as SNMP today (Nagaraja, Chittal and Kumar, 2007).

2.12.1. How Network Performance Monitor Works

Through ICMP, SNMP, WMI, and MIB communication and data collection, NPM continuously monitors the health and performance of a network, and it does this without interfering with the critical functions of network devices. Unlike many other network monitoring products, NPM helps maintain the overall performance of network in the following ways (Solarwinds, 2014):

- NPM does not install outside agents on mission-critical servers
- NPM does not employ services that take vital resources from critical applications
- NPM does not install any code on monitored network devices. Unmanaged or outdated code can open security holes in the network.

After installing NPM, can automate the initial discovery of network, and then simply add new devices to network. NPM stores gathered information in a SQL database (the SolarWinds database) and provides a user-friendly, highly customizable web console in which to view current and historical network status.

SolarWinds provides two types of Packet Analysis Sensors to monitor and analyze the network traffic (Solarwinds, 2014). Packet Analysis Sensors for Networks (network sensor) collect and analyze packet data that flow through a single, monitored switch for up to 50 discrete applications per node. Packet Analysis Sensor for Servers (server sensor) collect and analyze packet data of specific applications that flow through a single node. After a sensor is deployed and configured, it captures packets and analyzes them to calculate performance metrics for the monitored applications. An included communication agent allows the sensor to send back sampled packet data to the server, which include volume, transactions, application response time, and network response time for each application on a node. The packet data are then saved to the database.

Statistics collection retrieves data from the selected device, interface traffic usage. Statistics represent summaries, peaks, and averages collected over a period of time. Statistics do not represent only the activity occurring at the time of the collection, but are calculated values summarizing total activity since the last collection.

- Packet Statistics is collected from the Node every 10 minutes.
- Packet Statistics is collected from each Interface every 9 minutes.
- Detailed statistics is summarized into hourly statistics after 7 days

As depicted in figure 2.4, the following procedure provides an outline of how SolarWinds NPM monitors network performance (Solarwinds, 2014).

- After Network Sonar Discovery has populated the SolarWinds Database with the network objects want to monitor, object information is passed to the Business Layer.
- The Business Layer passes node and volume information to the Collector Polling Controller and provides licensing information to the SolarWinds Information Service (SWIS).
- The Collector Polling Controller creates the required polling jobs and then passes them on to the Job Engine v2.
- The Job Engine v2 performs requested polling jobs, using SNMP, ICMP and WMI, as configured in Network Sonar Discovery.

- The Job Engine v2 then passes the results of all requested polling jobs to the Collector Polling Controller.
- The Collector Polling Controller places all polling results into the Microsoft Message Queue (MSMQ).
- The Collector Data Processor pulls polling results from the MSMQ, and then performs the following operations:

  a. The Collector Data Processor performs any required calculations, and then inserts these "cooked" results into the SolarWinds database.

  b. The Collector Data Processor checks with the SolarWinds Information Service (SWIS) for any existing dependencies that are defined for the polled nodes.

  c. The Collector Data Processor checks polling results against existing basic alert definitions to determine if any basic alerts and corresponding actions should be triggered.

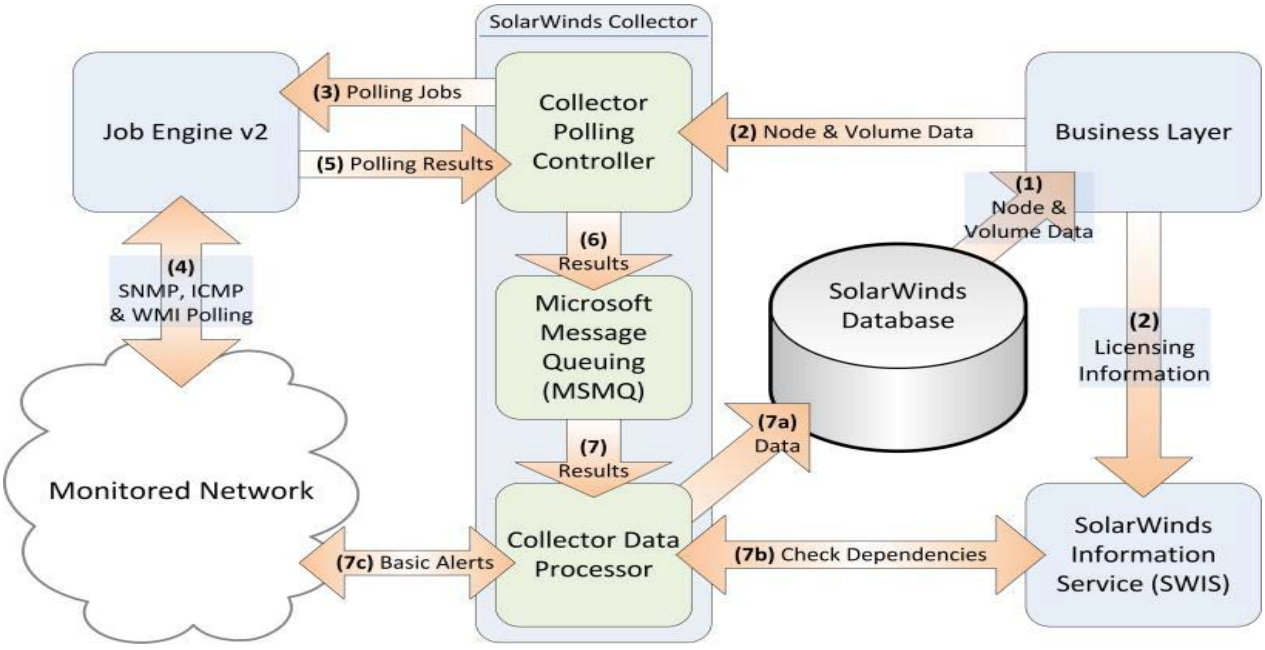

Figure 2.5: An outline of how NPM monitors the network (Solarwinds, 2014).

In this study, Solarwinds network performance monitor toolset is used for analyzing and evaluating AAU wide area network performance because of the following five advantages. Engineer's

Toolset delivers an advanced collection of monitoring, discovery, diagnostic, and Cisco® tools. Here are the top five reasons to use SolarWinds Engineer's Toolset:

- All the network tools you need in one complete package
- Monitoring tools include Real-Time Interface Monitor, SNMP Real-Time Graph, and more
- Diagnostic tools include Ping Sweep, DNS Analyzer, Trace Route, and more
- Network discovery tools include Port Scanner, Switch Port Mapper, Advanced Subnet Calculator, and more
- Cisco management tools include Real-Time NetFlow Analyzer, Config Downloader, and more

2.13. Review of Related Works

There are many studies done to explore the issue of Network performance globally and locally so as to evaluate network utilization and customer satisfaction.

2.13.1. Global Related Works

Kansanan (2009) defines several factors that affect wide area network performance. The focus of the study is evaluating the performance of commonly used application protocols such as TCP/IP in organization networks. After identifying the performance problems in organization WANs, the study concentrates on methods for improving WAN performance. Accordingly, WAN acceleration is proposed as a possible solution for improving WAN performance. The different acceleration methods are discussed on how the accelerators can improve. He conducted his research in Lappeenranta University of Technology, Finland.

Xu (2012) conducted his research on Evaluation of Wireless Network Performance in a Multi-Nodes Environment. The purpose of his research was to find the impact of increasing wireless nodes to the performance. Simulation tool was used to measure the network performance parameters such as throughput, jitter, packet drop rate, and delay. He compared the result obtained from the performance evaluation of test bed with the simulation tool (OPNET).The simulation tool is the better solution than the test bed that simulates hardware of networking environment and

monitors the traffic. In his experiment, Xu (2012) found that the throughput of OPNET simulation is always higher than any operating systems from test-bed experiment with the packet size increasing from 128 to 1408 bytes. Experiments have been done on different access points (APs), such as throughput and media access delay and the following result reported.

Throughput: The throughput increase with the increasing of packet size through all numbers of access points. The difference of throughput between each number of APs also increases during the increase packet size. The average throughput of each numbers of APs decreases when the numbers of APs climbs up from 20 nodes to 80 nodes. Delay: The delay decreases with the increasing of packet size through all numbers of APs. The difference of delay between each number of APs also decreases during the increase of packet size. The average delay of each numbers of APs increases when the numbers of APs climbs up from 20 nodes to 80 nodes.

Media Access Delay: The media access delay decreases with the increase in packet size through all numbers of APs. The difference of media access delay between each number of APs also decreases during the increase in packet size.

Manini (2005) conducted his research on Performance Evaluation of Complex Network Applications in Universita' Degli Studi Di Torino, Italy. He used a Fluid Stochastic Petri Net Model to evaluate performance of P2P file sharing applications of complex networks. Fluid model approach is possible to analyze performance, in terms of transfer time distribution, tuning many system parameters with low computational cost. P2P file sharing applications relay are cooperation among users and resource sharing. The main advantage of file sharing applications is the capacity to share large amounts of data without the necessity of expensive resources. This study found out the performance problems based on Cooperation, user dynamics, client-side user behavior and server-side peer behavior.

Alisa (2013) studied the performance of wireless networks using a performance tool called OPNET-14.5 simulator and various tests on it have been conducted. Performance of the proposed system is evaluated with different scenarios. Using optimized network engineering tool, OPNET 14.5 modeler, over four major physical layer technologies Infra-Red, Direct Sequence Spread Spectrum DSSS, Frequency Hopping Spread Spectrum FHSS and Orthogonal Frequency Division

Multiplexing OFD Mat multiple transmissionrates,1Mbps, 2Mbps, 5.5Mbps, 6Mbps, 9Mbps, 11Mbps, 36Mbps, 48Mbps, 54Mbps. DSSS can use 1,2,5.5,11 Mbps rates. FHSS and IR are able to operate at 1or 2 Mbps rates, while OFDM is capable to operate at 6, 9,11,36,48 or 54 Mbps rate. A trade off exists between the selected data rate and the physical technology. It is founded that at some transmission rate, the OFDM technology delay is better than in DSSS. FHSS delay is less than DSSS delay. IR delay is better than FHSS and DSSS delay. In addition, for one physical layer technology the delay can be reduced by increasing the transmission rate of the channel. She conducted her research in Baghdad University Electrical Engineering Department, Iraq.

Malhotra, Gupta & Bansal (2011) attempted Performance analysis of the Wireless and wired computer networks through simulation using OPNET as simulating tool. For wired networks, the performance parameters like delay and throughput have been investigated with varying transmission links and load balancer. The load-balancing has been analyzed through parameters like analysis of traffic sent and traffic received, while in wireless networks the metrics like delay, retransmission attempts and throughput have been estimated with varying physical characteristic and buffer size. From the obtained results, it is gathered that performance of the wired networks is good if high speed Ethernet links like 1000 Base X and server-load balancing policy are used whereas the performance of Wireless LAN can be improved by fine tuning and properly choosing the WLAN parameters. For the tested simulation scenarios the performance is observed to be better with wireless networks using infrared type physical characteristics and higher buffer size (1024Kb).

2.13.2. Local Related Works

Yacob (2013) investigated the impact of various WAN factors and developed a WAN optimization model that can improve the performance of business critical applications over the WAN enterprise environment. The developed WAN optimization model was aimed to give a solution for WAN enterprise environments given that the resources are from the company. This WAN optimization model is developed based on the experimental result of the study and considering of the selected metrics. The researcher has done his experiments in the head office of UNECA WAN environments, Addis Ababa, Ethiopia. In his experiment, he followed four steps. The first step is

collection of application's traffic flow data by using a tool called Net flow analyzer, the second step was to simulate and experiment the collected traffic flow data by using a tool called OPNET modeler, the third step was analysis of the simulated traffic data to investigate the challenges of the critical applications over the WAN by using different optimization methods. Finally, he developed of WAN Optimization model which is designed to identify the real-time bottleneck status based on the results of the analysis.

Gezahegn (2007) has done his research on Performance evaluation of wireless sensor network routing protocols for critical condition monitoring applications. In his work he discussed about the evaluation of the performance of different routing protocols selected from wireless sensor network (WSN) and mobile ad hoc network (MANET) for critical conditions monitoring applications with the help of important metrics. He found out that, the result of multi-path and multi-speed (MMSPEED) routing protocol has a very good performance of events delivery within a short end-to-end delay and high ratio of packet delivery by creating a reliable path for packets reaching to the destination in case of critical conditions monitoring applications with the limitation of moderate power dissipation when compared to other routing protocols.

Besides the above studies, different scholars attempt to investigate AAU network from the point of users browsing behavior and satisfaction (Tadele, 2011, Awet, 2011).

From the forgoing discussion, an attempt has been done to figure out the conceptual foundation related to WAN and its associated protocols such as IP, TCP, MAPI, SMTP ICMP, SNMP, MIB and WMI and performance metrics such as bandwidth, network availability, latency, packet loss and response time. Besides, the related works show that such research of the current type was given a minimal attention. What is more, the literature review was helpful to frame the methodology for this work such as how to use a performance monitoring tool and what are the performance metrics which affects the network performance such as bandwidth, packet loss, latency and delay, congestion and response time.

As to the researcher knowledge there is no study conducted to evaluate the performance of AAU wide area network. Hence, this research aimed to investigate the overall performance of the AAU network using network performance monitor toolset and propose a WAN optimization framework that can improve the performance of the network.

CHAPTER THREE

CASE DESCRIPTION AND ANALYSIS TOOL

In this study, business understanding is undertaken over the AAU WAN environment using methodologies and techniques described in chapter one. AAU WAN infrastructure and the AAU WAN architecture is taken from the official AAU website (AAU, 2014). In addition to this, an interview was conducted with the AAU ICT experts about the current status of the network performance problems and tried to access what are the applications and services hosted in the AAU data center.

3.1. Network Infrastructure of AAU (LAN/WAN)

AAU has fourteen campuses with a wide geographic distribution in and around Addis Ababa. All but the recently acquired Akaki campus have had inter-campus connection in the form of a hybrid wired and wireless connection for at least three years. There is a current effort underway, in collaboration with ETC, to convert the somewhat limited wireless connections to fiber. The main campuses (Sidist Kilo, Amist Kilo, and Arat Kilo) serve as the core of the network with redundant high speed connectivity.

Within the various campuses there are numerous existing and new buildings that do not yet have access to the network. Connecting more and more buildings (and rooms within the buildings) is an ongoing process. Today there are about 6,000 nodes connecting end-users to the network. These nodes include administrative and academic staff offices as well as computer labs.

The connectivity devices (routers, switches, etc.) which were predominantly CISCO devices in the past were upgraded and converted years ago to more capable, predominantly Huawai (a Chinese brand) devices through a donation form the Chinese company. The internal network (LAN/WAN) is connected to the internet via a 100Mbps link to the ETC exchange. The connection is managed internally through a gateway and protected from intrusion and virus and other attack by two firewalls. The existing network interconnection in Addis Ababa University (AAU), shown in figure 3.1 below has fourteen campuses with wide geographical distribution in and around Addis Ababa.

3.1.1. AAU Network Architecture

Addis Ababa University has three major campuses (Main Campus, Technology Campus, and Science Campus). These campuses form the core network and connected via fiber network. The remaining campuses are connected with virtual private network (VPN) provided by the national service provider, the Ethiopian Telecommunication Corporation (ETC). Addis Ababa University (AAU) has adopted information and communication technology (ICT) resources as strategic tools in advancing its mission of learning, teaching, and public service. The other eleven campuses are connected to the network through Sidist Kilo using Ethiopian Telecommunication Corporation Virtual Private Network (ETC VPN). Each of these campuses has core switches and local Servers.

A virtual private network (VPN) is a technology used to access private networks from a remote host that is not physically connected to the private network. VPNs can also be used to interconnect two or more private networks in order to build one larger virtual network. Since VPNs are much cheaper than leased lines, it is commonly used to interconnect offices spread out all over the world with an intranet. VPNs can be used for both data and voice communications hence it is of interest for the telecom business. VPNs does however not necessarily offer any security which makes the connection between networks and hosts fragile for malicious attacks, such as packet sniffing and identity spoofing. In order to fill these security gaps, various techniques and protocols are currently being used by different vendors (Englund, 2010).

As such, the proper integration, use, and management of ICT resources have become vital to the success of the university. Proper integration, use, and management of AAU's ICT resources entails, among others, equitable sharing of their limited capacity, protection of sensitive information to which they provide access, prevention of abusive practices enabled by their use, and ensuring their manageability through technology standardization.

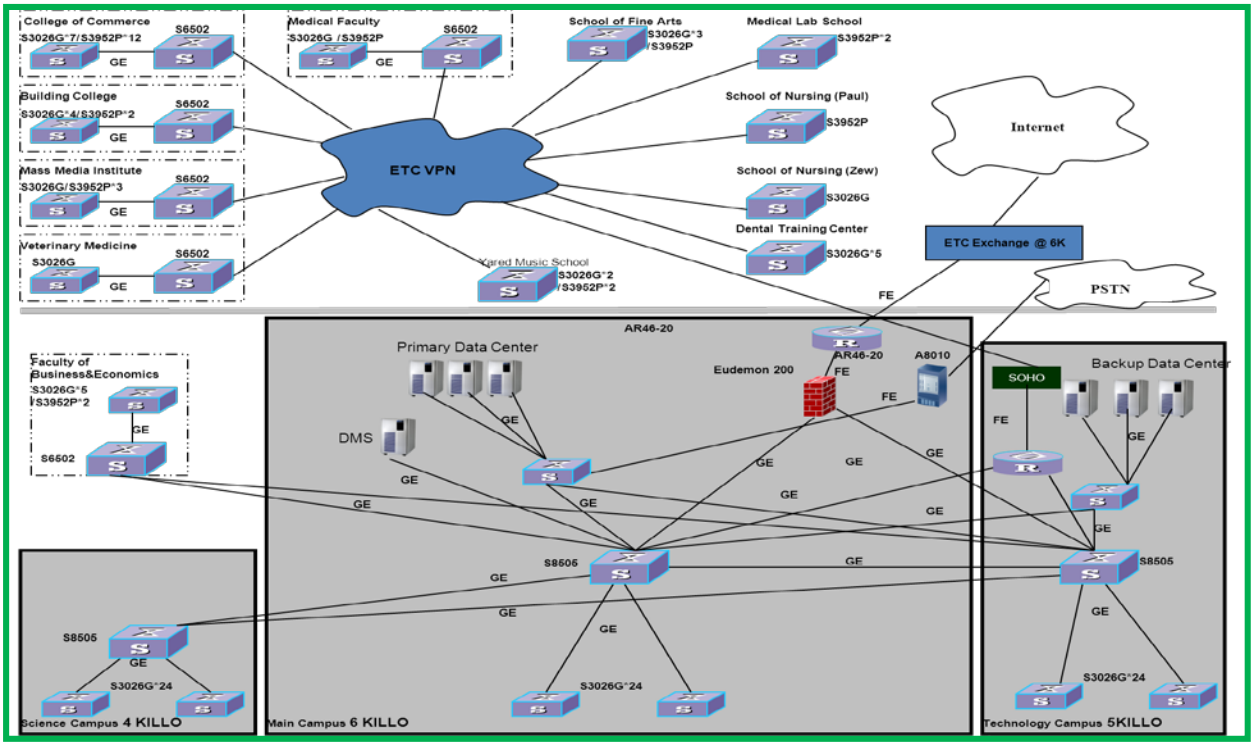

Figure 3.1: AAU Network architecture

3.1.2. AAU Data center

A data center (DC) refers to any large, dedicated cluster of computers that is owned and operated by a single organization. Data centers of various sizes are being built and employed for a diverse set of purposes today. On the one hand, large universities and private enterprises are increasingly consolidating their IT services within on-site data centers containing a few hundred to a few thousand servers. These service providers also employ some of their data centers to run large-scale data-intensive tasks, such as indexing Web pages or analyzing large data-sets, often using variations of the Map Reduce paradigm (Benson, Akella and Maltz, 2010).

The servers used to run the network and that provide other services (such as mail and web services) are housed in two main datacenters (AAU, 2014). The approach followed under the recent network upgrade was to have one main datacenter at Sidist Kilo with a backup and disaster recovery center at Amist Kilo. While there are about 35 servers (Sun & HP) of various specifications operated by the ICT office, there are numerous others scattered throughout the university under different

faculties, institutes, and administrative offices. This situation, while justifiable in rare cases, is not in line with best practice. The AAU ICT Development Office is making attempts to consolidate server maintenance by trying to persuade the various stakeholders that they are better served by having their servers hosted in the main datacenter.

Figure 3.2: AAU Data Center

3.1.3. AAU WAN Applications and Services

The services delivered by and through the ICT infrastructure, aside from the provision of connectivity, may be grouped into two. First, there are technical services that provide indirect support to the user community. These are services such as Domain Name Service (DNS), proxy servers for facilitating managed internet access, LDAP for identity management, and firewall for security. The second set of services is made up of services that provide direct support to end-users in their normal day-to-day activities. These are such services as email, internet, automation, e-learning, etc.

At AAU all the essential technical services are in place and operational. Of the direct support services, the most widely used and the ones that have been in place for some time are email and

internet access. There is limited automation in the form of registrar automation and library automation. There have also been previous attempts at finance automation that have not been successfully implemented. There is an initiative that has been underway for a little over a year to select and deploy an integrated and enterprise-wide automation (business process and student administration automation).

Different services and applications are consolidated in the AAU data center, housed in Sidist Kilo campus. And hence, the WAN is being used by different services and applications like Web Service using Joomla, e-mail Service using Google, e-Learning Service using Moodle, Doman Name System using BIND, Proxy Server, Library System, eGranary/Internet in a BOX, Video Conferencing, DHCP Service, FTP Service, UCIS /University College Information System, Sun Ray Server / Thin Clients and CISCO Networking.

3.2. Performance of AAU-WAN

Christner, Seils, and Jin (2010) asserts that before examining how to improve network performance challenges caused by WAN factors, it is important to deal with vital issues like identifying factors, which are the major source of performance problems for that specific WAN environment, and determining the performance reaction of each enterprise application for these factors.

Based on the results of the interviews which was conducted with end users and five domain experts (see Appendix), and moreover, the current practical WAN situation in AAU, the network faces network/application performance problems which affects the operation of the University. The problem includes email delivery delay, a delay to open an application on the web, packets fail to deliver timely, fluctuation of network availability, unable to transfer files using FTP are the network services/applications performance problems. Moreover, the results of the interview shows that the overall AAU network is underperformed. There are different reasons that contribute for the degradation of the AAU network. One possible reason is the AAU WAN infrastructure such as the router, servers, and switches in the data Center is old enough to forward network traffics. The layout of the internal network and devices attached to it also impact the network performance. Non-business Related Activity - In many cases, non-core activities occurring within the organization are found to be the cause of high network utilization. Casual web surfing, Internet

radio streaming, and viruses that have infected machines on the network can steal precious network resources and cause problems with critical business functions.

Other possible reason is the lack of network security management in the university that resulted in network congested because of high bandwidth usage by viruses. Organization's networks are susceptible to virus and worm attacks. These attacks can create a potential traffic nightmare for the routers and switches. Lack of proper cable Installation can also be the reason for network latency, since collision among network wires can be a cause to high congestion both at the end user and datacenter. The length of the communication media is another cause for network performance degradation. The AAU network has 14 campuses located at different places which are joined to the university network by VPN are the causes for the network congestion.

Therefore, the AAU network should be managed and monitored in order to know cause of the current network performance status and to identify what the network bottlenecks are by using different performance metrics. Knowing those network bottlenecks enables us to take an action on how to improve the performance of the network and to keep users happy by avoiding these network traffic bottlenecks. The network performance problem would be worse when the AAU WAN are not well equipped with the required network resources to run the different network applications. As new technologies and applications are being emerged, the AAU network may be impacted and therefore the network should be monitored to ensure the optimal performance of the WAN. When network resources like Bandwidth, powerful WAN devices are limited, the impact of WAN network factors including latency (Delay), congestion, packet loss, response time, are more significant. AAU network do not have the required financial power to supplement the WAN environment with latest network resources. Considering the limitations of the required resources in AAU Network, it is a hot issue to investigate and improve the performance problem.

Selection of Performance Metrics

Network Performance monitor toolset uses response time, packet Loss and network availability for analyzing network performance. After the researcher identifies the strategic network performance problems which are collected from an interview with end users and ICT experts, the three performance metrics are selected accordingly to analyze the performance of the AAU WAN.

Therefore, this study uses response time, Packet loss and network availability at the node detail view as a primary performance metrics for evaluation of the AAU network.

Response Time: The causes for longer response time (delay) are the following (Cole and Ramaswamy, 2000): (i) the distance between the nodes (propagation delay). Propagation delay becomes significant in long distance and this affects the overall network latency, (ii) The serialization delay is the time takes to move bits of packets (size of packets, medium, speed of interface) in to the line is another factor for longer response time, (iii) The processing delay like comparing a list of data by the router is also a factor for high response time, (iv) The time spent by routers on deciding where to forward the packet is also contribute for the overall latency. Therefore, average response time is the total time elapsed to measure the above delay types. High latency is the major cause for long network response time over a WAN.

Packet Loss: Packet loss is typically caused by network congestion. When content arrives for a sustained period at a given router or network segment at a rate greater than it is possible to send through, then there is no other option than to drop packets. If a single router or link is constraining the capacity of the complete travel path or of network travel in general, it is known as a bottleneck. Packet loss can also be caused by a number of other factors that can corrupt or lose packets in transit faulty networking hardware, or faulty network drivers. Number of lost packets is determined from the sequence numbers of correct received packets. Packet Loss measures the reliability of a connection. A known chunk of data is sent to the router and then the router is supposed to send the same data back unaltered. In process of pinging, several packets are sent out over the different time interval of seconds. For example, if 10 packets were sent out, but only 8 made it back, then that would be 20% packet loss; so the more packets that are sent, the more accurate the picture of what the actual packet loss is. In a perfect world 0% packet loss is what we all want - every packet we send out makes it to where it's supposed to go. In reality, some packet loss is probably going to happen, but as long as it is under 5% or so you shouldn't even notice. So just remember that the higher the packet loss percentage, the slower the connection will work because in most instances it has to send the same piece of information several times. Congestion causes packet loss and retransmission and that may seriously deteriorate the network performance.

Network Availability: Availability calculations are based on node status. Node availability is 0% when the node's status is DOWN or unknown. Availability is 100% when the node's status is UP. Availability calculations are based on packet. If the node is dropping packets (packet loss), availability may show less than 100% even though Node status is up. Toolset Network Performance Monitor provides two methods for calculating device availability. The default method uses the status of the node, UP or DOWN. As long as the device responds to a ping within the warning interval, Toolset Network Performance Monitor considers the node up. The other method offered bases node availability on packet loss percentage. Unless you need packet loss percentage based availability calculations, it is recommended that you retain the default method for availability calculation.

Thus, this study undertakes experiments to evaluate the impact level of WAN factors such as response time, packet loss and network availability on the entire network performance of the AAU by using network performance monitor. NPM Node Monitor can display a number of useful statistics by capturing and analyzing SNMP data from the selected router for analysis purpose. These evaluation results are then discussed and finally a WAN optimization Framework is proposed as a solution so that it will be effective to mitigate the impact of the analyzed network issues on the performance of WAN of the university.

3.3. WAN Traffic Analysis Tool

According to Hailay (2011), Network management is becoming an increasingly complex task due to the variety of network types and the integration of different network media. As networks become larger, more complex, and more heterogeneous, the costs of network management rise. In this scenario, automated tools are needed to support human effort, gathering information about the status and behavior of networked elements. Network monitoring is the most fundamental aspect of automated network management.

Tools and Performance metrics that are described in the methodology section, chapter 1, is used for experimentation. Network Performance Monitor is used in the data collection, analysis and evaluation phase. The Network Performance Monitor Toolset is a real-time network monitor able to track network latency at the node level and defined in terms of packet loss, Response time and

network availability. Toolset Network Performance Monitor provides two different levels of monitoring. Devices that support SNMP can be monitored for network latency and packet loss, provide traffic statistics, and supply detailed management information. Devices that do not support SNMP can provide network latency and packet loss information.

This experiment uses a Network performance monitor toolset version 9. SolarWinds develops and markets an array of network management, monitoring, and discovery tools to meet the diverse requirements of today's network management and consulting professionals. The Solar Winds Toolset provides the tools need as a network engineer or network consultant to get the job done. SolarWinds was founded by network professionals and continues to design tools for the network professional by listening to and enlisting the help of network professionals.

3.4. Network Traffic data Collection from AAU WAN Link

Understanding the traffic data helps to determine the major performance bottlenecks that affects AAU Network performance. However, the level of performance influence by these potential factors will be experimented in the analysis phase (Chapter 4) to achieve the objective of the study. Configuration of the selected monitoring tool is done in order to capture network traffic data from the selected node by running a series of wizards on the main router of the AAU WAN.

Network traffic Data is collected about various features of the AAU WAN environment, which includes response time, packet loss and network availability. The collected data is used to determine the major performance bottlenecks that affect the AAU's network performance. Charts (Views) provide a number of ways to analyze the collected real-time data. A view consists of network object information provided in an early scanned table that reflects the most recently collected data. The following terms are some of the attributes to measure the performance of the network.

The figure 3.3 below shows the architectural procedure on how the AAU network performance is analyzed and evaluated

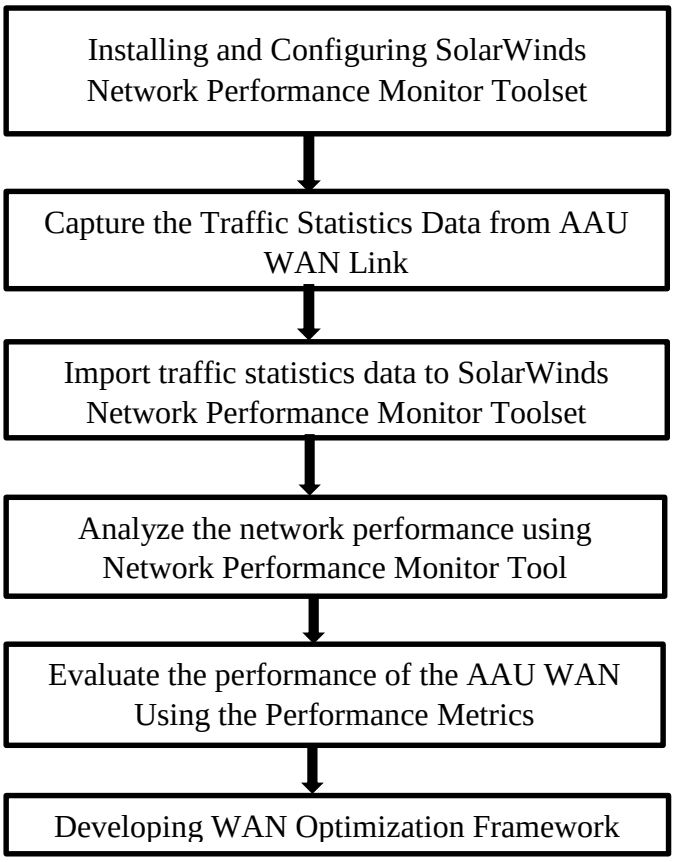

Figure 3.3: Procedural Architecture for WAN performance Analysis of AAU

To analyze the AAU network, the researcher follows the following steps. The first step is installing and configuring the Network performance monitor toolset on the AAU WAN link which is the Router device to ensure that it is exporting Network statistics data. Using a Web console technology on the local machine of the researcher, a network traffic data is captured from the managed device of AAU WAN Router from August 14, 2014 to October 14, 2014 since the 30 days data are representative for analysis. After capturing the network traffic flow from the WAN link, data analysis is done in order to understand the performance of the network using different network performance metrics such as network availability, packet loss, and response time. The analyzed data is interpreted and evaluated to have a clear understanding on how the AAU WAN is performing effectively and efficiency. Network traffic data Evaluation and interpretation is done using network performance monitor toolset. Finally, based on the findings of the experiment, a WAN optimization framework is developed to enhance the performance of the AAU network.

3.5. Overview of Network Performance Monitor (NPM)

To evaluate the performance of the AAU network, different performance metrics are used. The network performance monitor tool supports different performance metrics such as response time, current in/out bps, peak traffic load, percent utilization, bandwidth utilization, CPU utilization, memory utilization, volume utilization and packet loss for analyzing network performance measurements. Therefore, this study uses response time, packet loss, and network availability to analyze the performance of the AAU WAN.

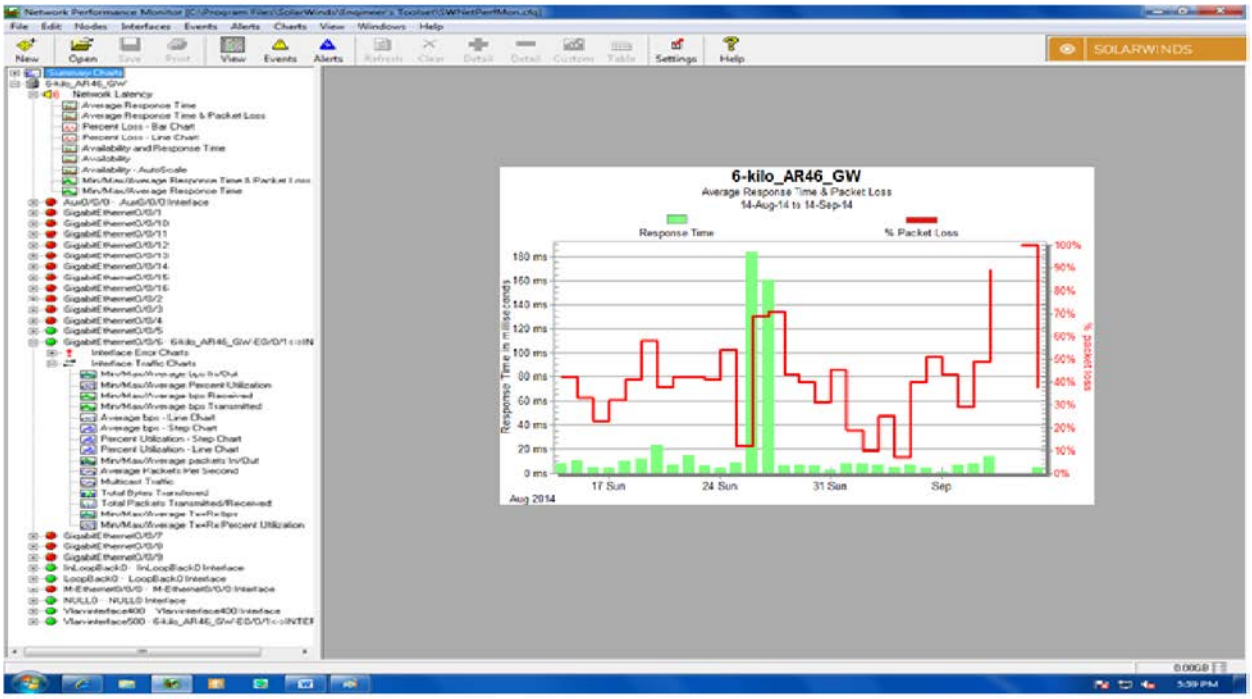

Figure 3.4: Overview of Network Performance Monitor

3.6. AAU WAN NODE

A Node is any Managed network device that is capable of participating in a network management system to capture network traffic data. The node which comes with a variety of resources and charts that provide details information about current the node being monitored. This experiment is done on the real environments of the AAU WAN scenario. To do the experiment on the given environment, a managed device with a name “6-kilo_AR46_GW” which has a unique IP address and a unique community string is added in order to be able to capture a network traffic statistics

data. The Simple Network management Protocol is enabled on the router in order to make the network flow. After the network performance monitor toolset is installed on the researcher's machine, an IP address and a Community string was added to access the remote router so that the real-time traffic flow is collected and analyzed in terms different performance metrics. The analysis is done using network performance monitor tool by examining the traffic flow through the managed device of the routers which is SNMP enabled.

Network Performance Monitor nodes typically include routers, switches, e-mail servers, application servers, and workstations. Nodes do not need to be on the same network, though it must be reachable from the computer on which Network Performance Monitor is installed.

IP Address	Community String	Node name	Router Type	Machine Type	Network Status	Status	Status Description	Current Response Time
10.1.0.6	aauro	6-kilo_AR46_GW		HUAWEI Technology Co.,Ltd	Down		Node status is Down, One or more interfaces are in an Unknown state.	No response
10.1.0.6	aauro	6-kilo_AR46_GW		HUAWEI Technology Co.,Ltd	Up		Node status is Up, 'GigabitEthernet0/0/4' is Down.	10.1.0.6

Table 3.1: Node selected for minoring

Quidway_AR_46 Series Routers are high-performance edge routers independently developed by Huawei Technologies for service providers and enterprise networks. Depending on the number of provided slots, the AR 46 series falls into three models: Quidway AR 46-80, AR 46-40, and AR 46-20. The AR 46 series can be deployed on networks of different sizes to provide different functionality. The series can function as edge routers on a provider IP network, high-performance backbone routers on a provider support network (such as a management network), or core routers or high-performance distribution routers in an enterprise network (H3C Technologies, 2014).

The router Node "6-kilo_AR46-GW**"** is selected by the researcher for the purpose network traffic data capturing using the Network performance monitoring tool. The following performance metrics are displayed in the form of charts and tables about the performance of the entire AAU WAN in terms of Network Latency at the link level**.**

- Average Response Time
- Response Time and Percent Loss
- Percent Loss – Bar Chart
- Percent Loss – Line Chart
- Availability and Response Time
- Availability
- Availability – Auto scale
- Min/Max/Average Response Time and Percent Loss
- Min/Max/Average Response Time

Therefore, this study uses response time, packet loss and network availability as a primary performance metrics to evaluate the effects of different WAN factors on network performance of the AAU at the link level. To select these performance metrics, the researcher of this study identifies the strategic network performance problems which are collected from an interview with end users and ICT experts.

CHAPTER FOUR

RESULTS AND DISCUSSIONS

Traffic flow data is collected from the 6-kilo_AR46-GW Router, which is the HUB for all the network campuses of AAU. Both the incoming (WAN IN) and outgoing (WAN OUT) interfaces of the router are configured to send traffic to NPM. Traffic flow statistics data is collected from the AAU WAN for about 30 days of time interval (from August 14, 2014 to September 14, 2014) and the statistics charts and tables are presented in the following section 4.1. Node Views provide a number of ways to analyze the real-time data and provided in an easily scanned table that reflects the most recently collected data.

4.1. Performance Analysis of Network Latency at AAU WAN Link

The analysis is done using a performance metrics such as response time, packet loss and network availability at the selected node. Network latency is the delay that is introduced by the network. Network latency can be measured either one-way (the time elapsed from the source sending a packet to the destination receiving it), or round-trip (which is the sum of the one-way latency from source to destination plus the one-way latency from the destination back to the source). In other words, Network latency is the time to deliver an entire data unit (packet) from one host to the other. When studying network performance, the most commonly used form of latency is the Round Trip Time (RTT). In order to analyze the AAU WAN link measured traffic, the researcher highlights a representative of 30 days of data capturing time in to four weeks. Week1 is from August 14-21, 2014, week2 is from August 22-28, 2014. Week3 is from August 29, 2014 to September 06, 2014 and week4 is from September 07-14, 2014.

4.1.1. Average Response Time

Average response time is the total time it takes after the client sends a request till it gets a response from the server. Response time report represents the values of the server response time. Average and maximum response times are the most important characteristics of measuring network performance.

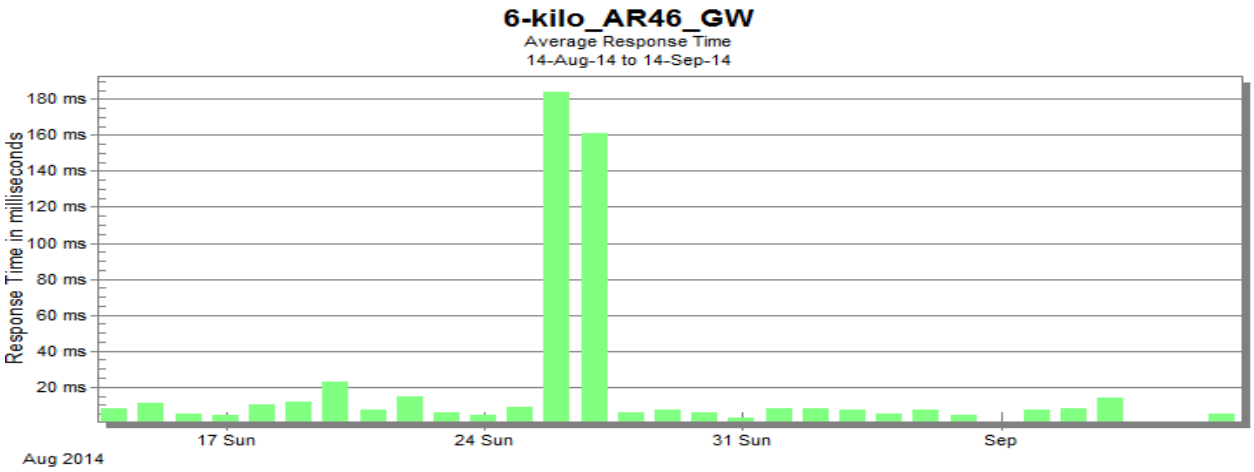

Figure 4.1: Average Response time

Average Response time (14-Aug 14 to 14-Sep-14)	**Week1 Aug 14-21**	**Week2 Aug 22-28**	**Week3 Aug 29-Sep 06**	**Week4 Sept 07-13**
Minimum Response Time	4.29	5.88	3.48	4.8
Maximum Response Time	23.91	184.94	8.39	15.78

Table 4.1: Maximum and Minimum Response Time

The above figure 4.1 and table 4.1 show the average response time in milliseconds for 30 days of data capturing period; from August 14, 2014 to September 14, 2014. During week1, the maximum average response time observed is 23.91 ms and the minimum response time recorded is 4.29 ms. During week2, a maximum of 184.67 ms and a minimum 5.88 ms of average response time is recorded. In week3, a maximum of 8.39 ms and a minimum of 3.48 ms is recorded. In week4, a minimum of 4.8 ms and a maximum of 15.78 ms response time is observed. From the above result, it is seen that, there is a high latency that causes noticeable delays during the week2 (August 26-27, 2014) and at this time the network was so overloaded due to the high utilization of the network by different application users. In September 7/10/11/12/13, 2014, the response time is 0.00 ms. This shows the response time was not processed because of different reasons such as network unavailability, shortage of electricity, or internet service failure from the service provider. The minimum average response time observed is 3.48ms during the week3 (August 31, 2014). At this day of the week, the network was at its lowest latency (high speed) since there was less number of network users or application users. At weekend (Saturday and Sunday), there are less number of

network users. Generally, the lowest response time observed is during the weekend especially on Sunday this is because the network is underutilized by the network users found at different campuses of the university. Therefore, one can say that the number of application users are the cause for network congestion and resulted in high latency which is the major cause for long network response time (Delay) over a WAN.

4.1.2. Percent Loss

When data is transmitting over computer network, one or more packets may fail to reach their destinations, and this is packet loss. In other words, Packet loss is the number of packets that fail to reach the destination. Packet loss is the ratio of the number of packets lost to the total number of packets transmitted.

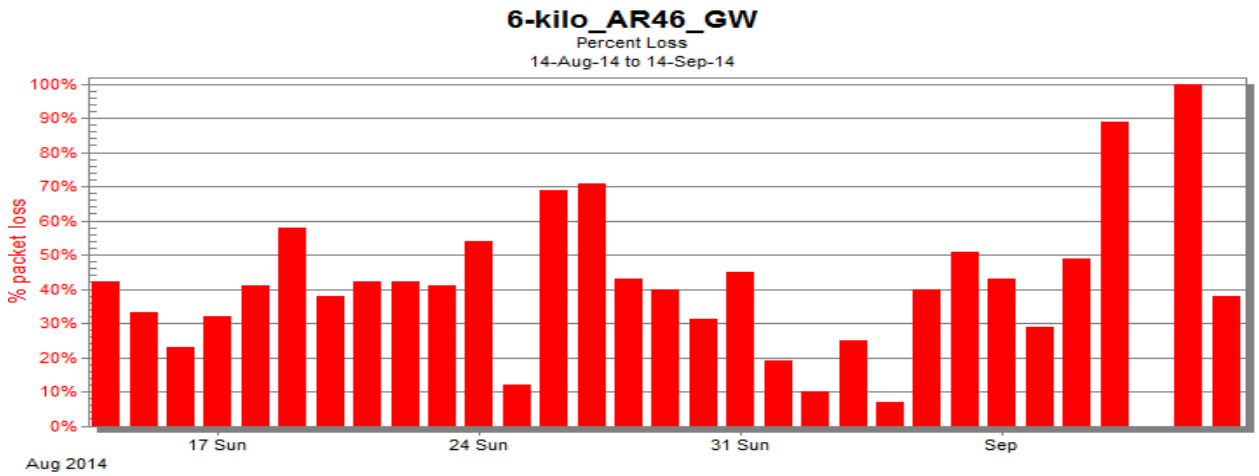

Figure 4.2: Packet Loss in Percentage

Packet Loss Rate (14-Aug 14 to 14-Sep-14)	Week1 Aug 14-21	Week2 Aug 22-28	Week3 Aug 29-Sep 06	Week4 Sept 07-13
Minimum Packet Loss	24%	13%	8%	28%
Maximum Packet Loss	58%	72%	52%	88.8%
Average Packet Loss	41%	42.5%	30%	58.4%

Table 4.2: Packet Loss in Percentage

The above figure 4.2 and table 4.2 present the packet loss rate observed in the AAU main router, located at the main campus (6kilo). In Week1, The maximum packet loss observed is 58% in the day of August 19, 2014 and the minimum packet loss recorded is 24% in August 4, 2014. In

Week2, a maximum of 72% and a minimum of 13% packet is lost. In Week3, a maximum of 52% and a minimum of 8% packet is lost. On week4, a maximum of 88.8% and a minimum of 28% of packet is lost. During the week4, a maximum data is lost due to network congestion. 100% packet lost happens when there was no traffic flow through the router due to network breakdown. 0% packet lost means there was no network availability. On average, a maximum of 67% and a minimum of 18% of data is lost during the 30 days of time interval. The reason for this packet lost is (Section 2.8.6) link failure, when all bits currently in transit on that link is lost or network breakdown, and buffer overflows (congestion) when the network router are temporarily sent more packets than their buffers can accommodate. Other reasons for packet loss is due to weak signal and collision at the link layer.

4.1.3. Average Response Time and Packet Loss

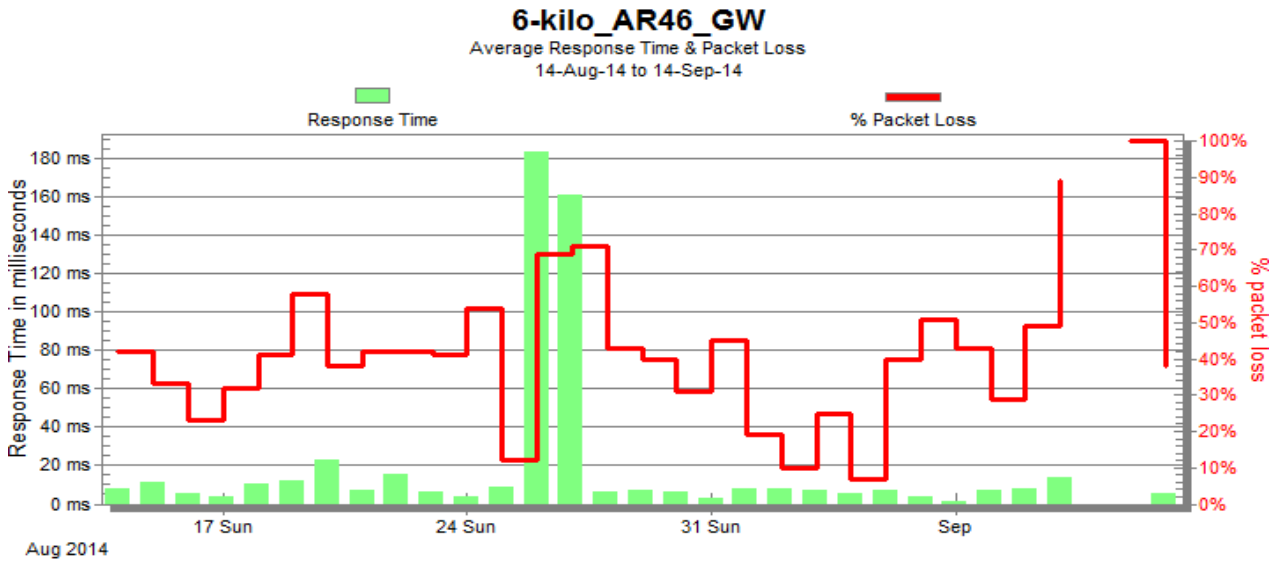

Figure 4.3: Average Response time and Packet loss

Response time & Packet Loss rate (14-Aug 14 to 14-Sep-14)	Week1 Aug 14-21	Week2 Aug 22-28	Week3 Aug 29-Sep 6	Week4 Sept 07-13
Minimum Response Time	4.29	5.88	3.48	4.8
Maximum Response Time	23.91	184.94	8.39	15.78
Minimum Packet Loss	24%	13%	8%	28%
Maximum Packet Loss	58%	72%	52%	88.8%

Table 4.3: Average Response time and Packet loss

The above figure 4.3 and table 4.3 of the bar charts display the average response time and packet loss rate over a period of 30 days of time interval from August 14, 2014 to September 14, 2014. The graph in green color indicates the average response time in milliseconds and the red color indicates the packet loss rate in percentage. The highest response time is recorded 184.94 ms at the beginning of the working days (Monday and Tuesday) and at this time a maximum of 72% of packet is lost. When an application takes long response time to reply to the request at the link level, the probability of occurrence of packet loss is high due to the network congestion or network overloaded. This happens because of the reasons described above which contributes for the high response time and high packet loss rate.

4.1.4. Network Availability

Availability is a Measure of what percentage of the time a network resource is available for use. It is clear that, high availability is better because down time is not welcome.

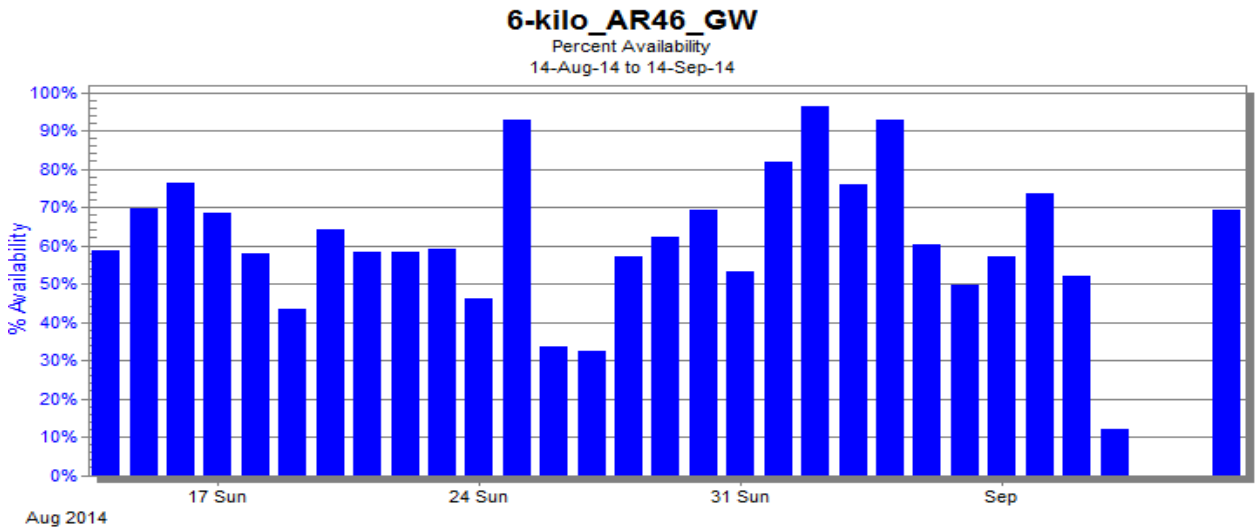

Figure 4.4: Network Availability in Percentage

Network Availability (14-Aug 14 to 14-Sep-14)	Week1 Aug 14-21	Week2 Aug 22-28	Week3 Aug 29-Sep 06	Week4 Sept 07-13
Minimum Availability	43%	33%	48%	10%
Maximum Availability	75%	94%	96%	65%

Table 4.4: Network Availability in Percentage

The above figure 4.4 and table 4.4 show the network availability during the 30 days of data capturing phase. In the week1 of this phase, a maximum of 75% and a minimum of 43% of network was available. In week2, a maximum of 94% and a minimum of 33% of network service was available. In week3, 96% was the maximum and 48% was the minimum was seen in network availability. In week4, the network was in its maximum of level 65% and minimum level of 10% available. One can say that, the highest availability is seen at the end of the week (Sunday and Saturday) since there are few services and application users of the AAU network which results in less packet lost. 0% network availability is to mean the network was unavailable. This could be due to the shortage of electricity at the Data center or a network failure from the internet service provider or it might be the router's problem which broadcasts the network traffic to all campuses of the university. On average, the network was 80% available. The maximum availability recorded was 96% and the minimum network availability observed was 10%. The reason for less network availability is due to the high latency and network overloaded (congestion) and this results for the occurrence of packet loss.

4.1.5. Average Response Time and Availability

The AAU's day-to-day operation relies heavily on uninterrupted network availability. To achieve this, it is necessary to monitor all network devices to make sure that the AAU critical IT processes continue to function. It is also a must to frequently check device availability statistics and traffic details across the network in order to stay running at peak performance. Diagnostics play an important role in maintaining high network availability by monitoring the performance of routers, switches, workstations, servers, or any other network device by using NPM. The network must deliver 100% availability of all network resources and to keep a close eye on the performance of every device, server and application in your infrastructure while, of course, everything is constantly changing.

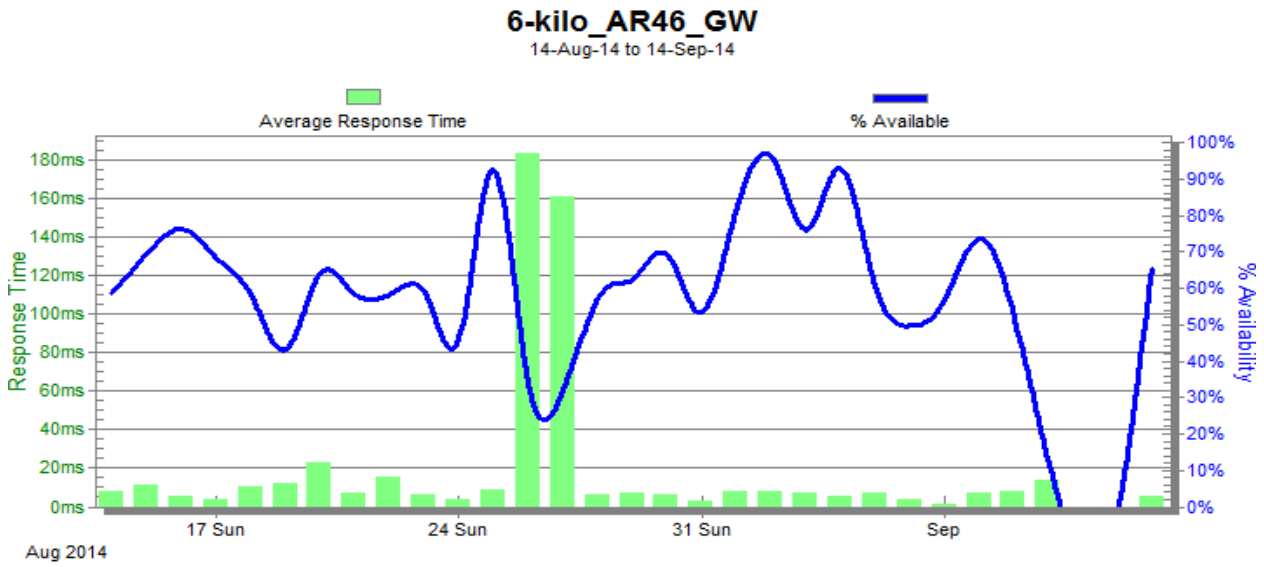

Figure 4.5: Average Response time and Availability

Response Time & Network Availability (14-Aug 14 to 14-Sep-14)	Week1 Aug 14-21	Week2 Aug 22-28	Week3 Aug 29-Sep 06	Week4 Sept 07-13
Minimum Response Time	4.29	5.88	3.48	4.8
Maximum Response Time	23.91	184.94	8.39	15.78
Minimum Availability	43%	33%	48%	10%
Maximum Availability	75%	94%	96%	65%

Table 4.5: Average Response time and Availability

The above figure 4.5 and table 4.5 show the average response time and network availability for about 30 days of time interval. In the above figure, the graph in blue color shows the network availability and the graph in green color shows the average response time. The maximum average response time is 184.94ms when the minimum network availability is 33%. Practically, when the network availability is going high, an average response time is slowing down and results for less packet loss rate. On the other side, when the network availability is less, the probability of response time is high due to network congestion and high utilization.

4.1.6. Average Response Time and Packet Loss

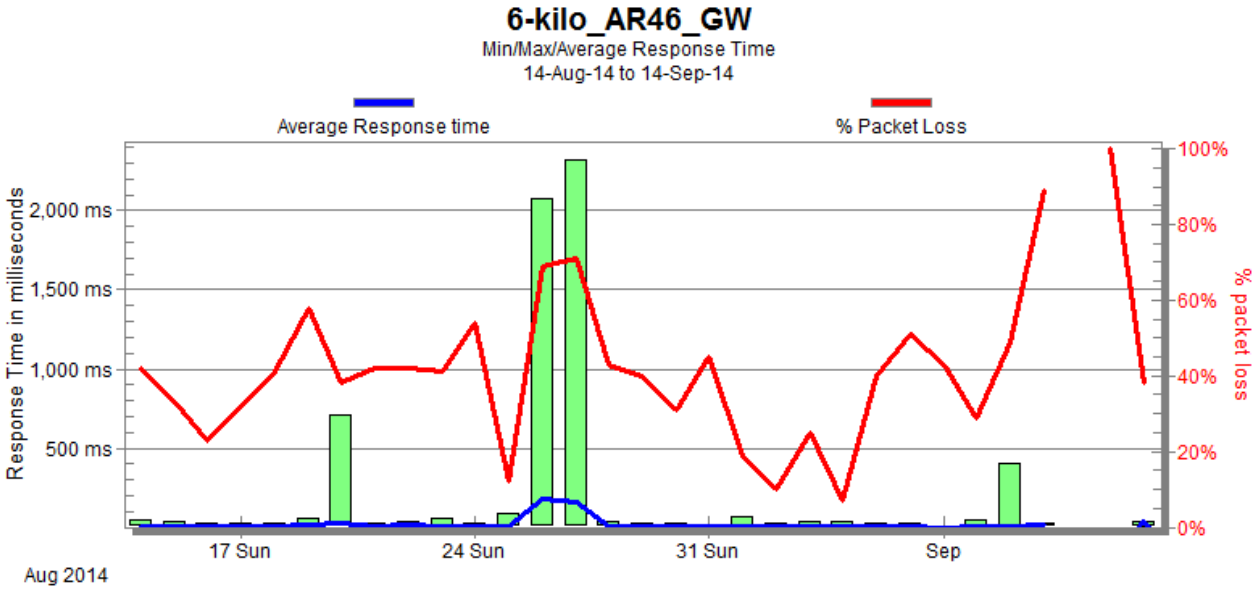

Figure 4.6: The Minimum, Maximum and Average Response Time and Packet loss rate

Response Time & Packet Loss (14-Aug 14 to 14-Sep-14)	Week1 Aug 14-21	Week2 Aug 22-28	Week3 Aug 29-Sep 06	Week4 Sept 07-13
Average Response Time	75	202	50	52
Packet Loss Rate	58%	70%	25%	40%

Table 4.6: The Minimum, Maximum and Average Response Time and packet loss rate

The above figure 4.6 and table 4.6 show average response time and the packet loss rate for 30 days of time interval. The maximum average response time is 202 ms which is seen in week1. This happens due to network congestion and the minimum average response time is 50 ms. The broken line in packet loss and average response time indicates that there was no network traffic follow to and from the router. The maximum average response time recorded is during the working days (Tuesday and Wednesday) since the network is occupied with large number of network application users throughout the university and this is resulted to a high utilization of the network. The minimum response time is observed during the weekend (Saturday and Sunday) because the network is at its low latency due to less number of network application users. As shown in the figure above, when the network exhibits long average response time, a high packet is loss.

4.2. Summary of Findings for AAU Network

Network latency, a synonym for delay, is an expression of how much time it takes for a packet of data to get from one designated point to another. Latency is measured by sending a packet that is returned to the sender and the round-trip time is considered the latency. In this experiment three performance metrics such as response time, packet loss and network availability are used to measure the Network Latency of AAU network.

The lowest response time observed is during the weekend mainly on Sundays because the network is underutilized by the network application users of the university's campuses and at this time the network is in its lowest latency. Therefore, one can say the number of application users are the cause for network congestion and resulted in high latency which is the major cause for long network response time over a WAN. On average, a maximum of 67% and a minimum of 18% of data is lost during the 30 days of time interval. The reason for this packet lost is link failure, when all bits currently in transit on that link is lost or network breakdown, and buffer overflows (congestion) when the network router are temporarily sent more packets than their buffers can accommodate. Other reasons for packet loss is due to weak signal and collision at the link layer. The highest response time is recorded 184.94 ms in one of the starting of the working days and at this time a maximum of 72% of packet is lost. When an application takes long response time to reply to the request at the link level, the probability of occurrence of packet loss is high due to the network congestion or network overloaded. This happens because of the different causes which contribute for the high response time and high packet loss rate. The highest availability is seen at the end of the week (Sunday and Saturday) since there are few users of the network which results in less packet lost. 0% network availability means that the network was unavailable. This could be due to the luck of electricity at the data center or the failure from the internet service provider or it may be the router's problem which propagates the traffic. On average the network was 80% available. The maximum availability recorded was 96% and the minimum network availability observed was 10%. The reason for less network availability is due to the high latency and network overloaded (congestion) and this results for the occurrence of packet loss.

The maximum average response time is 184.94 ms when the minimum network availability is 33%. Practically, when the network availability is going high, an average response time is slowing

down and results for less packet loss rate. On the other side, when the network availability is less, the probability of response time is high due to network congestion and high utilization. The broken line in packet loss and average response time indicates there was no traffic follow to and from the router. This happens when there is shortage of electricity or network breakdown in the datacenter. The maximum average response time recorded is during the working days (Tuesday and Wednesday) since the network is occupied with large number of network users throughout the university and resulted to high utilization of the network. The minimum average response time is observed during the weekend (Saturday and Sunday) because the network is at its low latency because of less number of network users at this time. When the network exhibits long average response time, a high packet is loss.

The results of the interview with the AAU ICT experts and the conducted experiments in this research show similar results. The findings of this experiment shows, the AAU network exhibits high response time, high packet loss rate and less network availability. Therefore, one can say that the AAU network is not satisfied the users need because of the WAN bottlenecks.

4.3. Proposed WAN Optimization Framework

The main objective of this study is to develop a WAN optimization framework which is capable of determining the performance problems and accordingly apply appropriate performance techniques to improve the network performance of AAU. The developed WAN optimization framework is based on the findings of experimental analysis of WAN factors and the interview conducted with the domain experts of the AAU ICT office. The analysis result shows that the AAU WAN has different bottlenecks as discussed in section 4.1. The following bottlenecks are the main source of network performance problems in AAU WAN environment: latency bottleneck and congestion bottleneck.

- **Latency bottleneck**: The problem is caused by long propagation delay and characterized by WAN conditions like long round trip time.
- **Congestion bottleneck**: This problem is due to over utilized link and characterized by WAN conditions like High Packet loss and Fluctuation of network Availability.

The proposed WAN optimization framework features are to identify Network bottlenecks in real-time and apply appropriate techniques for improving the performance of the entire network. Hence, the features of the proposed WAN optimization Framework is further elaborated as follows.

- The optimization framework Monitors and identifies the real-time WAN conditions of the link, and initiate optimization actions when the conditions changes in AAU's WAN link.
- From the identified conditions of the WAN, determine the category of the network bottleneck of the link. Hence, optimization technique is selected based on these categories of the network bottleneck.
- Identify the new network connection requests from the campuses of AAU WAN link at real-time (SYN packets), when WAN conditions changes.
- Select the appropriate optimization techniques based on the determined network bottleneck of the link and apply it to the network server, which is accepting new connection requests at that moment. The optimization techniques are prioritized on their effectiveness to reduce the performance impact of the real-time WANs conditions. Appropriate technique in this context means: the techniques applied should control the effect of the real-time WAN factors to a level that is required by the network traffic that is currently using the link. The networks' demand for various WAN factors has been determined by the experimentation and used to the performance of network. The optimization techniques used to control each bottlenecks category are also selected based on the network performance result achieved by each technique in the experiments conducted on AAU WAN environment.

4.3.1. Desirable Features of WAN optimization Framework

One of the main feature of the proposed Framework is to recognize the real-time WAN conditions of the link before initiating the appropriate optimization action. An intelligent approach to detect and solve network bottlenecks means that the proposed Framework is able to decide for the best decision on the way information is exchanged between the client and the server. The Framework is designed to determine the presence of network bottlenecks listed below which are selected based on the experimentation done as described in section 4.1:

- Latency bottleneck
- Congestion bottleneck

Once the network bottlenecks are determined at real-time in any given WAN link, these rules are used to pick the decision that should be taken by the Framework. The rules will decide on issues like what optimization technique to apply and how to apply it based on the behavior of the network traffic currently using the link.

4.3.2. Components of WAN Optimization Framework

The WAN optimization Framework is designed with four major building blocks to improve the network performance problem over a WAN link, which lacks sufficient resource for providing services for remote campuses of the AAU. As per the study, such kinds of campus WANs introduce various nature of bottlenecks when users try to access and use campus network applications through these WAN links. Thus, this study proposes an optimization Framework that can solve these network performance issues particularly for campus WAN that lacks sufficient network resources to provide resource access for their branch campuses. The major components of the Framework are listed below:

- Network Condition Recognizer (NCR)
- WAN Bottleneck Determiner (WBD)
- Optimization Solution Provider (OSP)
- Event Reporter (ER)

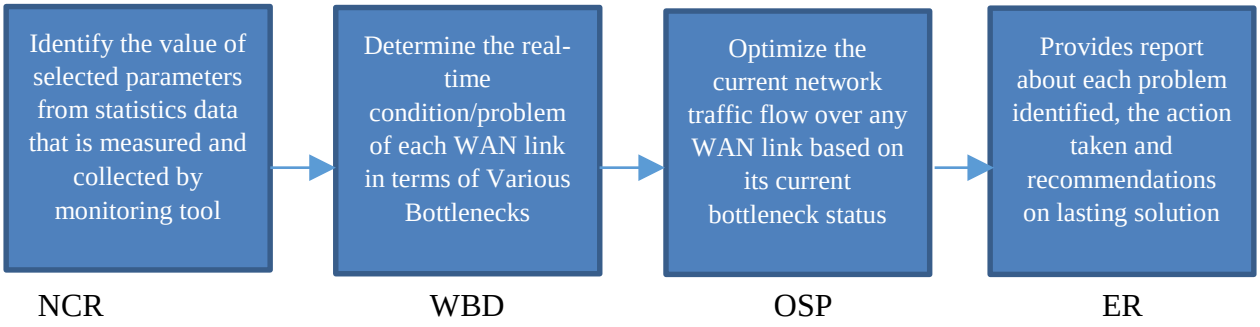

Figure 4.7: Building Blocks of the optimization Framework

Network condition recognizer (NCR)

This component is responsible for monitoring the AAU WAN environment to detect and recognize the presence of any WAN conditions that affects network performance. It also recognizes if there is a change in any of the WAN conditions on each links. The Network condition recognizer periodically monitors the status of the conditions and when it recognizes any change, it provides the information to WAN bottleneck determiner. This component is capable of recognizing any change on various WAN conditions (factors, selected from the AAU WAN environment under the study) in the AAU WAN link which includes, link response time rate, packets receiving rate and latency.

Monitoring link statistics including response time, packet loss and latency helps to recognize the real-time status of WAN factors in the WAN link. And from these real-time link statistics it is possible to determine which performance bottleneck category should be solved to optimize the performance of the network traffic flow which is using the WAN link at that point of time. This component uses link Statistics collector and Timer to accomplish its task.

WAN bottleneck determiner (WBD)

This component is responsible for determining the bottleneck status of each interface that comes from remote campuses AAU WAN links. Which means it determines the real-time bottlenecks status of each WAN link as per the WAN conditions reported by network condition recognizer. Then, it uses predefined rule that are described below for interpreting WAN conditions to bottleneck problem categories. The rules are defined based on the experimental results found from the study regarding the factors and their effect on network bottlenecks categories.

- If link Latency (delay) is high (based on link RTT), then Latency bottleneck

This component uses bottleneck determiner technique (WBD) to accomplish its task and is initialized by link statistics collector technique.

Optimization solution provider (OSP)

The task of this component is selecting and applying the optimization solution technique, as request is received from any of the WAN links. Its operation is based on the real-time bottleneck status of the WAN link that is determined by the pervious components. Like bottleneck determiner, it uses predefined rule but in this case, it is for selecting the optimization solution to be applied on the link, which is appropriate for the link bottleneck problem. This component has an element that can trace the incoming client network application SYN requests from any WAN link, so that optimization technique will be applied to the network application server before it starts sending applications object to the client. Some of the predefined rules are described below:

- If latency bottleneck is true, then increase the TCP window size of network application (to send more data in parallel) as new SYN request comes over the WAN link.
- If latency bottleneck is true then increase application segment size on the server as new SYN request comes over the WAN link.
- If congestion bottleneck is true in any WAN link and if non network traffics are actively using the link, then call method to reschedule these network traffics in an order starting from domain applications, Monitoring traffics, replication traffics and drop application traffics using unknown ports.
- If latency and congestion bottlenecks are true in any WAN link, then apply server based compression techniques on network traffic as new SYN request comes over the WAN link.

This component uses NPM, Packet classifier and optimization selection techniques to accomplish its task and it is initialized by NPM techniques. The next section will explain how these Framework components are integrated to accomplish the tasks.

Event Reporter (ER)

Event reporter provides a report about each problem identified and the action taken to solve it. Moreover, it periodically recommends a lasting solution for the most seen network problems.

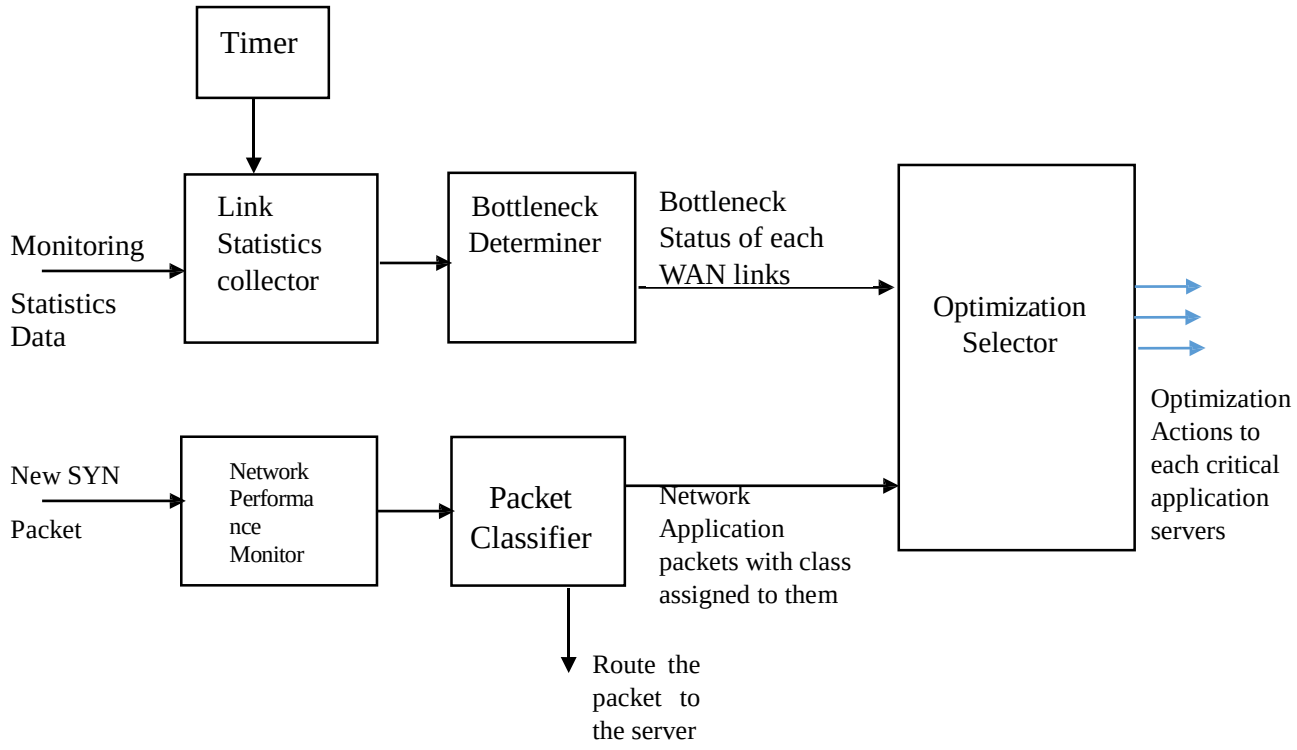

Figure 4.8: The system architecture of the WAN optimization Framework for AAU

4.3.3. Validation of the WAN Optimization Framework

Validation of the developed WAN optimization framework is performed in a real situation at AAU Network scenario. The developed framework was presented to the AAU ICT experts and discussion was done among the experts of AAU ICT office. Before starting the evaluation process of the framework using the questionnaire, the researcher first gave explanations about the developed WAN optimization framework to the domain experts of AAU. This explanation helped domain experts to avoid the variation of awareness among them about the developed WAN optimization framework. During validating the developed framework, the applicability of the developed framework is evaluated by domain experts of the AAU ICT office. Five senior AAU ICT domain experts are actively participated to evaluate the proposed WAN optimization framework. The validation of the framework process is conducted through consecutive discussions with AAU ICT System and Network administrator experts and getting feedback from the experts.

Table 4.7 below presents summary of the feedback obtained from the domain experts of AAU ICT office (evaluators) for the developed WAN optimization framework for the applicability of the framework.

Parameter of Evaluation	Performance Value					
	1	2	3	4	5	NA
1. How do you rate your general understanding of the Developed WAN optimization framework?				3	2	
2. Is the developed WAN optimization framework easy to implement?			1	3	1	
3. Is the developed WAN optimization framework's user interface interactive?				3	2	
4. Are the components in the framework appropriate enough to explain the AAU network situation?				3	2	
5. Is the overall framework design is suitable to further develop an Algorithm or Model?			1	3	1	

Table 4.7. Performance Evaluation of the Framework by Domain Experts

As depicted in table 4.7 above, 60% of the respondents rate the understandability of the developed WAN optimization framework as very good and the remaining 40% of the respondent's rate as excellent. Similarly, the ease of implementation of the WAN optimization framework rated as good by 20% of the respondents whereas the remaining 80% of the respondents rate it as very good. In the case of graphical user interface of the developed WAN optimization framework, 80% of the respondents rate the user interface as very good whereas only 20% of the respondents rate as excellent. Respondents rate the appropriateness of the components of the framework to explain the situation in AAU network is rated as 60% very good and 40% as excellent. Finally, the suitability of the overall framework design for further Model/Algorithm development is 20% rate as good, 60% rate as very good and the remaining 20% of the respondents" rate as excellent. Finally, based on the evaluation results of all the respondents, the average performance of the developed WAN optimization framework is 68%, which is above very good.

The overall performance result from the domain experts shows the developed WAN optimization framework has a promising and encouraging applicability for the enhancement and improvement of the network performance problem in AAU WAN scenario.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATIONS

5.1. CONCLUSION

Due to the centralization of services and applications and decentralization of network users in different university campuses, the WAN has become degraded in performance. Wide Area Network is one of the most important tool for organizations to run their day to day activities. Today, applications and services that are used to be given using local area network (LAN) are also provided using WAN. However, there is more impact on the WAN performance. Besides, most protocols designed for LAN environments do not perform well over the WAN. The factors affecting the WAN are network availability, network latency, congestion and packet loss. This study is aimed to investigate the effects of different WAN factors using performance analysis, with the view to develop a WAN optimization Framework that can improve the performance of the AAU network traffic flows over the WAN link.

To achieve the objective of the study, an experiment has been conducted using cases that are taken from the AAU WAN environment. The study followed the following approach to evaluate the AAU network. Network traffic data is collected using Network Performance Monitor (NPM) tool from the AAU WAN environment, then the collected data is analyzed and evaluated to investigate the network performance using metrics such as network availability, response time, and packet loss.

The purpose of this study is to figure out the network performance challenges and propose a WAN optimization framework solution. From the network perspective, identifying network performance issues is important. Network latency uses response time, packet loss and network availability are considered as performance factors. To understand how it is possible to improve WAN performance, it is important to recognize what is causing the poor performance. This study identifies the performance factors on the AAU WAN environment using a network performance monitor tool. With regards performance problems, this study investigated the main reasons of

WAN bottlenecks such as congestion, response time, and packet loss and network availability. The results of this study indicates the performance of the AAU network shows a high response time, high packet loss rate, and fluctuating network availability. This study developed a WAN optimization framework based on the experimental findings as a potential solution for improving WAN performance. The techniques it provides can improve the response time of network applications and minimize the amount of data travelling through the network. The framework developed for this study was valued and found to be encouraging.

5.2. RECOMMENDATIONS

Based on the finding of this research, the following recommendations are forwarded as a future research direction.

- To evaluate the performance of the AAU WAN, different performance metrics are used such as network availability, response time, packet loss. Further research is recommended to include other performance metrics such as bandwidth utilization, Quality of service, error rate, jitter, retransmission rate, reply rate and others which are not listed here.
- Another recommendation in the future attempts in this area is investigating the evaluation of application performance from the point view of user's experience.
- In this study it is not believed that full investigation is done on the AAU network due to the lack of time and budget. Hence future work is needed to investigate further analysis on the network traffic by using both the real-time traffic data and a log file traffic (historical statistics) data since it is important to give a brief understanding about the performance of the network and generate usage patterns on how the network is utilized.
- This study focuses on evaluating the performance of the AAU wired wide area network and developing a WAN optimization framework based on network traffic analysis in case of AAU University network. Network traffic is a complex; hence, this study covers only network traffic analysis at single Node from the data center located at 6kilo campus. Since the result of this study is promising, further work has to be done to investigate the network performance at other AAU campuses.
- Based on network traffic analysis the WAN optimization framework is proposed which takes in to consideration the geographic locations, the number and type of users and traffic types

passing through the link and different interfaces available, the number of network resources and protocols.

- Based on the proposed WAN optimization framework, this is a roadmap to design a WAN optimization algorithm and model.
- Another possible future work is applying and testing the proposed WAN optimization framework on the campus network since the proposed WAN optimization Framework is a prototype. It has to be tested and validate on the actual WAN environment.
- Further work is required to build a better understanding of network traffic analysis in other Ethiopian Universities network.

REFERENCES

Aberdeen Group. (2007). Optimizing WAN for Application Acceleration. Retrieved from www.aberdeen.com:http://www.riverbed.com/docs/WhitePaper-Riverbed-ApplicationDeliveryOverTheWAN.pdf. Accessed date: January 30, 2014.

Addis Ababa University. (2014). About AAU. Retrieved from AAU official website: http://aau.edu.et/index.php/administration/ict. Accessed date: September 30, 2014.

Addis Ababa University. (2014). About AAU. Retrieved from AAU official website:

http://aau.edu.et/index.php/aboutaau. Accessed date: September 30, 2014

Alisa, Z. (2013). Evaluating the Performance of Wireless Network using OPNET Modeler. Baghdad University Electrical Engineering Department, Iraq.

Arora, R. (2000). Voice over IP: Protocols and Standards. Retrieved from http://www.cis.ohio-state.edu/~jain/cis788-99/voip_protocols/index.html.

Awet, F. (2011). Web Usage: Exploring Navigational Behavior of Users, the Case of the Official website of Addis Ababa University. MSc Thesis, Addis Ababa University, Addis Ababa, Ethiopia.

Bai, G., Oladosu, K. & Williamson, C. (2007). Performance benchmarking of wireless Web servers. Science Direct.

Benson, T., Aditya, A. & Maltz, D. (2010). Network Traffic Characteristics of Data Centers in the Wild. University of Wisconsin–Madison, Microsoft Research–Redmond, Melbourne, Australia.

Bergfeldt, E. (2010). Available-Bandwidth Estimation in Packet-Switched Communication Networks. Master's Thesis, Linköping Studies in Science and Technology, Sweden.

Christner, J., Seils, Z., & Jin, N. (2010). Deploying Cisco Wide Area Application Services. Indianapolis, IN 46240 USA: Cisco Press.

Cisco. (2008). Internetworking with TCP/IP principles, Protocols and Architectures. Fourth Edition, retrieved from: http://docwiki.cisco.com/w/index.php?title=IntroductiontoWAN Technologies&action=edit. Accessed date: October 25, 2013.

Cisco. (2008). Internetworking Technology Handbook. Retrieved from Cisco: http://www.cisco.com/en/US/docs/internetworking/technology/handbook/Intro-to-LAN.html

Cole, R., & Ramaswamy, R. (2000). Wide-Area Data Network Performance Engineering. Boston. London: Artech House.

Comenr, E. & Oy, G. (2002). Internetworking with TCP/IP principles, Protocols and Architectures. Fourth Edition, (pp. 674-719), ISBN: 951-826-435-X.

Englund, H. (2010). Evaluation of traffic generation tools and implementation of test system. Master's Thesis, Umea University, Sweden.

Exinda Networks. (2009). Smarter Approach to WAN Optimization: Unified Performance Management. Retrieved from http://www.exinda.com.

Gezahegn, G. (2007). Performance Evaluation of Wireless Sensor Network Routing Protocols for Critical Condition Monitoring Applications. Master's Thesis, Addis Ababa University, Ethiopia.

Grevers, T, Christner, J. (2008). Application Acceleration and WAN Optimization Fundamentals. Cisco Press.

H3C Technologies. (2014). About Router. Retrieved From: http://www.h3c.com/portal/Products_Solutions/Products/Other_Products/Routers/Quidway_AR4600_Series_Routers/Accessory_Document/200701/194281_57_0.htm. Accessed date: November 3, 2014.

Hailay, W. (2011). Developing Dynamic Bandwidth Allocation Prototype Model for Campus Network Based on Network Traffic Analysis. Master's Thesis, Addis Ababa University, Ethiopia.

Harrison, F. (2010). Response times in Client-Server Systems.

IVANOVICI, M., BEURAN, R. & DAVIES, N. (2005). Assessing Application Performance in Degraded Network Environments: an FPGA-based Approach. Communicating Process Architectures 2005, (pp. 385-395). CERN, Geneva, Switzerland.

Jenson, S. (2009). Consolidated Tactical Network Analysis for Optimizing Bandwidth: Marine Corps Support Wide Area Network (Swan) and TCP Accelerators. Master's thesis, Naval Postgraduate School, MONTEREY, CALIFORNIA.

Juniper Networks, Inc. (2005). Accelerating Application Performance across the WAN. Retrieved from www.juniper.net: www.juniper.net/us/en/products-services/

Kakay, O. (2006). Performance Analysis of LAN, WAN & WLAN in Eritrea. Master's Thesis, University of KwaZulu-Natal (Westville Campus).

Kansanen, M. (2009). Wide Area Network Acceleration in Corporate Networks. Master's Thesis, Lappeenranta University of Technology.

Kokulan, P. (2010). Investigate the performance of a network looking at WAN technologies with differing traffic loads. Master's Thesis, NETWORK AND COMPUTER SYSTEMS SECURITY, University of Greenwich. London.

Lucio, G., Farrera, M., Jammeh, E., Fleury, M. & Reed, M. (no date). OPNET Modeler and Ns-2: Comparing the Accuracy of Network Simulators for Packet-Level Analysis using a Network Test bed. University of Essex, United Kingdom.

Malhotra, R., Gupta, V. & Bansal, R. (2011). Simulation & Performance Analysis of Wired and Wireless Computer Networks. International Journal of Computer Applications (0975-8887), Volume 14– No.7, Adesh Institute of Engineering & Technology, Faridkot.

Manini, D. (2005). Performance Evaluation of Complex Network Applications. Master's Thesis, Universita' Degli Studi Di Torino Dipartimento Di Informatica, Italy.

MEF Technologies. (2010). Understanding Carrier Ethernet Throughput. Retrieved from: www.metroethernetforum.org. Melbourne, Australia.

Moceri, P. (no date). SNMP and Beyond: A Survey of Network Performance Monitoring Tools. Retrieved from: http://www.cse.wustl.edu/~jain/cse567-06/net_traffic_monitors2.htm.

MS-MMSP. (2013). Microsoft Media Server (MMS) Protocol.

Nagaraja, N., Chittal, R. & Kumar, K. (2007). Study of Network Performance Monitoring Tools-SNMP. Journal of Computer Science and Network S 310 Security, R.V.C.E, Bangalore M.Tech, Dept. of C.S.E VOL.7 No.7, (pp. 310-314).

Oo, M. & Othman, M. (2011). The Effect of Packet Losses and Delay on TCP Traffic over Wireless Ad Hoc Networks. Applications, Prof. Xin Wang (Ed.), ISBN: 978-953-307-416-0, InTech. Retrieved from: http://www.intechopen.com/books/mobile-ad-hoc-networks-applications/the-effect-ofpacket-losses-and-delay-on-tcp-traffic-over-wireless-ad-hoc-networks.

Orike, S. & Okwoli, P. (2011). Wide Area Network Implementation Issues in Small and Medium Scale Enterprises. NIGERIA COMPUTER SOCIETY, INTERNATIONAL CONFERENCE, Information Technology for People-Centered Development, Rivers State University of Science and Technology, Nigeria.

Prabhavalkar, N. (2013). The design and evaluation of network services in an active network architectural framework. Master's Thesis, Graduate School-New Brunswick, New Jersey.

Raghavan, K. (2010). BIG-IP WAN Optimization Module Performance: Performance guidelines and testing results for the BIG-IP WAN Optimization module. F5 Networks, Inc. 401 Elliott Avenue West, Seattle, WA 98119 888-882-4447, retrieved from: www.f5.com

Rayanchu, S., Mishra, A., Agrawal, D., Saha, S. & Banerjee, S. (no date). Diagnosing Wireless Packet Losses in 802.11: Separating Collision from Weak Signal. Department of Computer Sciences, University of Wisconsin Madison, USA.

Rehman, S., Turletti, T. & Dabbous, W. (2011). A Roadmap for Benchmarking in Wireless Networks. Inria-00614167, version 1 – 9. Planete Project Team, INRIA Sophia Antipolis, France.

Riabov, V. (2005). Simple Mail Transfer Protocol. Rivier College.

Riverbed. (2013). WAN Optimization vs. More Bandwidth – why increasing bandwidth is not always the answer. Retrieved from http://www.cxo.eu.com/article/.

SAITO, H. (2002). Development and Evaluation of QoS Measurement System, for Internet Applications by Client Observation. Journal of information science and engineering, (pp.891-904).

Sanders, A. (2011). WAN optimization: a business process reengineering and knowledge value added approach.

Sevcik, P. & Wetzel, R. (2008). Improving Effective WAN Throughput for Large Data Flows. Retrieved from: www.silver-peak.com http://www.silver-peak.com/ assets/download/pdf/Net forecastwpEffectiveThroughput.pdf.

SolarWinds. (2014). Orion Network Performance Monitors. Retrieved from: http://www.extralan.co.uk/products/Diagnostictools/Solarwinds/orion.htm.

Tadele, A. (2011). Web Usage Pattern Discovery: The Case of Addis Ababa University Official Web Site. Master's Thesis, Addis Ababa University, Ethiopia.

Tuosto, E. (2003). Non-Functional Aspects of Wide Area Network Programming. Master's Thesis, Universit`a di Pisa Dipartimento di Informatica, Italy.

Wikipedia. (2014). retrieved from: http://en.wikipedia.org/wiki/Ping_%28networking_utility%29

Xu, X. (2012). Evaluation of Wireless Network Performance in a Multi-Nodes Environment. Master's Thesis, Guan Yue Hong.

Yacob, G. (2013). Developing WAN Optimization Model to improve the Performance of Business Critical Applications: The Case of UNECA. Master's thesis, Addis Ababa University.

Yusuff. A (2012). NETWORK MONITORING: Using Nagios as an Example Tool.

Zhang, Y. (2012). Wide Area Network Optimization. IEEE COMMUNICATIONS SURVEYS & TUTORIALS, VOL. 14, NO. 4, FOURTH QUARTER.

APPENDIX

Validation of the Developed WAN optimization framework

This is a validation form for the developed WAN optimization framework and is to be filled by the Domain experts of the AAU ICT staffs in order to evaluate the applicability of the framework

Performance Value	1	2	3	4	5	NA
Description	Poor	Fair	Good	Very Good	Excellent	Not Applicable/ Not Known

Instruction: please assign (X) on the appropriate value for the corresponding parameter of evaluation questions of the developed WAN optimization framework for AAU network environment.

Instruction: please assign (X) on the appropriate value for the corresponding parameter of evaluation questions of the developed WAN optimization framework for AAU network environment.

Parameter of Evaluation	**Performance Value**					
	1	**2**	**3**	**4**	**5**	**NA**
6. How do you rate your general understanding of the Developed WAN optimization framework?						
7. Is the developed WAN optimization framework easy to implement?						
8. Is the developed WAN optimization framework's user interface interactive?						
9. Are the components in the framework appropriate enough to explain the AAU network situation?						
10. Is the overall framework design is suitable to further develop an Algorithm or Model?						