ADDIS ABABA UNIVERSITY

FACULTY OF INFORMATICS

DEPARTMENT OF INFORMATION SCIENCE

DEVELOPING DYNAMIC BANDWIDTH ALLOCATION PROTOTYPE MODEL FOR CAMPUS NETWORK BASED ON NETWORK TRAFFIC ANALYSIS

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

BY

HAILAY WELDEGEBRIEL MARU

January, 2011

ADDIS ABABA, ETHIOPIA

ADDIS ABABA UNIVERSITY

FACULTY OF INFORMATICS

DEPARTMENT OF INFORMATION SCIENCE

DEVELOPING DYNAMIC BANDWIDTH ALLOCATION PROTOTYPE MODEL FOR CAMPUS NETWORK BASED ON NETWORK TRAFFIC ANALYSIS

BY

HAILAY WELDEGEBRIEL MARU

Approved by the Examining Board

_____	_____
Chairman, Examining Committee	Signature
_____	_____
Advisor	Signature
_____	_____
Examiner	Signature

ACKNOWLEDGEMENT

First and foremost, I offer my thanks to almighty GOD, who made this work possible and real. I would like to thank my advisor Workshet Lamenew for his constructive and uninterrupted comments and guidance. I would also like to thank those who influenced my thinking in this field, especially my advisor.

Special thanks go to Ato Teklay G/Michael, the ICT staff member of Mekele University, who assisted me to get the web server log file as well as to investigate the network features of Mekele University using ntop and MRTG.

Also, I offer my respects to the authors of many tools I used in this thesis such as ntop for analyzing the global protocol distribution on real network traffic data, MRTG for network traffic load analysis at different links such as routers, core switches and firewall; Weblog Expert 7.0 for offline web server log file analysis and C# for developing the interface conceptual framework for dynamic bandwidth allocation.

At last but not least, I offer my heartfelt appreciation and indebtedness to my respected family and friends.

Table of Content

Abstract

Enterprise or campus networks usually impose a set of rules for users to access the network in order to protect network resources and enforce institutional policies (for instance, no sharing of music files or no gaming). This leaves network administrators with the daunting task of identifying the application associated with a traffic flow as early as possible and controlling user's traffic when needed. Therefore, accurate and early network traffic analysis is an essential step for administrators to detect intrusion, malicious attacks, or forbidden applications. Hence, bandwidth management based on network traffic flow analysis is a researchable area which gives the owner of the research with lots of opportunity to assess the network behaviors.

In this work protocol distribution and network traffic load analysis are conducted on different links based on real data. Besides, it is tried to assess the potential application of ntop for measuring the global protocol distribution based on the real data. Moreover, the ingoing and outgoing traffic load is analyzed using MRTG which is a versatile tool for graphing network data and it can run on a Web server. Every five minutes, it reads the inbound and outbound octet counter of the gateway router, and then logs the data to generate daily, weekly, monthly and yearly graphs for web pages.

Weblog analysis consists in measuring the usage of relevant traffic activities. Weblog expert tracks web server log file, generating a series of statistics for each host, for operating system, for each browser, for each visitor and soon in the Mekele University (MU) inter campus network as a whole.

Based on the traffic analysis at different links and web server of MU, a dynamic bandwidth allocator algorithm is proposed which consists of modules such as administration tool, which provides a graphical interface for configuring bandwidth allocation based on the different bandwidth demand in the intercampus network; policy agent, which implements the configuration and handles communication with the kernel module; kernel module, which implements the packet classifier, packet selector, bandwidth estimator, unique IP address counter, Host DBA and timer.

Keywords: network traffic, Dynamic Bandwidth Allocation, Bandwidth Management.

List of Figures

List of Tables

List of Acronyms

ARP: Address Resolution Protocol

DHCP: Dynamic Host Configuration Protocol

DNS: Domain Name System

FTP: File Transfer Protocol

HTML: HyperText Markup Language

HTTP: HyperText Transfer Protocol

ICMP: Internet Control Message Protocol

IDS: Intrusion Detection System

IETF: Internet Engineering Task Force

IP: Internet Protocol

MRTG: Multi Router Traffic Grapher

NCSA: Network Computer Security Association

POP3: Post Office Protocol version 3

QoS: Quality of Service

RFC: Request For Comment

SMTP: Simple Mail Transport Protocol

SNMP: Simple Network Management Protocol

SSH: Secure Shell

TCP: Transmission Control Protocol

UDP: User Datagram Protocol

URL: Uniform Resource Locator

Chapter One

Bandwidth Management

1.1 Introduction

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. In recent years, the term bandwidth has been popularly used to refer to the capacity of a digital communications line, typically measured in some number of bits per second. 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. Generally, bandwidth refers to the transmission capacity of a computer channel or communications line or bus, usually stated in bits per second (bps) [1, 3].

Bandwidth in developing countries is so expensive that most universities cannot afford more than 1.544 Mbps – equivalent to the average Western household with ADSL connection [6]. The reasons for this situation include the following, according to [6]:

- In many cases, Internet access to the country is available only via satellite connections, which are much more expensive than cable.
- Where marine fiber cables do exist, they may not carry enough traffic to achieve the economies of scale that make transatlantic bandwidth to Europe, for example, so affordable. In some of the countries that are connected via a marine fiber cable, the telecommunications infrastructure for connecting it to most of the country does not exist.
- The wired telecommunications networks in many developing countries reach only a small part of the population, and many areas (even parts of cities) are not covered at all. The development of wired networks cannot follow the same course as it did in industrialized countries owing to small

populations or low population densities in some areas, poverty, the rise of mobile and satellite communications.

- Some telephone companies that have telephone lines lack the capacity (owing to low demand) to create leased-line connections. Low demand exists mainly because many companies and institutions bypass the national telecommunications grid by using VSAT.
- Leased lines are sometimes analogue instead of digital. On an analogue line, a modem is used for digital transition (such as connection to the Internet), resulting in a maximum speed of 56 Kbps. Digital lines are capable of much higher speeds.
- Bandwidth is also expensive due to the comparative weakness of the currencies of developing countries that have to pay in US dollars or other major currencies for most or all of their upstream international bandwidth.
- While the cost of the telecommunications link between two countries is generally shared, in the case of African countries (and possibly of many other developing countries) the cost of the international link is paid for entirely by the African country. This amounts to reverse subsidization of developed countries.

The modern network becomes complex as there are a number of users and much data to be shared. Hence, the bandwidth needs to be managed appropriately as it becomes the business source of many enterprises.

Bandwidth management is the process of measuring and controlling the communications (traffic, packets) on a network link, to avoid filling the link to capacity or overfilling the link, which would result in network congestion and poor performance [2]. It plays a critical part in traffic management of packet networks. Poor bandwidth management can result in congestion, packet loss, and application performance degradation and thus, affect the overall performance of a network [3].

1.2 African University Scenario: Bandwidth Management

Managing bandwidth improves the performance of an Internet connection by removing unnecessary traffic. Bandwidth is like a pipe. It doesn't matter how big the pipe is, if the traffic in the pipe is not managed it will clog up with unwanted traffic and be hijacked by peer-to-peer traffic, viruses and other malware [5].

"...improving bandwidth management is probably the easiest way for universities to improve the quantity and quality of their bandwidth for educational purposes" [4].

Bandwidth management, as it is described in [5], is essential for any institutional network. Universities understand that if they had a much smaller capacity connection and managed it correctly, the Internet would still be accessible. However, if the connection was increased and the management removed useful access to the Internet would decrease immediately and soon become impossible. The main challenges relating to bandwidth management can be categorized as increasing awareness, improving skills and providing appropriate tools. African Universities are suffering from these challenges of bandwidth management.

In many institutions, bandwidth can be thought of as a public good. By "public goods," economists generally mean a resource that can be consumed by an individual in arbitrarily large amounts, irrespective of the contribution made by that individual to conserving or renewing that resource. Public goods are notorious for being liable to over consumption, and it can be shown that the rational, self-interested individual will almost always choose to over consume – even though this leads to a collective outcome that is bad for everyone. A "public goods problem" is any problem that arises out of this paradoxical tendency. Public goods problems can be managed in a number of ways: for example, by rationing the good, by converting it from a public good into a private good, by coercing appropriate behaviour, by educating consumers, and by fostering community spirit. Those concerned with managing bandwidth need to be informed of this dimension regarding public goods. In particular, they should be

made aware that it only requires a small group of abusers to wreck the availability of 'the good' (or bandwidth) for the group at large. It is almost always the case that a small minority of users account for most of the consumption of an over consumed public good. Thus, 5-10% of users create 50-60% of the problems. Policy aims to manage the behaviour of this minority. If majorities are over consuming bandwidth, then the problem is probably of a different kind: most likely of undersupply (i.e., not enough of the bandwidth is being provided to meet the reasonable needs of the users). Good policy also has an enabling purpose. Policy is not just a set of arbitrary restrictions about how a network may or may not be used. Its central purpose is to govern usage of a resource to provide equitable access to all of its users. By enacting policy, we limit the ability of the minority abusing the network to infringe on the majority who need to use the network.

Bandwidth Optimization for African Universities (BOAU) aims for widespread improvement in how African research and education institutions manage their Internet bandwidth. Because African universities spend up to 100 times more for access compared to their counterparts in industrialized countries, these institutions have far less access to ICTs and online research resources.

For example at Mekele University of Ethiopia, students were not aware of the criteria that constituted acceptable use, because no relevant policy was in place. IT staff could not solve network congestion issues because they were unable to decide which services deserved priority, and which should be cut off altogether.

1.3 Mekele University Scenario: Bandwidth Management

The following discussion is given based on [7]:

Mekele University is one of the universities in Ethiopia with a student population of more than 20,000 in the regular and continuing education programs. At the level of academic programs, it caters both undergraduate and post-graduate programs in various areas of Agriculture, Engineering, Education,

Law, Business, and Health Sciences. The University has a Distance Education Institute (DEI) that runs various programs in areas of business and Economics. The University has three campuses, viz: Endayesus campus (main campus), Adi-Haki campus and newly built Aider campus. To-date, all of the campuses are interconnected in a triangular connection. There is an optical fiber backbone link (about 3km) that connects the main campus and the Adi-Haki campus. There is also a fiber interconnection between Adi-Haki and Aider with about 5 km optical fiber backbone link. Thirteen buildings in the main campus, 4 buildings in the Adi-Haki campus and some buildings in the Aider campus are connected to the Data center using fiber link. The data center is a room that hosts the server machines, CISCO core and catalyst switches and the Main Distribution Frames, MDFs. There are three data centers located in each campus, which is air conditioned and supplied with two 10kVA UPSs and backed by generators in case of power blackout. In each building, there are Intermediate Distribution Frames, IDFs (12" cabinet) where CISCO 2940 switch are installed for the final UTP cabling at each out lets. A good number of LX, SX and 1000-BaseT GBICS are used as the interface between the fiber and the switches, and at the crossover link between the switches in a given IDF. From the IDFs, the trunking of more than 1000 outlets in both campuses follow, which gives connectivity of each node to the data center. Moreover, the switches in each IDF are supplied with UPS to protect the switches and supplied and controlled independently. The data centers at the three campuses houses the server computers which host the network services, manage the network and computer-computer communication within the campus and over the www. Before 2005 G.C the Campus Network was operating with a nominal 256K leased-line connection, via the main Ethiopian Telecommunication Corporation (ETC). During that time as more users connected to the network after the completion of the network, the response of the service for Internet browsing started to slow down. Hence, it needed urgent bandwidth management policy, bandwidth analysis, bandwidth monitoring and bandwidth management in the intercampus network. Currently, the university is requesting for about 20Mbit per second internet line connection.

The generalized available network infrastructure is depicted as follows.

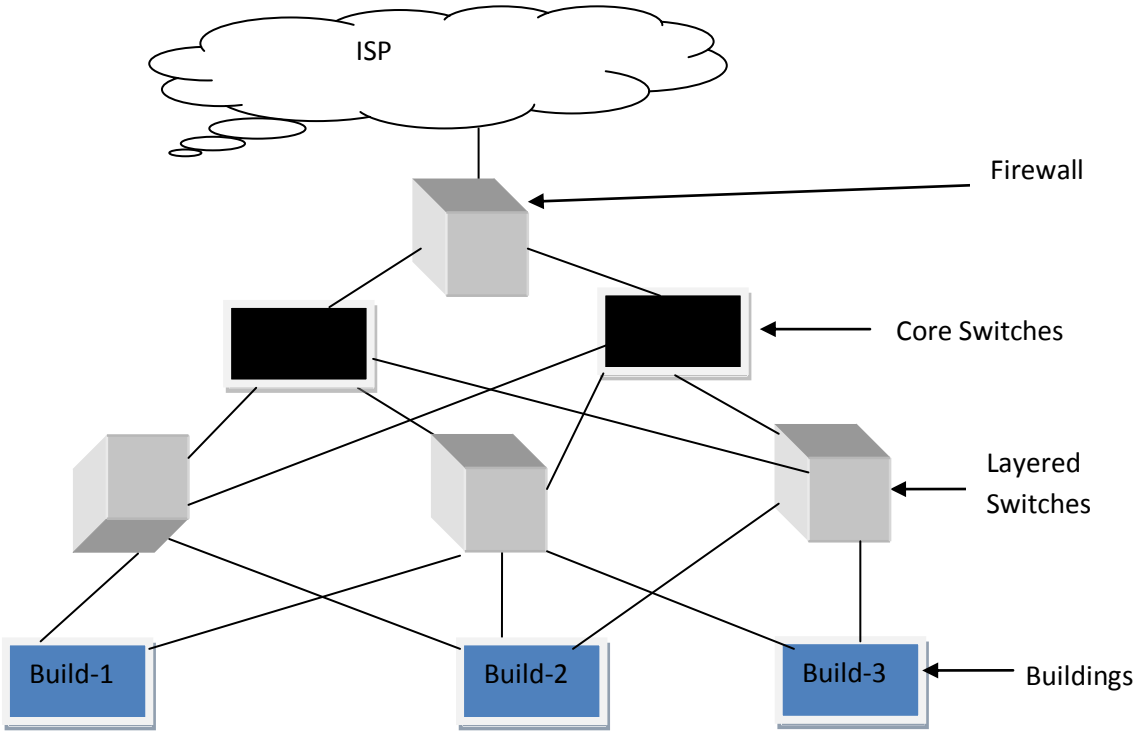


Figure 1.1: Mekele University Network Infrastructure

1.4 Statement of Problem

Web access through a web server at an academic institution can be analyzed using inexpensive software (Weblog expert in this case). Since the web server log keeps a record of every Web site visited, as well as the amount of bandwidth consumed, the top sites by bandwidth can be identified (sites can also be sorted by number of visits, but that is of less interest for bandwidth analysis). From the analysis it is obtained that there are connection attempts by worms to URLs on the Internet that were programmed into the worm's code such as BaiduSpider, Yahoo!Slurp, etc. This shows how the bandwidth usage of an academic institution can be monopolized by malicious worms and unnecessary files. The academic purpose of the network is not fully reflected in the bandwidth usage, and it is clear that this network is out of control.

Other problems particularly associated with university campuses include, according to [5]:

- Students typically have more time, are less supervised, and are under less pressure from work targets than, for example, office workers. Therefore, a university network is one of the most challenging environments to manage.
- People use the Internet in many different ways, some of which are inappropriate or do not make the best use of the available bandwidth. For example, while it may not generally be a problem if a student downloads a music file, plays on-line games; it becomes a problem when the bandwidth consumed by this activity prevents a researcher from downloading or viewing a scientific article.
- Even where high amounts of bandwidth are available, monitoring and optimization are necessary because users (especially students) will always find a way to fill the available amount of bandwidth.
- Hacking: students experiment with their computing knowledge, and connect to exposed computer systems both on the campus and elsewhere in the world.
- Peer-to-peer (P2P) networking (using Kazaa and other programs) is very popular among students eating available bandwidth.
- Universities may need to police the amount of bandwidth they are getting from a shared system because they might be competing for bandwidth with other customers. For example, if a university gets its bandwidth from an ISP, it is likely that the ISP also sells bandwidth to other users, such as local companies. An organization should have a clear understanding of the nature of the shared system – how much minimum bandwidth they are paying for.
- Staff issues. Without proper management, IT staff may themselves become part of the problem.
- Control and monitoring is necessary in order to make informed decisions about how much bandwidth is needed. If usage graphs show that bandwidth is used mainly for recreational activities, or is consumed by virus activity and Windows updates, then control is more urgently required than additional bandwidth.

In view of the above and other associated problems, the university bandwidth should be controlled, managed and optimized so as to make university aware of its utilization capacity; to spread fair internet service without facing aggregated bandwidth starvation in the campus network.

1.5 Objectives

1.5.1 General Objective

The general objective of this work is to develop dynamic bandwidth allocation conceptual framework in intercampus network based on network traffic analysis and propose algorithms for different bandwidth allocator components.

1.5.2 Specific Objectives

In line with achieving the general objective, the study deals with the following specific objectives:

- Analyzing the network traffic generated by network applications
- Identifying top talkers and conversations in the network; determine what applications are using maximum bandwidth.
- Identify bandwidth usage and application usage for each host in university campus daily, weekly, and monthly.
- Developing bandwidth management conceptual framework for intercampus networks.
- Identifying bandwidth management components and developing their algorithms.

1.6 Methodology

In order to accomplish the above objectives, the researcher followed several procedures as described below.

1.6.1 Literature Review

The researcher conducted literature review to assess the major issues and concepts in the field of network traffic analysis, network protocol distribution, Web log analysis and bandwidth management. Various books, journals, articles, and papers from the Internet as well as books were consulted to assess the importance and applications of network traffic analysis, protocol distribution and bandwidth management. Moreover, the researcher reviewed several articles regarding various network traffic analysis tools which are deployed to accomplish the research successfully such as ntop, MRTG, Weblog Expert and other tools to develop interface for bandwidth management in intercampus network.

1.6.2 Data Collection Method

In this paper we analyzed the online and offline network traffic of MU intercampus network so as to develop conceptual framework for bandwidth management in intercampus network. For online traffic analysis the researcher used ntop and MRTG which are used for analyzing protocol distribution and Network traffic analysis at different links respectively. The offline network traffic analysis is based on the server log file which is taken from Mekele University web server. This Server log files are records of web server activity. They provide details about file requests to a web server and the server response to those requests. In the access log, which is the main log file, each line describes the source of a request, the file requested, the date and time of the request, the content type and length of the transferred file, and other data such as errors and the identity of referring pages.

This study takes as its raw data the log files of a single web site from Mekele University web server over a period of 6 days. It reports an attempt to extract as much useful information as possible, in terms of bandwidth usage, browsers usage, website errors, and the like to the site owner and wider information about web surfing behaviour. This includes an analysis of thousands of search engine queries that resulted in visits and the discovery of pages with links to the site.

1.6.3 Network Traffic Analysis Method

The behavior of modern computer networks is fundamentally complex: many users, many uses, numerous protocols, massive operating systems, complex applications, and a wide variety of connected devices [15]. Due to this fact we categorized network traffic analysis in online traffic analysis and offline traffic analysis. The online traffic analysis is implemented using ntop and MRTG to measure the global protocol distribution and to visualize the network traffic usage at different links of Mekele University intercampus network. The offline network traffic analysis implemented using Weblog Expert so as to analyze the website usage and to understand the user behaviors in the intercampus network. The offline network traffic analysis accomplished based on the server log file which is collected from MU web server.

1.6.4 Developing Dynamic Bandwidth Management Conceptual Framework

The Dynamic bandwidth allocation conceptual framework is developed using structural English.

1.7 Scope

Due to shortage of time and finance the study focused on developing dynamic bandwidth allocator conceptual framework based on network traffic analysis in MU intercampus network. The web log file used in this analysis is only of six days as it is only used to assess university bandwidth usage and unnecessary path traversals.

1.8 Significance

Enterprise or campus networks usually impose a set of rules for users to access the network in order to protect network resources and enforce institutional policies (for instance, no sharing of music files or no gaming). This leaves network administrators with the daunting task of identifying the application

associated with a traffic flow as early as possible and controlling user's traffic when needed. Therefore, accurate and early network traffic analysis is an essential step for administrators to detect intrusion, malicious attacks, or forbidden applications. Due to network traffic analysis the dynamic bandwidth allocator helps to allocate the available link bandwidth amongst users or building efficiently. Moreover, It can be used as a springboard for other researchers who intend to continue their studies in the area or other related areas.

1.9 Thesis Outline

The following chapter covers literature survey or theoretical framework; it reviews published and unpublished works, journals and books and to gather any information relevant to the research work. It gives background knowledge about network traffic, network protocol distribution, bandwidth management and web log analysis. Chapter three demonstrates the discussions and results of network traffic analysis using different network traffic analysis tools. Chapter four focuses on developing conceptual framework of the dynamic bandwidth allocator; and, developing algorithms for the different modules of bandwidth allocator. The last chapter is devoted for the final conclusions and recommendations based on the research findings.

Chapter Two

Theoretical Framework

This chapter discusses the need for network traffic analysis, bandwidth management in enterprises and bandwidth management policies. It also gives a broad overview of web log file analysis.

2.1 Network Traffic Analysis

The rapid growth of the Internet in size, complexity and traffic types has made network management a challenging task. The ability of a monitoring system to provide accurate information about the nature and type of the network traffic cannot be over emphasized. Information about who is generating the most traffic, what protocols are in use, where is the traffic originating from or where is the destination of the traffic can be very important to solving congestion problems. Many network administrators spend a lot of time trying to know what is degrading the performance of their network. A typical solution to congestion problem is to upgrade network infrastructure, i.e. replace servers with high end servers and increase the bandwidth. This solution is expensive, short term and does not scale. As soon as the upgrade is done the congestion problem will improve for a while and later gradually deteriorate as the users change their behavior in response to the upgrade. The alternative solution to this problem is to deploy a scalable network traffic monitoring and analysis system, in order to understand the dynamics of the traffic and changes in the internet and overall stability of the network. In addition to knowing the health status of the network, monitoring of network activity also has the benefits of detecting denial of service (DoS) and bandwidth theft attacks. In order to conduct analysis of wide range of network behaviors, it is necessary to collect network traffic on a continuous basis rather than as a onetime event which only captures transient behaviors that provides insight into network problems. Collecting long

term network traffic data will provide valuable information for improving and understanding the actual network dynamics [37].

2.2 TCP/IP Networking Protocols

The TCP/IP suite of protocols is the set of protocols used to communicate across the internet. It is also widely used on many organizational networks due to its flexibility and wide array of functionality provided [16].

These TCP/IP suites of protocols are the sources of different networked applications. In this thesis set of TCP/IP suite of protocols are assessed by showing the percentage each of the network protocols. Moreover, the TCP/IP protocols analysis can be used to clearly show the percentage of traffic consumption which can be used as input for bandwidth allocation in intercampus network. The TCP/IP suites of protocols that is interested in are: FTP, Telnet, SMTP, HTTP, TCP, UDP, ICMP, ARP, DNS, eDonkey, SNMP, DHCP, SSH, Telnet, NetBIOS, and POP3. Strictly speaking, these are not applications themselves, but protocols that are implemented by various applications; however, for this purposes, it is assumed that the protocol is designed for a particular purpose, and that the purpose drives the behaviour more than the implementation does.

The standards for network interactions, or protocols, of these are defined in documents called RFC, published by IETF. The IETF develops and organizes documentation such as protocol standards, best practice documents, and the like for the purpose of improving the Internet [18]; the RFC system is a major vehicle to this work.

FTP, as its name suggests, is a protocol designed to allow the transfer of files between nodes on an internetwork. FTP is defined in RFC959 [17]; this discussion draws primarily on that document. An FTP session consists of two communications channels, the control channel and the data channel. The two channels are used for different purposes and are expected to exhibit entirely different behaviours, but the

ntop considers as general FTP during the Global Protocol Distribution. FTP-control is registered as using port TCP/21 and FTP-data as using port TCP/20 [46].

Telnet is a network protocol used on the Internet or local area networks to provide a bidirectional interactive text-oriented communications facility using a virtual terminal connection. User data is interspersed in-band with Telnet control information in an 8-bit byte oriented data connection over the TCP [32]. Moreover, the Telnet protocol is a protocol designed to provide an interface between terminal devices and processes such as terminals slaved to a mainframe and the mainframe terminal process; on the Internet, it is often used for remote access to a command shell, which is the use with which this work is concerned. It is also used for the control channel of FTP. Telnet is defined in RFC854 [19]; this discussion draws primarily on that document. The purpose of the telnet Protocol is to provide a fairly general, bi-directional, eight-bit byte oriented communications facility. Its primary goal is to allow a standard method of interfacing terminal devices and terminal-oriented processes to each other. It is envisioned that the protocol may also be used for terminal-terminal communication linking and process-process communication distributed computation. A Telnet connection is a TCP connection used to transmit data with interspersed telnet control information. The Telnet Protocol is built upon three main ideas: first, the concept of a Network Virtual Terminal; second, the principle of negotiated options; and third, a symmetric view of terminals and processes. When a Telnet connection is first established, each end is assumed to originate and terminate at a Network Virtual Terminal, or NVT. An NVT is an imaginary device which provides a standard, network-wide, intermediate representation of a canonical terminal. This eliminates the need for server and user hosts to keep information about the characteristics of each other's terminals and terminal handling conventions. All hosts, both user and server, map their local device characteristics and conventions so as to appear to be dealing with an NVT over the network, and each can assume a similar mapping by the other party. The NVT is intended to strike a balance between being overly restricted (not providing hosts a rich enough vocabulary for mapping into their local character sets), and being overly inclusive (penalizing users with modest

terminals). Telnet is registered as using port TCP/23 for communication [46]. In this work the ntop generated the protocol distribution of Telnet from Mekele University network.

SMTP is a protocol used to implement the handling of e-mail on an internetwork; SMTP is defined in RFC821 [23]. This discussion draws primarily on that document. The objective of SMTP is to transfer mail reliably and efficiently. SMTP is independent of the particular transmission subsystem and requires only a reliable ordered data stream channel. While this document specifically discusses transport over TCP, other transports are possible. Appendices to RFC 821 describe some of them. An important feature of SMTP is its capability to transport mail across networks, usually referred to as SMTP mail relaying. A network consists of the mutually-TCP-accessible hosts on the public Internet, the mutually-TCP-accessible hosts on a firewall-isolated TCP/IP Intranet, or hosts in some other LAN or WAN environment utilizing a non-TCP transport-level protocol. Using SMTP, a process can transfer mail to another process on the same network or to some other network via a relay or gateway process accessible to both networks. SMTP is registered as using port TCP/25 [46]. Here, the ntop generated the SMTP coverage of the network protocols. Moreover, The Post Office Protocol, version 3, is a protocol designed to allow e-mail clients, used directly by a human user, to interface with the mail hosts responsible for forwarding mail across an internetwork (via a protocol such as SMTP). POP3 is defined in RFC1939, and registered as using port TCP/25 [46].

The HTTP is a protocol designed to allow the distribution of content via an internetwork. The content distributed via HTTP is not limited, but often consists of documents in HTML and images. The HTTP is an application-level protocol for distributed, collaborative, hypermedia information systems. It is a generic, stateless, protocol which can be used for many tasks beyond its use for hypertext, such as name servers and distributed object management systems, through extension of its request methods, error codes and headers [22]. HTTP version 1.0 is defined in RFC1945 [24], and version 1.1 is defined in RFC2616 [25]; this discussion draws primarily on those document. Both versions are registered as using port TCP/80 [46].

SNMP is an internet-standard protocol for managing devices on IP networks. Devices that typically support SNMP include routers, switches, Servers, workstations, printers, modem tracks, and more [20]. It is used mostly in network management systems to monitor network-attached devices for conditions that warrant administrative attention. SNMP is a component of the Internet Protocol Suite as defined by the IETF. It consists of a set of standards for network management, including an application layer protocol, a database schema, and a set of data objects [21]. SNMP is defined in RFC1157 [26]. This discussion draws primarily on that document. Implicit in the SNMP architectural model is a collection of network management stations and network elements. Network management stations execute management applications which monitor and control network elements. Network elements are devices such as hosts, gateways, terminal servers, and the like, which have management agents responsible for performing the network management functions requested by the network management stations. The SNMP is used to communicate management information between the network management stations and the agents in the network elements. Here, in this thesis, the ntop generated the protocol percentage of the Mekele University network.

TCP is intended for use as a highly reliable host-to-host protocol between hosts in packet-switched computer communication networks, and in interconnected systems of such networks. TCP is defined in RFC793 [27]. This discussion draws primarily on that document. TCP is a connection-oriented, end-to-end reliable protocol designed to fit into a layered hierarchy of protocols which support multi-network applications. The TCP provides for reliable inter-process communication between pairs of processes in host computers attached to distinct but interconnected computer communication networks. Very few assumptions are made as to the reliability of the communication protocols below the TCP layer. TCP assumes it can obtain a simple, potentially unreliable datagram service from the lower level protocols. In principle, the TCP should be able to operate above a wide spectrum of communication systems ranging from hard-wired connections to packet-switched or circuit-switched networks.

As strategic and tactical computer communication networks are developed and deployed in Mekele University campuses, it is essential to provide means of interconnecting them and to provide standard inter-process communication protocols which can support a broad range of applications. In anticipation of the need for such standards, the Mekele University ICT deployed the network and described TCP to be a basis for Mekele University-wide inter-process communication protocol standardization.

UDP is one of the core members of the Internet Protocol Suite, the set of network protocols used for the Internet. With UDP, computer applications can send messages, in this case referred to as datagram, to other hosts on an IP network without requiring prior communications to set up special transmission channels or data paths [28], unlike to the TCP. UDP is defined in RFC768 [29]. This discussion draws primarily on that document. This UDP is defined to make available a datagram mode of packet-switched computer communication in the environment of an interconnected set of computer networks. This protocol assumes that the IP is used as the underlying protocol. This protocol provides a procedure for application programs to send messages to other programs with a minimum of protocol mechanism. The protocol is transaction oriented, and delivery and duplicate protection are not guaranteed. Applications requiring ordered reliable delivery of streams of data should use the TCP. The ntop generated the UDP protocol distribution.

SSH is a network protocol that allows data to be exchanged using a secure channel between two networked devices. The two major versions of the protocol are referred to as SSH1 or SSH-1 and SSH2 or SSH-2. Used primarily on Linux and Unix based systems to access shell accounts, SSH was designed as a replacement for Telnet and other insecure remote shells, which send information, notably passwords, in plaintext, rendering them susceptible to packet analysis. The encryption used by SSH is intended to provide confidentiality and integrity of data over an unsecured network, such as the Internet [30]. SSH is defined in RFC4253 [31]. This discussion draws primarily on that document. The SSH transport layer is a secure, low level transport protocol. It provides strong encryption, cryptographic host authentication, and integrity protection. Authentication in this protocol level is host-based; this protocol

does not perform user authentication. A higher level protocol for user authentication can be designed on top of this protocol. The protocol has been designed to be simple and flexible to allow parameter negotiation, and to minimize the number of round-trips. The key exchange method, public key algorithm, symmetric encryption algorithm, message authentication algorithm, and hash algorithm are all negotiated. It is expected that in most environments, only 2 round-trips will be needed for full key exchange, server authentication, service request, and acceptance notification of service request. The worst case is 3 round-trips.

NetBIOS is an acronym for Network Basic Input/Output System. It provides services related to the session layer of the OSI model allowing applications on separate computers to communicate over a local area network. As strictly an API, NetBIOS is not a networking protocol. Older operating systems ran NetBIOS over IEEE 802.2 and IPX/SPX using the NetBIOS Frames (NBF) and NetBIOS over IPX/SPX (NBX) protocols, respectively. In modern networks, NetBIOS normally runs over TCP/IP via the NetBIOS over TCP/IP (NBT) protocol. This results in each computer in the network having both an IP address and a NetBIOS name corresponding to a host name [33]. Even though this TCP/IP protocol runs over older operating systems, in our case we assessed its existence in the Mekele University intercampus network and the ntop generated the protocol distribution of NetBIOS over TCP/IP (NBT) protocol. The ntop is installed in Mekele University server and collects the necessary network usage statistics. Installation of ntop on Debian is given in Appendix I.

2.3 Network Traffic Analysis Tools

2.3.1 Ntop

Ntop [40] provides a display similar to the UNIX top command, but for network traffic, it can be used for traffic measurement and monitoring. Features such as the embedded HTTP server, support for various network media types, light CPU utilization, portability across various platforms, and storage of

traffic information into an SQL database makes ntop versatile. However, ntop is limited by its high memory requirements when operating in a continuous monitoring environment. The extensive cache usage has the drawback that memory usage is increased [41]. This makes ntop a memory and computational intensive application. Like several of the other tools, ntop uses the same packet capture library to obtain the network data. Since ntop operates in continuous mode, it is designed to operate on networks with speeds of less than 100 Mbps [42]. A continuous network tracing infrastructure was described in [43], it is multi-user based and capable of collecting archives and analyzing network data captured. The system is limited by not providing a universal web interface to display the results of its monitoring. The system requires huge storage, 11TB of shared disk space for data repository since it is not web based and also there is high overhead since the storage system is accessed through an NFS server over an Ethernet link.

2.3.2 Multi Router Traffic Grapher

MRTG is a versatile tool for graphing network data [38], this tool can run on a Web server. Every five minutes, it reads the inbound and outbound octet counter of the gateway router, and then logs the data to generate graphs for web pages. These graphs can be viewed using a web browser. Although MRTG gives a graphical overview, it however does not give details about the host and protocol responsible for the traffic monitored. On the other hand, Windmill [39] is a modular system for monitoring network protocol events; it is useful for acquiring the data from the network, however it is limited in its capability by not providing any facility to aid in the analysis of those events or non protocol events acquired.

Many traffic measurements meter throughput on time scales in the order of 5 minutes, for instance, using the MRTG tool. The time scale in which users and machines perceive QoS is, obviously, orders of magnitudes smaller. This is one of many possible reasons for degradation of the perceived quality, congestion on links along the path network packets traverse. In order to prevent quality degradation due

to congestion, network links have to be dimensioned in such a way that they appropriately cater for traffic bursts on time scales similarly small to the time scale that determines perceived QoS. It is well-known that variability of link load on small time scales (e.g., 10 milliseconds) is larger than on large time scales (e.g., 5 minutes) [36] .

MRTG is free software licensed under the terms of the GNU General Public License for monitoring and measuring the traffic load on network links. It allows the administrators to see traffic load on a network over time in graphical form. It was originally developed by Tobias Oetiker and Dave Rand to monitor router traffic, but has developed into a tool that can create graphs and statistics for almost anything. MRTG is written in Perl and can run on Windows, Linux, UNIX, Mac OS and NetWare [34]. The MRTG can be easily installed on Debian using methods discussed in Appendix II.

2.3.2.1 Features of MRTG

The basic features of multi router traffic Grapher are described in [34] as follows.

- Measures two values (Ingoing traffic and Outgoing traffic) per target.
- Gets its data via an SNMP agent, or through the output of a command line.
- Typically collects data every five minutes (it can be configured to collect data less frequently).
- Creates an HTML page per target that features four graphs (GIF or PNG images).
- Results are plotted vs. time into day, week, month and year graphs, with the ingoing traffic plotted as a full green area, and the outgoing as a blue line.
- Automatically scales the Y axis of the graphs to show the most detail.
- Add calculated Max, Average and Current values for both ingoing traffic and outgoing traffic to the target's HTML page.
- Can also send warning emails if targets have values above a certain threshold.

2.3.2.2 MRTG Measurement Setup

In this study we have concentrated on network technologies that are common in both production and academic environments. We assume such networks to consist of (normal, Fast and Gigabit) Ethernet links, which are connected by hubs, switches and routers. In general it is important that normal network operation should not be disturbed by the measurements. For example, it may not be acceptable to interrupt network operation to install optical splitters that copy all network traffic to a measuring device. It will often be better to rely on the mirror facilities provided by current middle- to high-end switches, and copy all traffic that should be monitored to one of the free interfaces of that switch. The measurement device can be connected to that interface, and all traffic can be analyzed without disturbing the network. See Figure 2.1 for a schematic representation of the measurement setup.

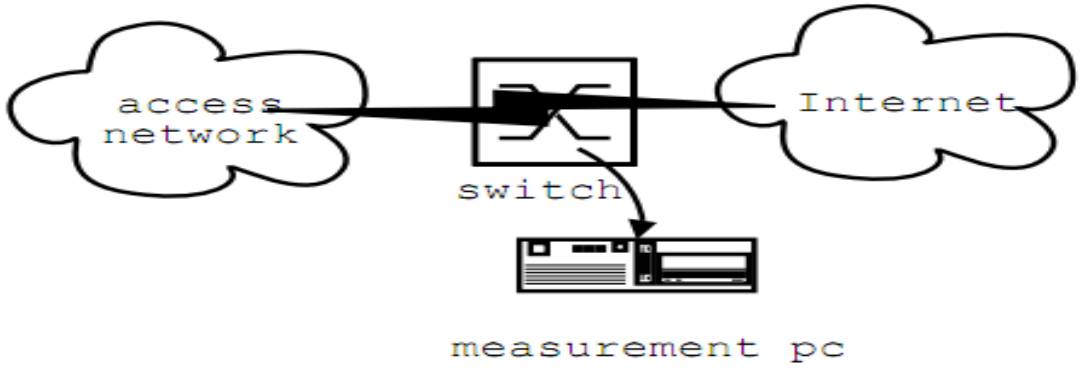


Figure 2.1: MRTG Measurement Setup

MRTG is a free performance management application for Unix/Windows; it monitors SNMP statistics from any SNMP-capable device on any network and:

- Captures, stores, and graphically presents SNMP data. By default, a web page with four graphs per network is created by MRTG. The graphs show the variation of network data over time.

- Runs from the crontab. Every five minutes, a cron job runs MRTG to query a user-configured list of OIDs and network devices. After each data collection cycle, the MRTG perl script posts updated graphs to a web page.
- Efficiently compresses and archives data samples to create graphs.
- Enables you to determine if trending data is useful for monitoring your environment before you invest in costly network performance software. If trending data is critical to manage your network, it may be necessary to purchase a commercial network performance package, such as Concord Network Health.

For each objects referenced in the configuration file, MRTG creates the following graphs:

- Daily graph—5 minute average data points with approximately 33 hours of data presented.
- Weekly graph—30 minute average data points with approximately 8 days of data presented.
- Monthly graph—2 hour average data points with approximately 5 weeks of data presented.
- Yearly graph—1 day average data points with approximately 1 year of data presented.

2.4 Weblog Analysis

The WWW continues to grow at an astounding rate in both the sheer volume of traffic and the size and complexity of Web sites. The complexities of tasks such as Web site design, Web server design, and of simply navigating through a Web site have increased along with this growth. An important input to these design tasks is analysis of how a Web site is being used. Usage analysis includes straightforward statistics, such as page access frequency, as well as more sophisticated forms of analysis, such as finding the common traversal paths through a Web site. Usage information can be used to restructure a Web site in order to better serve the needs of users of a site. Long convoluted traversal paths or low usage of a page with important site information could suggest that the site links and information are not laid out in an intuitive manner. The design of a physical data layout or caching scheme for a distributed or parallel

Web server can be enhanced by knowledge of how users typically navigate through the site. Usage information can also be used to directly aid site navigation by providing a list of popular destinations from a particular Web page [52, 54].

Seeking to improve internet service, many companies particularly higher institutions, use statistics to discover the types of people who tend to visit their sites, and the pages that they tend to visit within the site. Some companies do their own statistical analysis, while many more either buy software or hire another company to perform the analysis for them. Less than a decade ago, the only products performing web analysis were shareware and freeware. Commercial web analysis programs arrived on the market around 1994, thereby reducing the number of programmers willing to continue to produce the software for free [61]. Nevertheless, there are still some free log analysis tools, which are downloadable from the Internet such as Web Log Expert which is used for offline analysis of the web server log file. The commercial tools on the market are generally of better quality and range from $500 to $1,000 [60]. For an even steeper price, service companies will perform the analysis for other online companies using their own software and upload the statistics to the web. These companies typically charge a few hundred dollars per month, yet they take much of the work out of performing the analysis oneself. Moreover, whenever the service company upgrades their software, the customer benefits from the upgrade [61].

Web analysis software and companies obtain their statistical information from Server Access logs. These logs are available on all servers, and consist of transfer logs, error logs, referer logs, and agent logs. The transfer log provides a list of all hits on the server and the times at which they occurred. The error log, as the name implies, lists all errors that were made to the server. The referer log provides a list of the locations from which users came before entering the site, and the pages that the users hit first within the site. The agent log, finally, lists the web browsers or search engines used by the users in the site [61].

According [55], Web sites are generating a big amount of Web logs data that contain useful information about the user behavior.

Web log file analysis began as a way for IT administrators to ensure adequate bandwidth and server capacity on their organizations' web sites. Log file analysis has advanced considerably in the past years, with companies now mining log files for finer-grained detail about visitor profiles and buying activity. Organizations are now seeking ways to use log files to learn about the usability of their web sites—that is, how successfully visitors meet their specific information or transaction goals there. Log file data can offer valuable insight into web site usage. It reflects actual usage in natural working conditions, compared to the artificial setting of a usability lab. It represents the activity of many users, over a potentially long period of time, compared to a limited number of users for an hour or two each [51, 52].

2.4.1 Data Sources for Web Log Analysis

According [59], the data sources used in Web Usage Analysis may include web data repositories like web server logs, proxy logs which are explained as follows.

Web Server Logs – These are logs which maintain a history of page requests. The W3C maintains a standard format for web server log files, but other proprietary formats exist. More recent entries are typically appended to the end of the file. Information about the request, including client IP address, request date/time, page requested, HTTP code, bytes served, user agent, and referrer are typically added. These data can be combined into a single file, or separated into distinct logs, such as an access log, error log, or referrer log. However, server logs typically do not collect user-specific information. These files are usually not accessible to general Internet users, only to the webmaster or other administrative person. A statistical analysis of the server log may be used to examine traffic patterns by time of day, day of week, referrer, or user agent. Efficient web site administration, adequate hosting resources and the fine tuning of sales efforts can be aided by analysis of the web server logs. Marketing departments of any

organization that owns a website should be trained to understand these powerful tools. Detailed Explanation about the information obtained from the web server log file will be given below.

Proxy Server Logs- a Web proxy is a caching mechanism which lies between client browsers and Web servers. It helps to reduce the load time of Web pages as well as the network traffic load at the server and client side. Proxy server logs contain the HTTP requests from multiple clients to multiple Web servers. This may serve as a data source to discover the usage pattern of a group of anonymous users, sharing a common proxy server.

In this study we used the web server log which is discussed below in detail. A server log is a log file (or several files) automatically created and maintained by a server of activity performed by it. A typical example is a web server log which maintains a history of page requests. The W3C maintains a standard format for web server log files, but other proprietary formats exist. More recent entries are typically appended to the end of the file. Information about the request, including client IP address, request date/time, page requested, HTTP code, bytes served, user agent, and referrer are typically added. These data can be combined into a single file, or separated into distinct logs, such as an access log, error log, or referrer log. However, server logs typically do not collect user-specific information. These files are usually not accessible to general Internet users, only to the webmaster or other administrative person. A statistical analysis of the server log may be used to examine traffic patterns by time of day, day of week, referrer, or user agent. Efficient web site administration, adequate hosting resources and the fine tuning of sales efforts can be aided by analysis of the web server logs [51, 53]. Moreover, we analyzed the log file obtained from firewall of Mekele University so as to analyze the security level of the campus.

2.4.2 Web Log Analysis Applications

Letizia: is an application that assists a user browsing the Internet. As the user operates a conventional Web browser such as Mozilla, the application tracks usage patterns and attempts to predict items of interest by performing concurrent and autonomous exploration of links from the user's current position.

The application uses a best-first search augmented by heuristics inferring user interest from browsing behavior [57].

WebSIFT: the WebSIFT (Web Site Information Filter) system is another application which performs Web Usage Mining from server logs recorded in the extended NSCA format (includes referrer and agent fields). The preprocessing algorithms include identifying users, server sessions, and identifying cached page references through the use of the referrer field. It identifies interesting information and frequent item sets from mining usage data [56].

Web Log Expert: Weblog Expert is a powerful Apache log file analyzer and IIS log analyzer. It can help the network administrators reveal important statistics about website usage: activity of visitors, access statistics, and paths through the site, visitors' browsers, and much more. The program can read log files of the most popular web servers: Apache (Combined and Common log formats are supported) and IIS 4/5/6/7. Weblog Expert can also read ZIP, GZ, Text and BZ2 compressed logs so the network administrators do not need to unpack the logs manually before analyzing. Before network administrators start analyzing they need to create a profile that contains the path to your log files, information about web site domain, filters, tracked files, etc. Network administrators can create as many profiles as they wish regarding their organization's webs site [52].

In this study Web Log Expert is deployed to analyze Mekele University web server log file so as to assess web statistics in the campus network. These network traffic analyses show there are unwise bandwidth utilization in the campus network. Hence, it is a necessary condition to implement appropriate bandwidth management so as to reduce the bandwidth starvation in the campus network.

2.5 Bandwidth Management

Bandwidth is the maximum amount of data that can travel a communications path in a given time. Bandwidth is typically measured in bits per second. If you think of the communications path as a pipe, then bandwidth represents the width of the pipe that determines how much data can flow through it all at

once [13]. Inbound traffic is data that is received by your computer from another computer. Outbound traffic is data that is sent from your computer to another computer.

With the ever-increasing number of users that use high bandwidth for broadband multimedia traffic over the Internet, such as interactive games, videoconference, high-definition television (HDTV) and other high-speed services, high growth of access network is the most glaring issue in the communication industry [12]. Bandwidth management is the process of measuring and controlling the communications (traffic, packets) on a network link, to avoid filling the link to capacity or overfilling the link, which would result in network congestion and poor performance [2].

Bandwidth management plays a critical part in traffic management of packet networks. Poor bandwidth management can result in congestion, packet loss, and application performance degradation and thus, affect the overall performance of a network [3]. Bandwidth management is a general term given to a collection of tools and techniques that an institution can use to reduce demand on critical segments of networks. Often bandwidth management may be applied on the wide area network segment that connects the institution to the greater internet. It may also be applied on critical internal segments, such as segments connecting campus residence halls to the rest of the network [1].

In order to effectively manage a network connection of any size, you will need to take a multifaceted approach that includes effective network monitoring, a sensible policy that defines acceptable behavior, and a solid implementation that enforces these rules. Each component is important for effective bandwidth management in any network that consists of more than a few users [1]. Accordingly, it is a necessary condition for any enterprise to define bandwidth management policy, to define bandwidth analysis and monitoring approach and its implementation based on the current status of network traffic analysis.

Bandwidth management requires three activities: Policy, Monitoring and Implementation. If any one of these activities is missing then the management of bandwidth is significantly compromised. The

activities inform and reinforce each other [5]. The figure below shows the triangular components of bandwidth management.

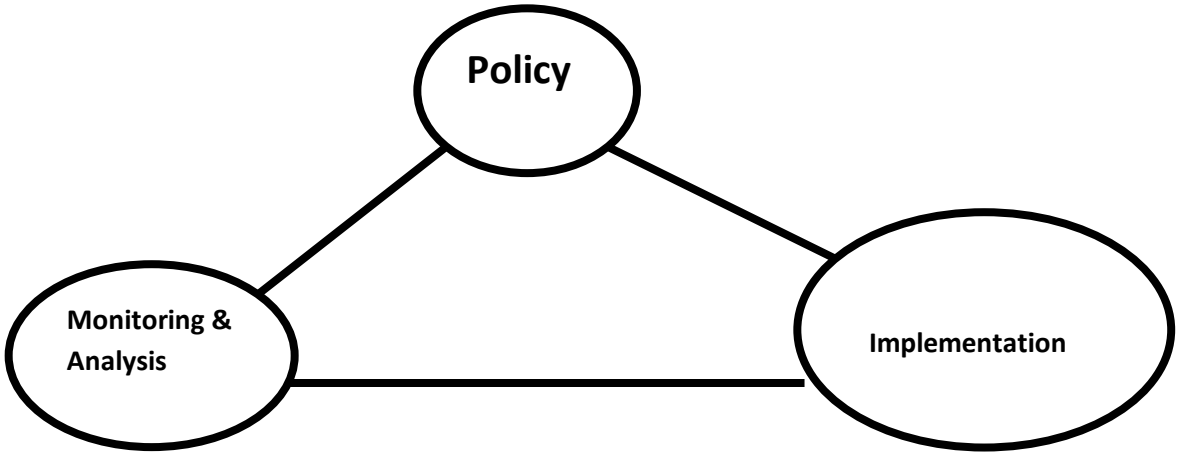


Figure 2.2: Policy, Monitoring & Analysis, and Implementation of bandwidth management

The above figure shows that bandwidth management consists of defining policy of bandwidth utilization management, monitoring and analyzing the allocated bandwidth and implementation of bandwidth allocation in campus network. Each of the components is explained as follows, according to [5].

A policy is a statement of opinions, intentions, actions and procedures that guide the overall use of the network. An acceptable use policy is a subset of this, setting out in technical detail what uses of the network are believed by the network operators to be acceptable, and what they intend to do to anyone who uses it in a manner that they consider unacceptable. It should be a written document that defines acceptable forms of network access, as well as guidelines for how network problems are dealt with, definitions of abuse, and other operational details. The policy also typically includes definitions of legal constraints for network users (such as the exchange of copyrighted material, requesting inappropriate materials, etc.). Having a policy makes it much easier to enforce certain types of network behavior, as you will be able to hold people to a set of agreed rules [1].

User education is obviously critical to every stage of implementing a plan to manage your bandwidth. While users can be forced to adhere to certain behavior patterns, it is always far easier to implement a plan with their voluntary compliance. But how does such a plan come into being? If you simply order people to change their behavior, little is likely to change. If you install technical hurdles to try to force them to change, they will simply find a way around the obstacles. Network monitoring is the ongoing process of collecting information about various aspects of network operations. By carefully analyzing this data, you can identify faults; find cases of waste and unauthorized access, and spot trends that may indicate future problems [1].

Implementation is the step of implementing traffic shaping, filtering, caching, and other technologies within your network to help bring actual usage in line with policy. The actions you need to take are indicated by the data collected through monitoring and analysis, and are constrained by the network policy. Many people expect to begin the task of bandwidth management by starting with this step. But without good monitoring techniques, network administrators are effectively blind to the problem [1].

An abundance of bandwidth enables electronic collaboration, access to informational resources, rapid and effective communication, and grants membership to a global community. An absence of bandwidth prevents access to the aforementioned global community, restricts communications, and slows the speed at which information travels across the network. Therefore, bandwidth is probably the single most critical resource at the disposal of a modern organization. Because bandwidth is a valuable and costly resource, demand usually exceeds supply. In many environments, unrestrained access and usage of bandwidth results in degraded service for all users. This is partly a supply problem (not enough bandwidth is available to meet demand), partly a demand problem (too many demands are being made on the limited resource), and partly a technical problem (little or no technical management and optimization of the resource is happening). The end result is a poor user experience when trying to use resources and tools that rely on bandwidth (e.g., browsing the web, sending emails, using network applications, etc.). Bandwidth management and optimization are often seen as technical issues.

However, policy is an essential component of any bandwidth management strategy. Without it, technical solutions will be difficult to implement and much less effective. Policies are essential, in that they provide the framework for defining how a network is to be used and detail how technical solutions should be implemented. Policy should be thought of as guidelines concerning network usage for both the users and those responsible for maintaining the network itself. Without a plan, unrestricted access to the campus network would push its management into total chaos [1].

Bandwidth allocation has been a classical problem since long time in communication networks. Dynamic allocation schemes effectively utilize the bandwidth as compared to static bandwidth allocation schemes. Allocation of fixed amount of network bandwidth to different nodes of a network based on the various changing parameters and different constraints instantly with time is dynamic allocation. The constraints could be real-time application requests, error rate, pricing, priority, node's allocation state information, congestion status, etc. Thus, system software must be developed that will be placed in the router to allocate bandwidth dynamically (to nodes/applications) within an organization based on various constraints/parameters. Hence, every enterprise is nowadays in a need of guaranteed bandwidth to run their day to day activities either in business, in education or in manufacturing.

2.5.1 The Need for Guaranteed Bandwidth

Corporate networks are rapidly evolving from a classic client/server paradigm towards an intranet-based model, founded on information-sharing and Web navigation. The resulting increased use of the network is causing increased use of bandwidth. Simply shackling on new connections and fatter pipes is not cost effective, and won't guarantee availability where it's most needed. The emerging intranet-based computing model implies many new factors. And MIS managers and service providers will have to take these into account when planning their network, and in offering service to their customers:

- Very different types of traffic co-exist, with different requirements in terms of quality of service. For example, interactive telnet traffic, real-time audio and video traffic, and bulk FTP traffic.
- It is becoming increasingly common for mission-critical applications to be deployed on the intranet. For example, enterprise resource planning (ERP) applications such as SAP and Oracle. Also, use of e-commerce and business-to-business transaction packages on extranets is increasing.
- On servers, key resources such as CPU and memory are managed by system management tools. Network bandwidth, which is an even more essential resource, also requires management.
- There is demand for increased amounts of bandwidth: People may stay on the link for extended periods of time and download large amounts of data.
- There is demand for guaranteed quality of service in terms of bandwidth and minimum delay. Emerging Internet applications are both bandwidth-intensive and time-sensitive, often requiring support for voice, video, and multimedia applications across the network infrastructure.
- User attitudes are changing: Users expect instant access to information without delays or restrictions, especially if that information is critical to their work.

2.5.2 Importance of Bandwidth Management for Higher Education

A robust campus network with good connectivity to the internet is no longer a luxury to higher education institutions; it is a necessity. Students consider good access to networked resources as a factor in their choices of institutions. It impacts both recruitment and retention. Students increasingly come to college expecting that the network will meet their educational needs, but they also expect it to provide their entertainment. Institutions wishing to factor this in to their decision making should favor prioritization schemes rather than blocking schemes. It is better from a management and public relations perspective to throttle recreational traffic than to block it completely. Some bandwidth management products allow for recreational traffic to increase when business traffic is low. As video and voice

services also move to the network, managing bandwidth to provide quality of service for time-sensitive applications will become even important. Research demands on bandwidth are also growing. Today we can buy more bandwidth for less money than in the past and this trend will probably continue. Nevertheless, it is often not practical to meet the increased demand for bandwidth by simply buying more. Each campus must decide when the cost of investing in a bandwidth management strategy will cost less than buying more bandwidth. Investing in neither a bandwidth management strategy nor more bandwidth will leave campus network at risk of being hopelessly bogged down, to the point where no user is well served [4].

According to [7], bandwidth management becomes essential when the cost of adding enough bandwidth to meet demand exceeds the cost of investing in bandwidth management technology. Some campuses usually smaller ones can appeal to their students, faculty, and staff to be good citizens. This approach may work for a while, depending upon the values of the campus community. Other campuses with larger staffs, can sometimes afford the staff time to track down users of excessive bandwidth and encourage or discipline them directly to control the problem. Automated bandwidth management tools work well, and are worth the expense, when these other techniques are no longer sufficient to solve the problem.

2.5.3 Issues to be addressed with Bandwidth Management

Campus network units getting ready to employ bandwidth management should develop a communication plan for the campus. Explain to the campus what is done regarding bandwidth management, emphasize the importance of giving all users fair access to resources as well as ensuring better access to resources that are more central to the mission of the campus. In a campus network if more money is invested on bandwidth to support unlimited recreational applications, it will drive up the cost of student's education. If unlimited bandwidth access is allowed to recreational applications on limited bandwidth, students will not be able to reach the legitimate educational resource materials. From

the analysis it is found that there are a number of applications which consumes the campus bandwidth inappropriately. Hence, such applications should be identified based on the time intervals and make dynamic bandwidth allocation to have fair bandwidth amongst the users in a specific geography.

2.5.4 Bandwidth Challenges

The main challenges relating to bandwidth management can be categorized as increasing awareness, improving skills and providing appropriate tools. These challenges will be described as follows according to [5].

Increasing Awareness: Although there are technical issues relating to bandwidth management, the biggest challenge is to raise awareness of the importance of managing bandwidth. Bandwidth is a limited resource that needs to be shared. Bandwidth has a cost and policy should govern its use. Just as with a phone line, use of bandwidth should be monitored and managed.

Improving Skills: Capacity building and skills development are fundamental to improving bandwidth management practice within developing world institutions. Institutions have put forward a strong demand for bandwidth management training, even amongst those institutions that have organized some form of training in the past. The challenge is to provide comprehensive training on policy and the purpose of bandwidth management for managers, integrated with hands-on technical training for network administrators.

Providing Appropriate Tools:- In many institutions the necessary tools are not yet present. And if they are present, they may not be used to their full extent. Any institution is expected to make investigation on bandwidth management tools and would answer the remaining questions:

- Why are tools not being used at many institutions?
- Are the tools appropriate for this audience?
- Are there significant gaps in the functionality of existing tools?

- How can the most appropriate tools be integrated into a wider programme of bandwidth management and training?
- Can any existing tools be leveraged for use by small or overstretched IT teams?

Hence, in this work it is developed bandwidth management conceptual framework by defining some network bandwidth policy in case of Mekele University. These policies are developed based on online network traffic analyzed using ntop and MRTG in Mekele university intercampus network and offline web server log file analysis using web log expert. For further details how the network traffic is analyzed next chapter of this thesis.

Chapter Three

Discussions and Results

3.1 Introduction

In campus network it is noticed a significant increase in the amount of streaming media to and from the Internet. While we realize the increase in popularity of streaming media, YouTube, etc., the campus network should not support the entertainment needs of the population. Limits have been placed on the amount of bandwidth that can be used for Peer-to-Peer software, games, VoIP, and streaming media, as these applications have the ability to overwhelm the intercampus Internet connections quickly. These limits are in place in all areas of the university. While it is realized this may cause inconvenience to some, it is necessary to ensure the critical applications of the university can continue to function, and that Internet remains usable for all. Traffic flows are being constantly analyzed and limits are subject to change. The most prevalent reason for the change is to protect the availability of Internet resources to the University community. The ability of users outside the University community to access information and resources within the University is also crucial. Also, the inability to access the University's main web server can be detrimental to the public perception of the University, and hinder the ability to deliver information to the public at large.

3.2 Bandwidth Management Based on Network Traffic Analysis

Network traffic analysis-based Bandwidth Management helps enterprises to establish priorities based on users, web category, groups and applications with precise bandwidth allocation based on usage and time of the day. Network traffic analysis based bandwidth management complements bandwidth management by blocking access to high bandwidth-consuming audio and video downloads gaming, tickers, ads and more. This ensures that business and bandwidth critical applications like CRM, VoIP and more gain

guaranteed bandwidth. Enterprises can fine tune their bandwidth policies based on changing user requirements as well as their usage for continually improved network performance.

Bandwidth management is the process of measuring and controlling the communications (traffic, packets) on a network link, to avoid filling the link to capacity or overfilling the link, which would result in network congestion and poor performance [2]. Network monitoring is the ongoing process of collecting information about various aspects of network operations. By carefully analyzing this data, you can identify faults; find cases of waste and unauthorized access, and spot trends that may indicate future problems [1].

3.3 Managing Network Capacity

Almost all network links are used by more than one user or application. This means that the available bandwidth must be shared between them. Bandwidth management tools provide management and control for how this is done. If a network link is continuously congested, the link needs to be upgraded to provide greater capacity. In many cases, however, the typical load on a link is within the link capacity, and the link is congested only temporarily. Temporary congestion is sometimes predictable. For example, typically there are peaks in network use at particular times of the day or following a particular event. Other causes of temporary congestion, such as the transfer of a large file, are not possible to predict. If the average use of a link is within the link capacity, considerable improvements can be made in the performance of the network link by managing how the available bandwidth capacity is used. Allocating bandwidth to a particular type of traffic helps to optimize the usage of the available bandwidth.

Hence, in this work it is developed bandwidth management conceptual framework by defining some network bandwidth policy in case of Mekele University. These policies are developed based on network traffic analyzed using ntop and MRTG in Mekele university intercampus network.

In this work the network traffic is analyzed at different links so as to observe or understand the network traffic utilization status in the intercampus network using different tools such as MRTG to measure the traffic load at different links and ntop to measure the global protocol distribution i.e. the protocol distribution in the intercampus network is not host specific. This protocol distribution shows the percentage of protocols. The following discussions show how the network traffic protocol distribution is analyzed in Mekele University intercampus network using ntop which is used as input for developing conceptual framework of bandwidth management.

3.4 Network Traffic Analysis –Case Mekele University

3.4.1 Network Traffic Analysis using ntop

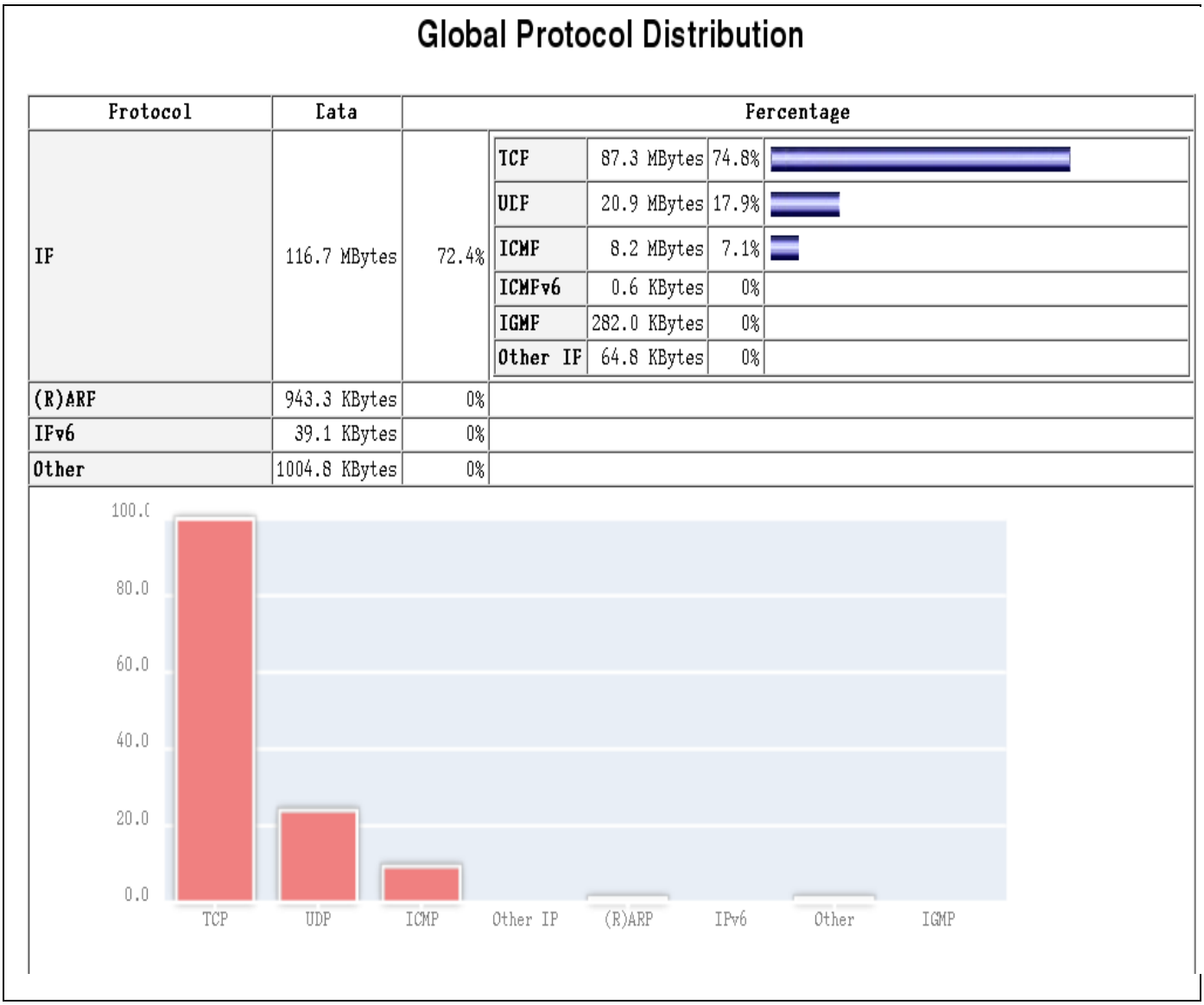
Network management is becoming an increasingly complex task due to the variety of network types and the integration of different network media. As networks become larger, more complex, and more heterogeneous, the costs of network management rise. In this scenario, automated tools are needed to support human effort, gathering information about the status and behavior of networked elements. According to [9], network monitoring is the most fundamental aspect of automated network management.

In this thesis we introduced ntop to measure the global protocol distributions which are important for network planning and optimization and to server as input for Dynamic bandwidth management, MRTG to measure the bandwidth usage daily, weekly, monthly and yearly, so as to predict the bandwidth demand for bandwidth prediction in the intercampus networks.

Table 3.1: Platforms, Media and Protocols Supported by ntop

PLATFORMS	UNIX, Win32
MEDIA	Ethernet, Token Ring, PPP,FDDI, Raw IP, Loopback
PROTOCOLS	IP, IPX, NetBIOS, OSI, AppleTalk, DecNet, DLC
IP PROTOCOLS	Fully user configurable (NFS, HTTP, X11, DNS, FTP, SMTP, POP, IMAP, SNMP, Telnet, etc.

As it is seen from Table 3.1 ntop can run in both UNIX and Win32 operating system; hence, it installed using different installation steps for each operating system. Both UNIX and Win32 versions are developed under a single source-code tree, and require a library called libcap, which can also be downloaded from the official homepage. In the supported UNIX platforms, after having downloaded ntop's source code and installed libcap, ntop should be compiled and installed. In our case we installed ntop in Debaian Operating system using sequential steps. Refer Appendix I.



Global Protocol Distribution

Protocol	Data	Percentage			
IP	116.7 MBytes	72.4%	TCP	87.3 MBytes	74.8%
			UDP	20.9 MBytes	17.9%
			ICMP	8.2 MBytes	7.1%
			ICMPv6	0.6 KBytes	0%
			IGMP	282.0 KBytes	0%
			Other IP	64.8 KBytes	0%
(R)ARP	943.3 KBytes	0%			
IPv6	39.1 KBytes	0%			
Other	1004.8 KBytes	0%			

Figure 3.1: MU Global Protocol Distribution

The above Figure shows the traffic statistics for IP protocols. The traffic statistics report general information about the observed traffic. The traffic is considered from a global perspective, with no host-

specific information. From the Figure, it is possible to view the Global IP Protocol Distribution table and graph. The data collected by ntop shows that TCP and UDP are the highest bandwidth consuming protocols currently present in the Mekele University network. Together they account for 92.7% of the network usage. This sort of statistics is important for the administrator to understand the traffic, associating it to specific applications. Hence, in this way, it will be possible to manage the available bandwidth appropriately according the different applications running in MU network. Moreover, this protocol distribution can serve as input for developing bandwidth management tools which is going to be proposed in this thesis. The ntop installed in Mekele University web server generated the global protocol distribution for the TCP flow and the associated protocols as shown in the Figure below.

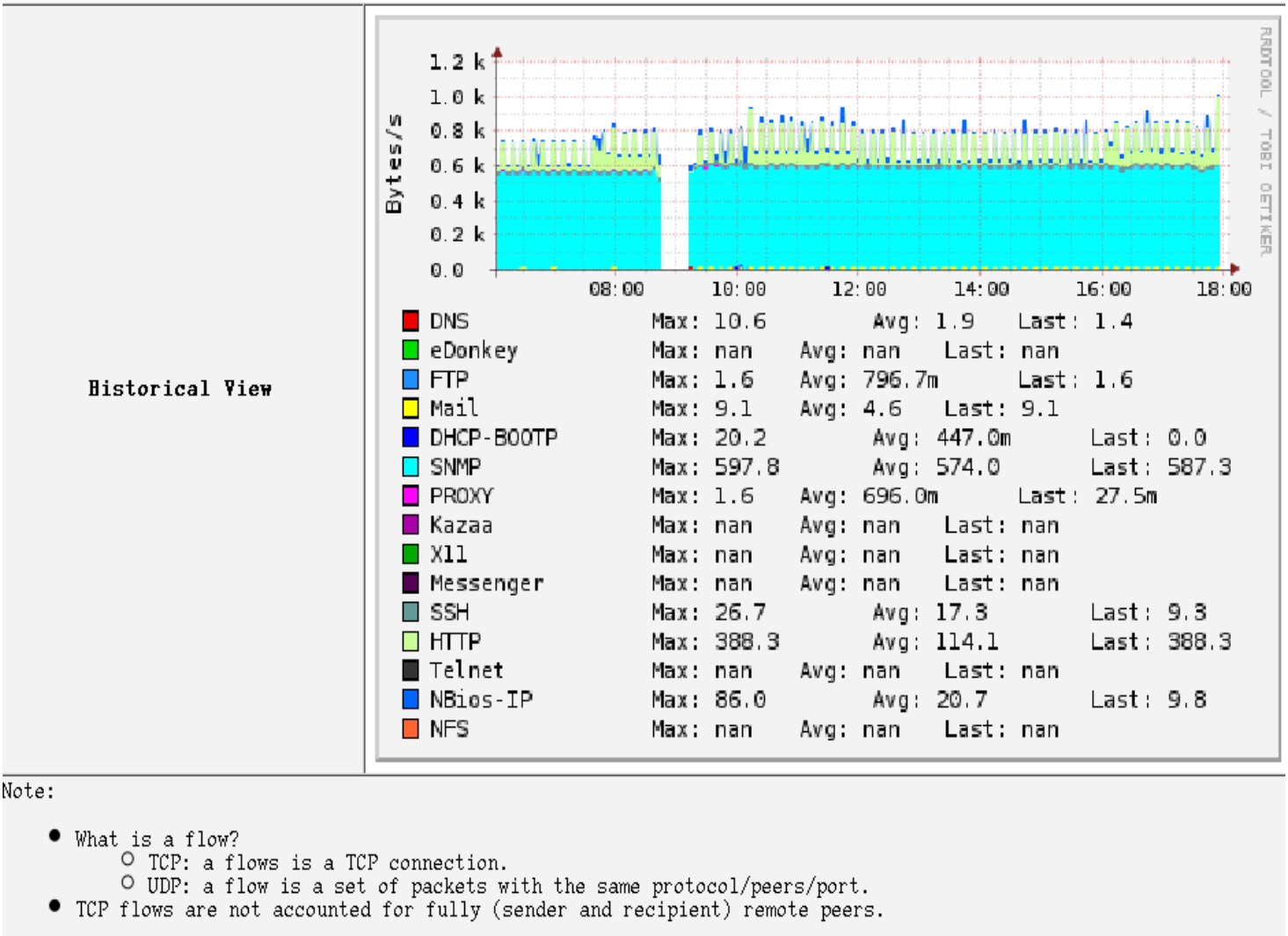


Figure 3.2: Historical View of TCP Flow

The above Figure shows the historical traffic statistics for TCP flow. The traffic statistics report historical TCP flow information about the observed traffic. The traffic is considered from a global perspective, with no host-specific information. The data collected by ntop shows that SNMP and HTTP are the highest bandwidth consuming TCP flows currently present in the Mekele University network. Together they account for 980.8 bytes/sec of the maximum network usage. This sort of statistics is important for the administrator to understand the traffic, associating it to specific applications and to propose possible solution for the bandwidth management in intercampus networks. As it is observed from Figure, ntop which is installed in Mekele University network do not collect data about all TCP flows. In the Figure, the network protocols with "nan" instead of some numeric value shows that the protocols or TCP flows such as eDonkey, Kazza, x11, Messenger, Telnet and NFS, are not required in the network. The ntop provides a guarantee to allow or disallow the analysis of specific TCP protocols i.e. this helps us to distinguish the most important protocols and unimportant protocols distributed over the intercampus network. Hence, the global protocol distribution is used to identify the important protocols with the percent coverage so as to be input for bandwidth management in the intercampus network. Moreover, the ntop collects data about each and every specific protocol in the network. In this case it collects data about DNS with maximum of TCP flow 10.6 bytes/sec, FTP with maximum of TCP flow 1.6.bytes/sec, DHCP-BOOTP with maximum of TCP flow 20.2 bytes/sec, Proxy with maximum of TCP flowb1.6 bytes/sec, SSH with maximum of TCP flow 26.7 byes/sec and NBios-IP with maximum of TCP flow 86.0 bytes/sec.

Having this protocol distribution we analyzed the ingoing and outgoing traffic in terms of chronological periods i.e. in terms of days, weeks, months and years. Hence, this helps us to develop the bandwidth management with predictive capability based on the historical traffic analysis. The following sections give brief analysis on network traffic based on daily, weekly, monthly and annually using MRTG.

3.4.2 Network Traffic Analysis using MRTG

Displaying statistical graphics to assess the use of the internet access band is considered an optional feature of a router. Yet, it is important to know this information to understand whether in internet access there are inefficiencies due to poor band distribution among the traffic types (VoIP, WWW, P2P, FTP ...) competing to use the internet connection. Lots of routers use SNMP to export the value of incoming and outgoing traffic counters for each of the network interfaces. Using software such as MRTG it is possible to repeatedly, and at regular time intervals, run SNMP queries towards these routers and save the traffic counters. Once this is done, MRTG enables the graphic analysis, via a browser, of incoming and outgoing traffic progression from the router interfaces [35].

Institutively, the Ethiopian universities do not deploy any network traffic analysis techniques in their Campus Network. This is due to the fact that

- Network administrators do not pay adequate attention for the traffic load analysis as they concentrate on network infrastructure expansion and its facilities.
- Moreover, the Universities are not as such informed about this valuable task of network administrators.
- Network administrators do get enough training even though they do have internal ambition.
- Network traffic analysis tools can be costly even though there are open source software which are licensed under the GNU.

In October 2010 first week, the Mekele University, Ethiopia, had one 3.04 Mbit Internet link for more than 2000 networked computers. As it was not possible to get a faster Internet link for another year, it was desirable to at least provide the administration and users on Mekele University campuses with current and detailed information about the status of the link.

This situation prompted the deployment of the Multi Router Traffic Grapher at Mekele University campus. Every five minutes, it queried the octet counters of the university's Internet gateway core switch, firewall and router. From this data, the average transfer rate of the internet link was derived for every five minute interval and a web page was generated with four graphs for each device showing the transfer rates for the last day, week, month, and year. The visual presentation on the Web allowed everyone with a web browser to monitor the status of the link. In next pages we presented an MRTG generated web page about each link with full description. While the availability of these graphs did of course not increase the capacity of the link, the performance data provided by MRTG proved to be a key argument to convince management that a faster Internet link was indeed needed.

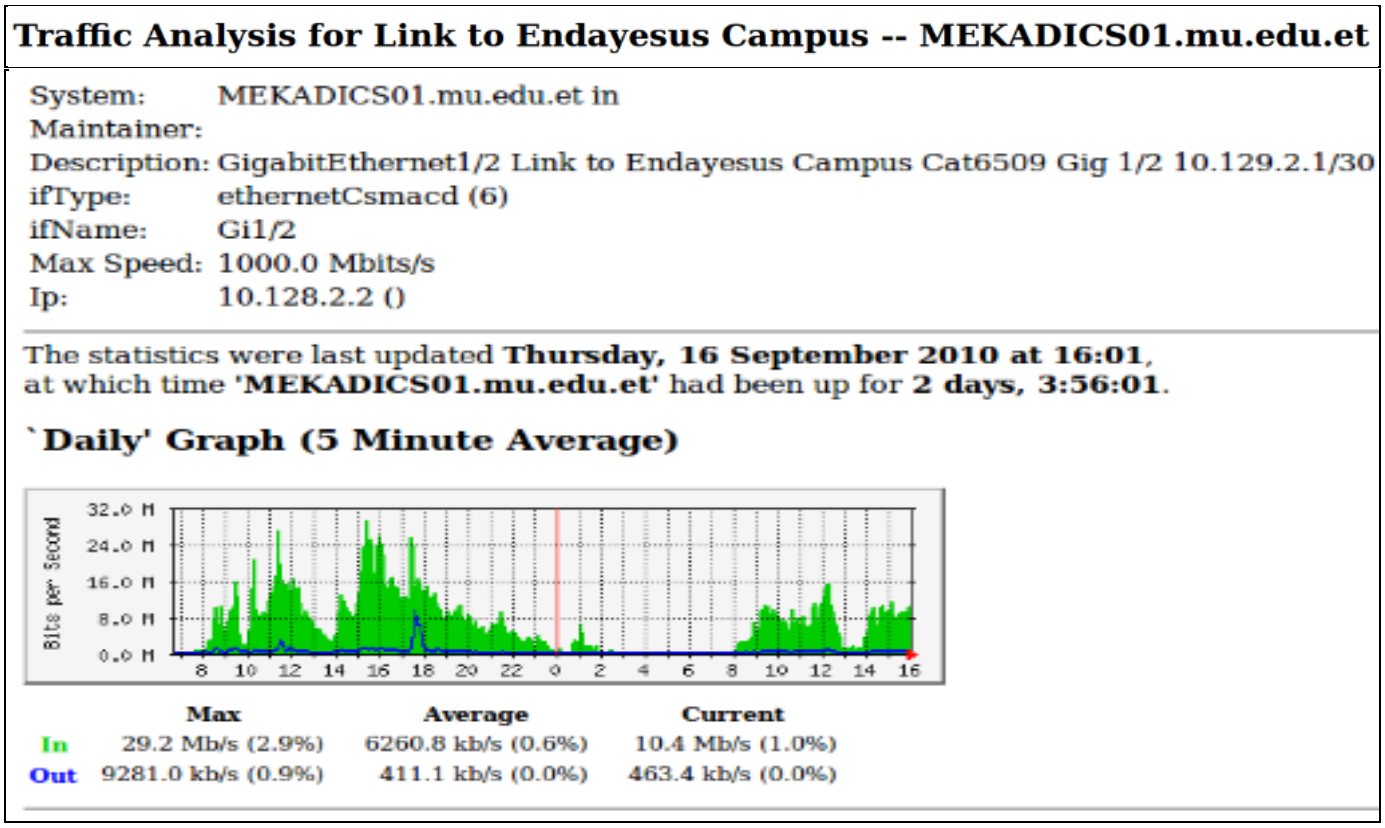


Figure 3.3: MRTG Daily Traffic Graph for MU Network at Links

The MRTG graph is influenced by the socio-economic activities on the university campus. Academic activities start at 8.00GMT at the morning for regular program; and 18:00 GMT at noon for extension

program. Here is a pronounced increase start at 8.00GMT when most lecturers arrive in the offices along with non academic staffs; by 16:00 GMT the campus is fully active and usage slightly decreases till 0:00 noons, with a dip, when people leave their offices during the evening. A steady rise in traffic by 14 GMT, when the break or lunch time is over and users are back in their offices and then a steady drop when users finally leave at about 18:00 GMT. Some academic staff and students typically work till midnight. Network usage is at its lowest from 0:00 GMT midnights local time till 8:00 GMT at the dawn local time and starts its cycle from 8:00 GMT once again. As it is depicted in the graph there is high traffic load usage after lunch time till evening.

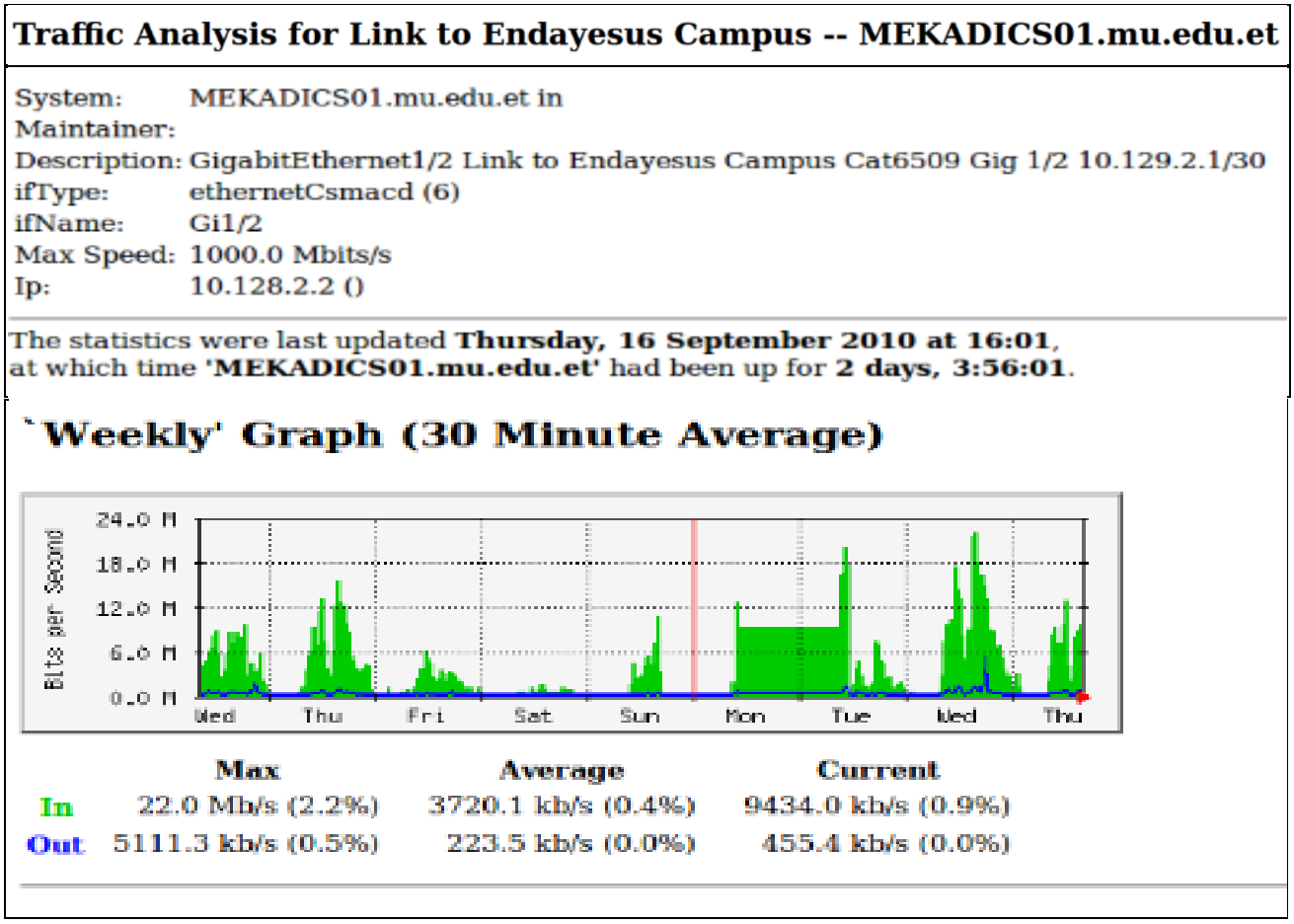


Figure 3.4: MRTG Weekly Traffic Graph for MU Network at Links

As it is stated above the MRTG is influenced by socio economic of the community of the University. Due to this fact the above Figure 3.4 shows that the weekly traffic load of the router link to Endayesus

Campus measured at 16 September 2010 reaches its peak by Wednesday. As it is shown from the MRTG graph there was about 22.0 Mbps, which is 2.2% of the traffic, recorded on Wednesday as maximum ingoing traffic. Moreover, we can notice the following facts from the Weekly MRTG graph of Mekele University Campus Network.

- There was 5.11Mbps, which is 0.5% traffic, outgoing traffic load.
- The graph shows flat traffic load during Monday and Tuesday of the week of September. This situation happens due to absence of electric power in the campus. If this happens the MRTG keeps the maximum traffic load attained steadily till the power is on in the campus. From the graph we can see that the MRTG attained 6.6 Mbps from Monday till Tuesday.

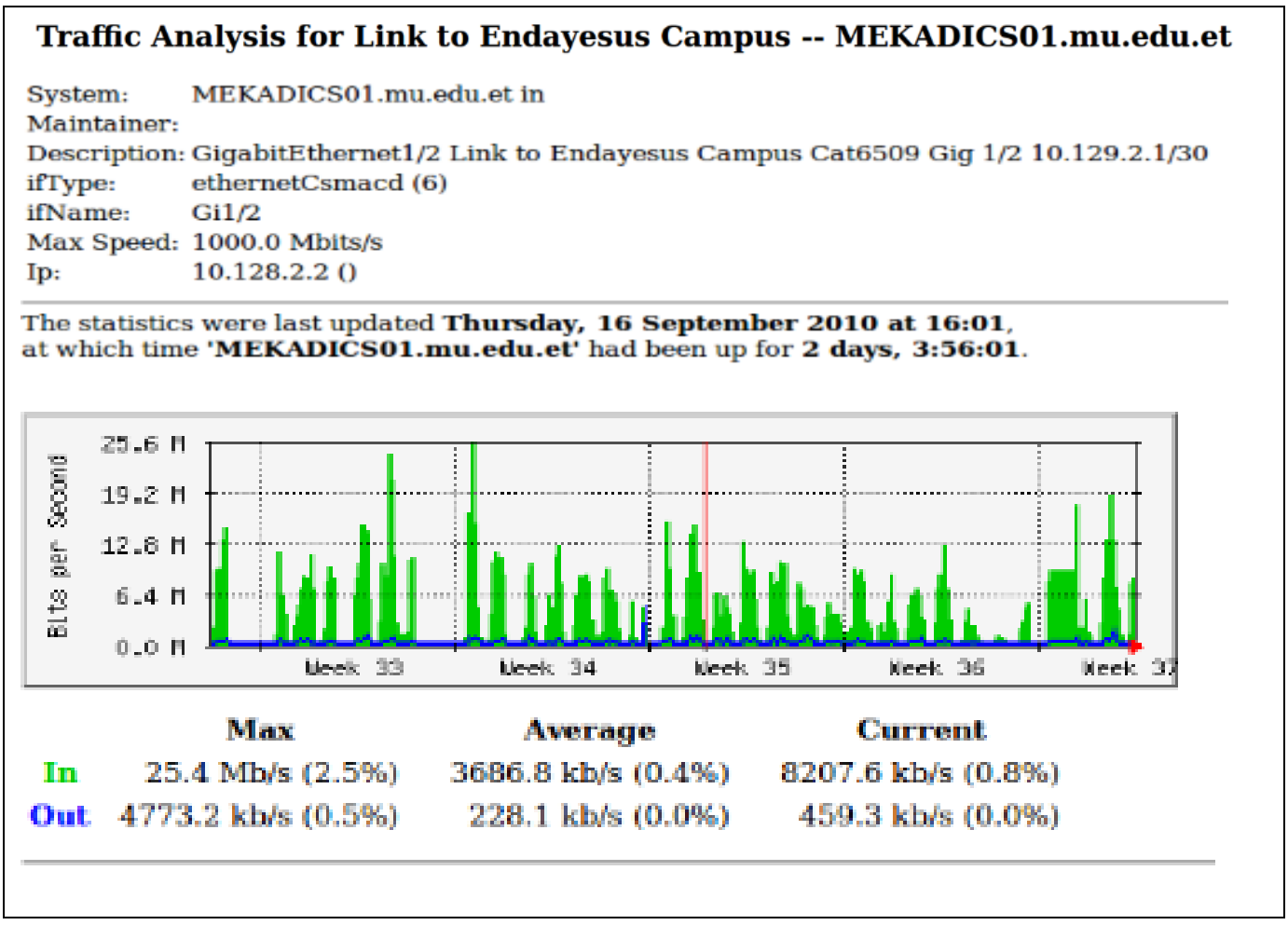


Figure 3.5: MRTG Monthly Traffic for MU Network at Links

From the above Figure 3.5 we derived the following facts.

- There was 25.4 Mbps maximum ingoing traffic load which is recorded at the week of 34.
- There was 4.773 Mbps maximum outgoing traffic load which is recorded at the end of week 34, 2010.
- There was 3.6868Mbps average ingoing traffic load.
- There was 228.1 kbps average outgoing traffic load.

3.4.3 Network Traffic Analysis at Firewalls using MRTG

NCSA believes that a firewall is a system or a complex of several systems that does some limitation between two or several networks. As a matter of fact a firewall tries to protect one inner network from another one by limiting the accesses between them. The firewall processes the packets and recognizes the unacceptable transfers according to its predefined security policy. Of course the firewalls are not the total security solutions. They have some drawbacks, and some insecurities or intrusions are out of firewall abilities, so the administrators have to use some physical security issues or host dignities, and so on [44]. In this paper we investigated the visualization of network traffic across the Firewall installed in Mekele University network infrastructure.

[45] describes a firewall is simply a system designed to prevent unauthorized access to or from a private network. Firewalls can be implemented in both hardware and software, or a combination of both. Firewalls are frequently used to prevent unauthorized Internet users from accessing private networks connected to the internet. All data entering or leaving the Intranet pass through the firewall, which examines each packet and blocks those that do not meet the specified security criteria.

Generally, firewalls are configured to protect against unauthenticated interactive logins from the outside world. This helps prevent hackers from logging into machines on your network. More sophisticated firewalls block traffic from the outside to the inside, but permit users on the inside to communicate a little more freely with the outside. Firewalls are also essential since they can provide a single block point where security and audit can be imposed. Firewalls provide an important logging and auditing function;

often they provide summaries to the admin about what type/volume of traffic that has been processed through it. This is an important point: providing this block point can serve the same as an armed guard cans [45].

In our work we tried to see online point firewall traffic statistics graphed using MRTG and offline descriptions based on the firewall access or traffic log file.

As the primary perimeter defense for most networks, firewalls can often be an import intrusion detection and forensic tool. So, for those serious about information security, understanding firewall logs is extremely valuable [45]. Accordingly we collected the most useful log entries from the firewall access log and we found, in the main log file, “Built”, “Deny”, and “Teardown” to show the accessed entries, denied entries and terminated calls respectively. These entries are especially useful for seeing port scans, host sweeps, and general probing. Check point gives you deny or drop alerts when traffic is not allowed and accept alerts when it is. Teardown means to drop the packet, whereas deny means to send a TCP reset or ICMP port/protocol unreachable message.

```
4|Sep 15 2010|16:35:06|106023|65.49.42.199|10.136.192.36|Deny tcp src outside:65.49.42.199/80 dst inside:10.136.192.36/43239 by access-group "outside_access_in" [0x0, 0x0]
4|Sep 15 2010|16:35:06|106023|65.49.42.199|10.136.192.36|Deny tcp src outside:65.49.42.199/80 dst inside:10.136.192.36/43239 by access-group "outside_access_in" [0x0, 0x0]
6|Sep 15 2010|16:35:06|302013|63.111.29.131|10.128.5.25|Built outbound TCP connection 2943733 for outside:63.111.29.131/80 (63.111.29.131/80) to inside:10.128.5.25/34246 (10.136.192.36/43464)
6|Sep 15 2010|16:35:06|305011|10.128.5.25|10.136.192.36|Built dynamic TCP translation from inside:10.128.5.25/34246 to outside:10.136.192.36/43464
6|Sep 15 2010|16:35:06|305012|10.128.5.25|10.136.192.36|Teardown dynamic TCP translation from inside:10.128.5.25/40013 to outside:10.136.192.36/42801 duration 0:00:30
6|Sep 15 2010|16:35:06|302013|69.147.97.211|10.128.5.25|Built outbound TCP connection 2943732 for outside:69.147.97.211/80 (69.147.97.211/80) to inside:10.128.5.25/40187 (10.136.192.36/65269)
4|Sep 15 2010|16:35:06|106023|65.49.42.199|10.136.192.36|Deny tcp src outside:65.49.42.199/80 dst inside:10.136.192.36/43239 by access-group "outside_access_in" [0x0, 0x0]
4|Sep 15 2010|16:35:06|106023|65.49.42.199|10.136.192.36|Deny tcp src outside:65.49.42.199/80 dst inside:10.136.192.36/43239 by access-group "outside_access_in" [0x0, 0x0]
```

Figure 3.6: MU Sample Firewall Log File

As it is seen from the above MU sample firewall log file Sep 15 2010|16:35:06|106023|65.49.42.199|10.136.192.36|Deny tcp src outside:65.49.42.199/80 dst inside:10.136.192.36/43239 by access-group "outside_access_in" [0x0, 0x0] is a log entry triggered by the unauthorized hitting the outside of MU firewall. In the full Internet-connected firewall log file we saw lots of these type log entries, as the number of these alerts went from nearly zero on average to hundreds of thousands per day on September 15th, 2010. Hence, security mechanism of MU network architecture should be as strong as possible and the university should not take firewall as sole network security mechanism.

As it is seen from the above MU sample firewall log file 6|Sep 15 2010|16:35:06|302013|69.147.97.211|10.128.5.25|Built outbound TCP connection 2943732 for outside:69.147.97.211/80 (69.147.97.211/80) to inside:10.128.5.25/40187 (10.136.192.36/65269) is a log entry for permitted/built HTTP traffic sourced from inside to the outside.

As it is seen from the above MU sample firewall log file 6|Sep 15 2010|16:35:06|305012|10.128.5.25|10.136.192.36|Teardown dynamic TCP translation from inside:10.128.5.25/40013 to outside:10.136.192.36/42801 duration 0:00:30 is a log entry for Teardown HTTP traffic sourced from inside to outside of the network.

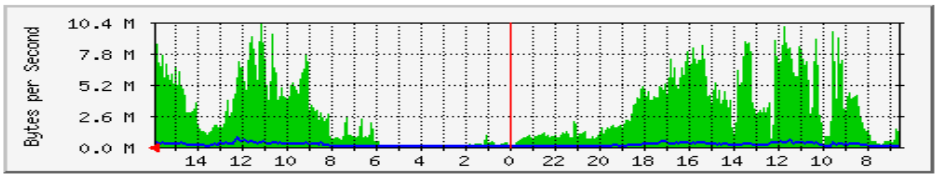
The above firewall log entry illustrates how using Firewall security system in MU for HTTP traffic allows for more in-depth IDS and forensic analysis. Moreover, in this paper we tried to assess the online visual Network traffic analysis of Firewall security system. In this case we deployed MRTG in the MU hardware firewall and generated the firewall traffic. The MRTG works in the same principle as it is used for the other network devices such as core switches, routers and servers. Accordingly, the MRTG generated the daily, weekly, monthly, and yearly traffic for both inside and outside interfaces of firewall.

Traffic Analysis for 1 -- mu-pix.mu.edu.et

System: mu-pix.mu.edu.et in
Maintainer:
Description: Cisco-PIX-Security-Appliance-'outside'-interface
ifType: ethernetCsmacd (6)
ifName: outside
Max Speed: 12.5 MBytes/s
Ip: 10.136.192.36 ()

The statistics were last updated **Thursday, 16 September 2010 at 15:55**, at which time **'mu-pix.mu.edu.et'** had been up for **2 days, 22:21:36**.

`Daily' Graph (5 Minute Average)



	Max	Average	Current
In	10.4 MB/s (83.0%)	3012.5 kB/s (24.1%)	6878.1 kB/s (55.0%)
Out	738.8 kB/s (5.9%)	165.5 kB/s (1.3%)	313.7 kB/s (2.5%)

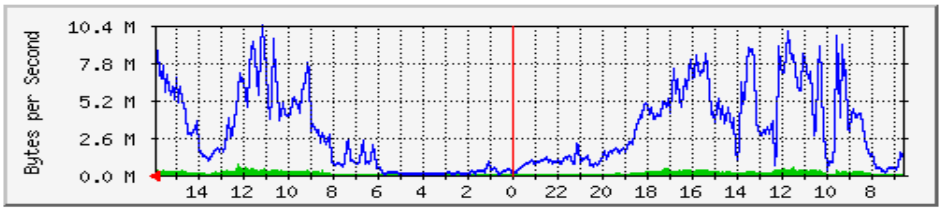
Figure 3.7: MRTG Daily Traffic for MU Network Firewall outside Interface

Traffic Analysis for 2 -- mu-pix.mu.edu.et

System: mu-pix.mu.edu.et in
Maintainer:
Description: Cisco-PIX-Security-Appliance-'inside'-interface
ifType: ethernetCsmacd (6)
ifName: inside
Max Speed: 12.5 MBytes/s
Ip: 10.128.2.17 ()

The statistics were last updated **Thursday, 16 September 2010 at 15:55**, at which time **'mu-pix.mu.edu.et'** had been up for **2 days, 22:21:36**.

`Daily' Graph (5 Minute Average)



	Max	Average	Current
In	739.8 kB/s (5.9%)	166.5 kB/s (1.3%)	315.4 kB/s (2.5%)
Out	10.4 MB/s (82.9%)	3002.6 kB/s (24.0%)	6857.7 kB/s (54.9%)

Figure 3.8: MRTG Daily Traffic for MU Network Firewall inside Interface

As stated above the MRTG generate the traffic for the inside and outside interfaces of the firewall. Hence, Figure 3.7 shows the firewall outside interface daily traffic. It clearly depicts that the daily maximum input traffic amounts to 10.4 MB/S and the daily maximum out traffic amounts to 738.8 KB/S. MRTG generate balanced traffic in the firewall inside interface as shown in Figure 3.8. The graph clearly depicts the daily maximum input traffic which amounts to 739.8KB/S and daily maximum out traffic which amounts to 10.4 MB/S which are the vice versa of the outside interface of the firewall.

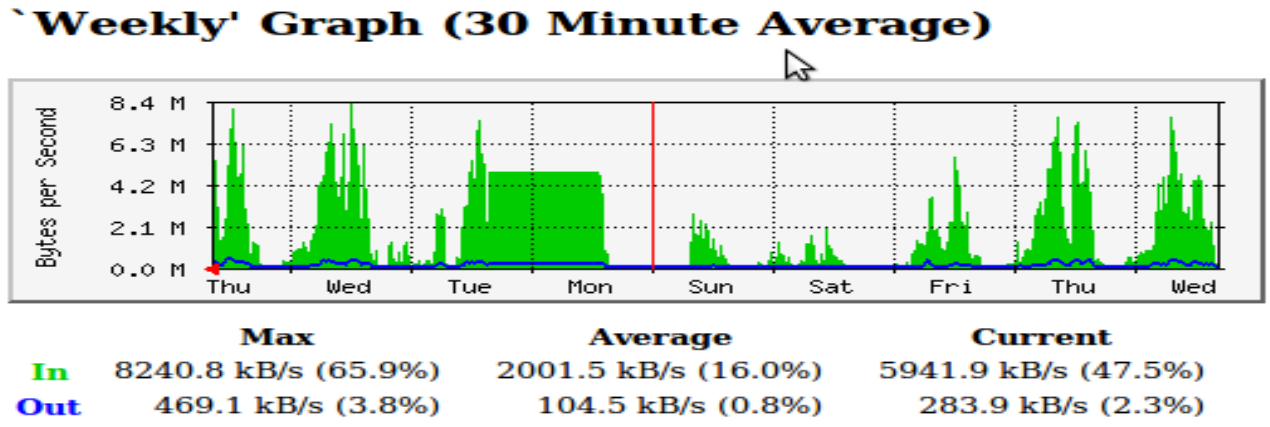


Figure 3.9: MRTG Weekly Traffic for MU Network Firewall outside Interface

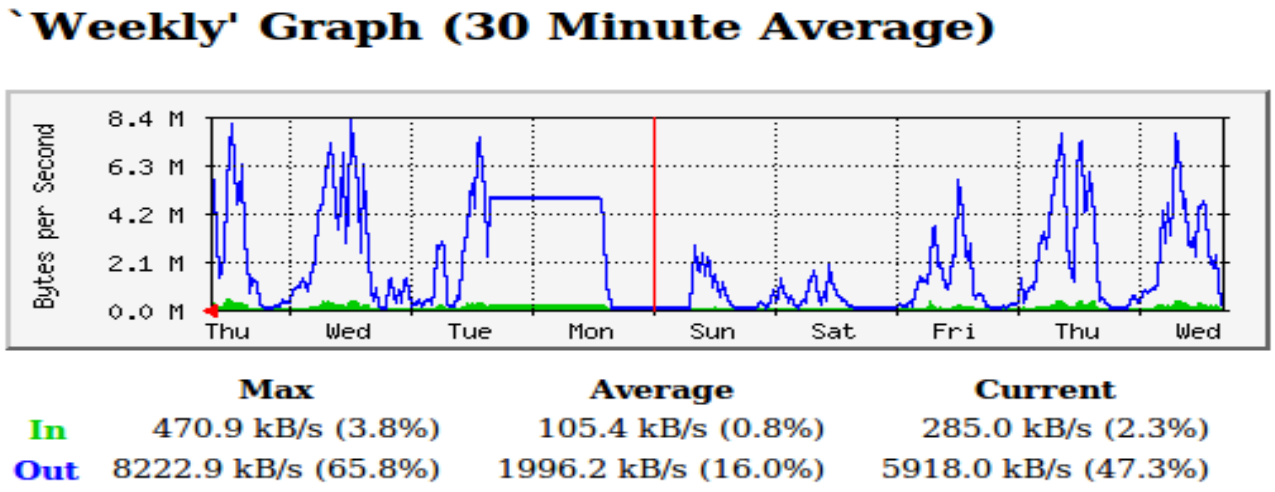


Figure 3.10: MRTG Weekly Traffic for MU Network Firewall inside Interface

As it is stated above the MRTG is influenced by socio economic of the community of the University. Due to this fact the above Figure 3.9 shows that the weekly traffic load of the MU firewall measured at 16 September 2010 reaches its peak by Wednesday, Thursday of the week with weekly average input traffic of 2001.5kB/s and weekly average output traffic of 104.5kB/s. MRTG generate balanced traffic in the firewall inside interface as shown in Figure 3.10. It clearly depicts the weekly average input traffic which amounts to 105.4 KB/S and weekly average output traffic which amounts to 1996.2 kB/S which are the vice versa of the outside interface of the firewall. Figure 3.9 shows flat traffic load during Monday noon and Tuesday Morning of the second week of September. This situation happens due to absence of electric power in the campus. If this happens the MRTG keeps the maximum traffic load attained steadily till the power is on in the campus. From the graph we can see that the MRTG attained about 4.6 Mbps from Monday noon till Tuesday morning.

Generally, MRTG enabled us to assess daily, weekly and monthly network traffic flow at the hardware firewall of MU campus network. Moreover, we tried to visualize the yearly traffic flow of the campus; but, we found it as it is insufficient enough for analysis of our paper.

3.5 Analysis of Web Server Logs: Case MU

We used web server log analyzer called Web Expert 7.0 to analyze sample web server logs obtained from Mekele University web server. The key information obtained include Total Hits, Visitor Hits, Average Hits per Day, Average Hits per Visitor, Failed Requests, Page Views, Total Page Views, Average Page Views per Day, Average Page Views per Visitor, Total Visitors, Average Visitors per Day, Total Unique IPs, Bandwidth, Total Bandwidth, Visitor Bandwidth, Average Bandwidth per Day, Average Bandwidth per Hit, Average Bandwidth per Visitor, Access Data like files, images etc., Referrers, User Agents etc. A few things to note about web log analysis:

- Each line in the file represents a single hit on a file on the web server, and consists of a number of fields.
- A web page hit is a page view, not same as a web server "hit". For example, if a web page contains 5 images, a hit on that page will generate 6 "Hits" on the web server, one hit for the web page, 5 hits for the images.
- A unique visitor is determined by the IP address or cookie. By default, a visit session is terminated when a user falls on inactive state for more than 30 minutes. So a unique user may visit your web site twice and get reported as two visits. If the visitor left the web site and came back 30 minutes later, Web Log Expert Analyzer will report 2 visits. If the visitor came back within 30 minutes, Web Log Expert Analyzer will still report 1 visit.
- The log file is in NCSA combined log format. The W3C maintains a standard format for web server log files, but other proprietary formats exist. The combined Log Format (CLF) is a fairly basic form of web server logging. It tracks nine different elements of the web transaction. Each request is written to one line, with the different elements of the request separated by spaces (items in quotes or square brackets are considered one item), and items that aren't sent are listed as a hyphen or dash (-).

The web server log file obtained from Mekele University is Combined Log Format. Combined Log Format is quite similar to the NCSA Common log format which contains only basic HTTP access information. The NCSA Combined Log is the second of three logs in the NCSA separate log format. The Combined log contains the requested resource and a few other pieces of information like referral, user agent, and cookie information. The information is contained in a single file. The fields in the combined log file format are host, login, username, date: time, request, status code, bytes, referrer, and user agent [1] .The following example shows these fields populated with values in a combined log file record.

125.125.125.125 - -

[03/Oct/2010:21:15:05 +0300]

"GET /index.html HTTP/1.0"

200

1043

“http://www.mu.edu.et/“

"Mozilla/4.0 (compatible;MSIE 7.0; Windows NT 6.0; SLCC1; .NET CLR 2.0.50727; Media Center PC 5.0;.NET CLR 3.5.30729; .NET CLR 3.0.30618; InfoPath.2)"

Fig 3.11: MU Web Server Sample Hit Log file

The following is a description of the fields in the fields in the above sample hit of combined log format which is obtained from Mekele University web server.

Host (125.125.125.125 in the example): the IP address or host/sub-domain name of the HTTP client that made the HTTP resource request.

Remote User Name ("-" in the example): the identifier used to identify the client making the HTTP request. If no value is present, a "-" is substituted.

Username ("-" in the example): the user name (or user ID) used by the client for authentication. If no value is present, a "-" is substituted.

Date: Time time-zone ([03/Oct/2010:21:15:05 +0300] in the example): the date and time stamp of the HTTP request. The fields in the date/time field are:[dd/MMM/yyyy:hh:mm:ss +-hhmm] where the fields are defined as dd is the day of the month, MMM is the month, yyy is the year, :hh is the hour, :mm is the minute, :ss is the seconds, +-hhmm is the timezone.

Request ("GET /index.html HTTP/1.0" in the example): this is the HTTP request. The request field contains three pieces of information. The main piece is the requested resource (index.html). The request field also contains the HTTP method (GET) and the HTTP protocol version (1.0).

Status code (200 in the example): the status is the numeric code indicating the success or failure of the HTTP request.

Transferred Bytes (1043 in the example): the bytes field is a numeric field containing the number of bytes of data transferred as part of the HTTP request, not including the HTTP header.

Referring URL ("http://www.mu.edu.et/" in the example): this is the page the visitor was on when he/she clicked to come to this page.

Crawlers: A web crawler is a computer program that browses the World Wide Web in a methodical, automated manner or in an orderly fashion. Other terms for Web crawlers are ants, automatic indexers, bots, and worms or Web spider, Web robot, or Web scutter. This process is called Web crawling or spidering [58]. During analysis of MU web server log we found web crawlers such as MSN Bots, Yahoo Slurps, Google Bots and Baidu Spiders etc.

3.5.1 Weblog Analysis: General Website Statistics

Weblog expert provides general statistics that includes Hits, page views, visitors, and bandwidth usage of the Mekele University as shown in Table 3.2.

Hits: A log entry will generate a hit on the web server. This can include pages, images, animations, audio, video, downloads, PDF or Word documents or anything else that allow visitors to access. When a web browser loads a page, it also loads all the components referenced by that page. For example, if a web page contains 5 images, a visit on that page will generate 6 hits on the web server, one hit for the

web page, 5 hits for the images [60]. When trying to view what happened in a web log using Weblog Expert 7.0, there is a definite problem: each interaction from a user is a discrete transaction. The browser requests a page, and the server provides it. From the perspective of the web, the transaction is done. If the user asks for another page, there isn't any information retained from the last transaction to hook this transaction to the last one. According to the analysis the web log expert generates the total hits of 440,902, visitor hits of 428,702, spider hits of 12,200, Average Hits per Day of 73,483, Average Hits per Visitor of 35.80, Cached Requests of 24,475, and Failed Requests of 293,248.

Visit: A visit happens when someone or something or robot visits a site. It consists of one or more page views/hits. One visitor can have many visits to a site. A unique visitor is determined by the IP address or cookie. By default, a visit session is terminated when a user falls on inactive state for more than 30 minutes. So a unique visitor may visit a web site twice and get reported as two visits [60]. When trying to view what happened in a web log using Weblog Expert 7.0, it generates, as shown in Table 3.2, the Total Visitors of 11,974, Average Visitors per Day of 1,995, and Total Unique IPs of 2,313.

Page views: Page is any file or content delivered by a web server that would generally be considered a web document. This includes HTML pages (.html, .htm, .shtml), script-generated pages (.cgi, .asp, .cfm, etc.), image files (.jpeg, .gif, .png), javascript (.js) and style sheets (.css) are generally not considered to be pages. A page view is a request to load a single page of an internet site. On the World Wide Web a page request would result from a web surfer clicking on a link on another HTML page pointing to the page in question. This should be contrasted with a hit, which refers to a request for a file from a web server. There may therefore be many hits per page view since a page can be made up of multiple files [48]. Generally**,** each line in the log file is a hit. If the request is for an asp or html page, then it is considered a page view. It could be a request for graphic or other supplementary information. So, one page view could result in many hits if there are a number of GIF or JPG files on the page [47]. When trying to view what happened in a web log using Weblog Expert 7.0, it generates, as shown in Table

3.2, the Total Page Views of 67,673, Average Page Views per Day of 11,278, Average Page Views per Visitor of 5.65.

Bandwidth: Custom reports let network administrators search for specific hosts, or conversations using bandwidth. When trying to view what happened in a web log using Weblog Expert 7.0, it generates, as shown in Table 3.2, Total Bandwidth of 35.78 GB, Visitor Bandwidth of 35.57 GB, Spider Bandwidth of 213.32 MB, Average Bandwidth per Day of 5.96 GB, Average Bandwidth per Hit of 85.08 KB, and Average Bandwidth per Visitor of 3.04 MB.

Table 3.2: Summary Web Log Statistics Report for MU Web Server

Summary

Hits	
Total Hits	440,902
Visitor Hits	428,702
Spider Hits	12,200
Average Hits per Day	73,483
Average Hits per Visitor	35.80
Cached Requests	24,475
Failed Requests	293,248
Page Views	
Total Page Views	67,673
Average Page Views per Day	11,278
Average Page Views per Visitor	5.65
Visitors	
Total Visitors	11,974
Average Visitors per Day	1,995
Total Unique IPs	2,313
Bandwidth	
Total Bandwidth	35.78 GB
Visitor Bandwidth	35.57 GB
Spider Bandwidth	213.32 MB
Average Bandwidth per Day	5.96 GB
Average Bandwidth per Hit	85.08 KB
Average Bandwidth per Visitor	3.04 MB

3.5.2 Weblog Analysis: Website Activity Statistics

Weblog expert is a powerful weblog analysis tool that can help network administrators to generate website activity statistics. Website activity statistics includes daily, by hours of a day, by day of week.

Daily: shows activity by date of the week. The weblog expert generates custom report about website daily activity in terms of total hits, total page views, total visitors, total average visit length and

bandwidth per date of the duration as shown Table 3.3. When trying to view what happened in a web server log using Weblog Expert 7.0, it shows, as shown in Table 3.3, that there were high numbers of hits during Mon o4-oct-10 which is amounted to 107, 387 as well as Wed 06-Oct-10 which is amounted to 107,506. The lowest number of hits recorded is during Sat 02-Oct-10. From the collected web log data of Mekele University the highest page views recorded during Mon 04-Oct-10 and the lowest page views are recorded during Sat 02-Oct-10. The web log expert returns 67,673 total page views. The total visitors are determined by the IP address or cookies; accordingly the highest number of visitors recorded was on Wed 06-Oct-10 as well as the lowest number of visitors were recorded on Sat o2-Oct-10. The total average visit length is the average of visits of the visitors in the web site. Accordingly, the highest visit length is recorded in Sat 02-Oct-10 as well as the lowest visit length recorded during Thu 07-Oct-10. The bandwidth is the rate of data transfer in the network. The highest bandwidth recorded is on Mon 04-Oct-10 as well as the lowest bandwidth recorded is Thu 07-Oct-10.From the above analysis the network administrators should use Dynamic Bandwidth Allocator tools to allocate the available bandwidth based on week distribution demand. Moreover, network administrators should make daily activity analysis as the allocated bandwidth.

Table 3.3: Web Activity Statistics by Day of a Week for MU Web Server

Daily Activity

Date	Hits	Page Views	Visitors	Average Visit Length	Bandwidth (KB)
Sat 02-Oct-10	10,311	909	282	28:33	2,725,865
Sun 03-Oct-10	55,063	2,891	1,670	22:50	2,124,561
Mon 04-Oct-10	107,387	20,412	2,830	18:27	26,597,847
Tue 05-Oct-10	97,811	11,995	2,631	19:24	2,767,615
Wed 06-Oct-10	107,506	16,148	3,113	18:52	2,393,856
Thu 07-Oct-10	62,824	15,318	1,448	17:58	903,486
Total	**440,902**	**67,673**	**11,974**	**19:34**	**37,513,233**

By Hour of Day: shows Activity by hour of day. Weblog Expert generate website activity statistics of Mekele University website based on total number of hits, total number of page views, total number of visitors, and total bandwidth usage as, shown in Table 3.4, per the access time. According to Table 3.4 the highest number of hits recorded is 48,410 during 04:00 - 04:59, the highest number of page views recorded is 11,468 during 04:00 - 04:59, the highest number of visitors recorded is 1,151 during 04:00 - 04:59 and the highest number of bandwidth recorded is 22,793,478 during 23:00 - 23:59. Moreover, Table 3.4 shows the lowest number of hits recorded is 2,474 during 17:00 - 17:59, the lowest number of page views recorded is 98 during 17:00 - 17:59, the lowest number of visitors recorded is 78 during 17:00 - 17:59, and the lowest web site bandwidth usage recorded is 66,927 KB during 17:00 - 17:59. Generally, the weblog expert depicts that the Network administrators should allocate adequate enough bandwidth during 00:00- 05:59 and 20:00- 23:59 with minimum bandwidth 1 MB.

Table 3.4: Web Activity by Hour of a Day for MU Web Server

Activity by Hour of Day

Hour	Hits	Page Views	Visitors	Bandwidth (KB)
00:00 - 00:59	47,692	9,758	1,148	1,069,454
01:00 - 01:59	38,809	8,154	848	599,070
02:00 - 02:59	18,722	745	593	253,949
03:00 - 03:59	45,578	8,170	1,251	1,422,367
04:00 - 04:59	48,410	11,468	1,151	873,403
05:00 - 05:59	43,223	9,361	1,078	775,989
06:00 - 06:59	24,263	1,570	724	404,611
07:00 - 07:59	11,180	528	276	214,575
08:00 - 08:59	9,151	734	235	494,237
09:00 - 09:59	8,590	375	226	353,183
10:00 - 10:59	8,330	531	195	550,506

3.5.3 Weblog Analysis: Website Access Statistics

Weblog expert is a powerful weblog analysis tool that can help network administrators to generate website access statistics. Web stats are useful for web administrators to get a sense of the actual load on the server. This is useful for diagnostics and planning, and for detecting unusual behaviour that may require planning action. The goal of the administrator is to keep the server running smoothly under expected loads, while improving the speed and reliability of obtaining documents from the site. The best way to achieve this is to have browsers retrieve documents from places closer to where they will be used (and even from memory) than to get them from the disk on the server. It is only when the file is retrieved from the server that the server has the ability to keep track of the access [60]. [60] describes web access statistics includes custom reports such as on pages, files…

Pages: Accessed pages. It can be changed the list of page file extensions in Options > Analysis > Files in the Weblog Export Window. In this paper the Weblog expert generates the most popular pages and least popular pages. The Most popular pages are set of documents that are accessed by high number of visitor; the least popular pages with least number of visitors for each page in the Website as shown in Fig.3.12.

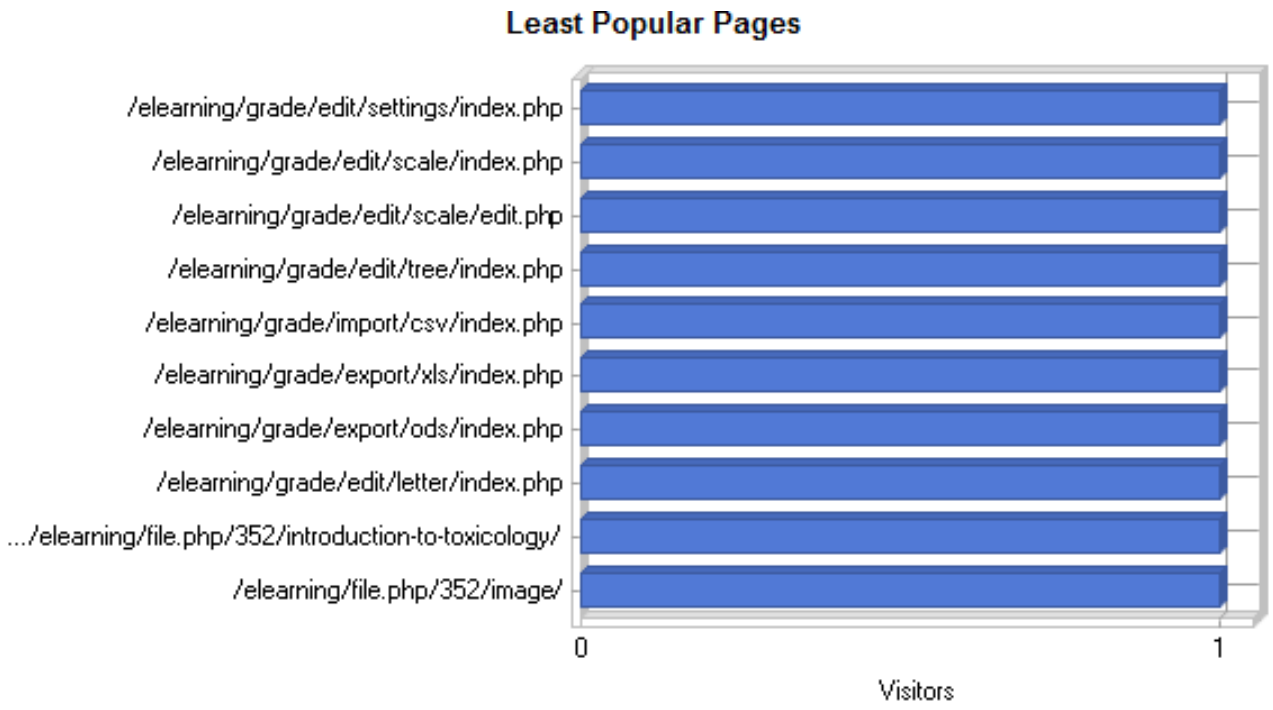


Figure 3.12: MU Least Popular Pages

Files: Downloaded files. You can change the list of download file extensions in Options > Analysis > Files in the Weblog Export Window. In a website, visitors write the approximate search query and the browsers provide with a number of responses; then, they can download as per the need of the user by clicking the links. As the website users download files, they need appropriate file readers such as Office packages, PDF readers. From MU website analysis we found that most of the most downloaded files are PDF files, as shown in Figure 3.16. Hence, each client in MU network infrastructure should have Acrobat Reader installed.

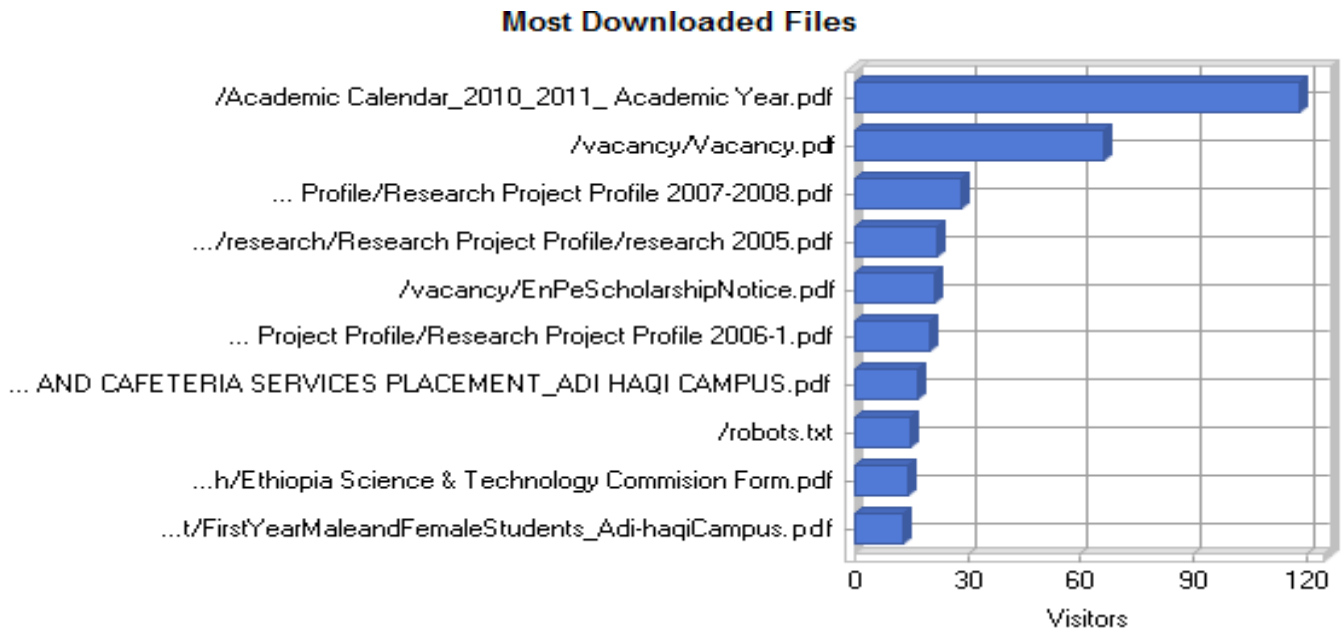


Figure 3.13: MU Most Downloaded Files

The web server log file analysis shows that a number of facts which clearly directs bandwidth allocation mechanism should be developed a specific for campus networks. In a number of researches has been conducted on this issue with different parameters and considerations. In this work we tried to develop dynamic bandwidth allocation conceptual framework with detailed description of the proposed model.

Current networking technology has two major limitations, [11], such as Availability of bandwidth cannot be predicted in terms of quantity or quality. Applications that require a specific quality of service, specifically bandwidth and delay, cannot currently reserve the quality of service they need. Moreover, it is difficult to control which applications or users get a share of the available bandwidth because it is impossible to differentiate between different types of traffic. An application or a user can take control of all the available bandwidth, preventing other applications or users from using the network.

To alleviate the above limitations of today's network a number of software tools have been developing such as, Solaris Bandwidth Manager i.e. a software product that controls the bandwidth allocated to particular applications, users, and organizations sharing the same intranet or Internet link.

Chapter Four

4.1 Developing Dynamic Bandwidth Management Conceptual Framework

Dynamic Bandwidth Allocator works by sorting traffic into classes based on application type; and the allocation of bandwidth is based on number of users at specific link, link traffic size and the available bandwidth within specified period according to the specified bandwidth policy in the intercampus network. It then schedules the traffic according to the waiting time or response time defined for each packet traffic class.

4.2 Proposed Bandwidth Allocator

The proposed Dynamic Bandwidth Allocator is going to be a software product that controls the bandwidth allocated to particular applications, users, and organizations sharing the same intranet or Internet link with especial focus in intercampus network. By allowing network administrators to guarantee minimum bandwidth and prioritize traffic, it enables them to prevent a small number of applications and users from consuming all available bandwidth and causing major network congestion. The proposed Dynamic Bandwidth Allocator is expected to meet the needs of campus ICT department who need to offer their campus community proper internet services. It enables the department to:

- Guaranteed bandwidth for education applications in the campus network.
- Reduce traffic congestion and increase network efficiency in the campus.
- Control users and applications in their access to network resources.
- Gather detailed network-use statistics based on campus, department level or VLAN level.

4.3 Dynamic Bandwidth Allocator Architecture

The proposed conceptual framework for Dynamic Bandwidth Allocator in intercampus network contains the following major components:

- The administration tool, which provides a graphical interface for configuring bandwidth allocation based on the different bandwidth demand in the intercampus network.
- The policy agent, which implements the configuration and handles communication with the kernel module.
- The kernel module, which implements the packet classifier, packet selector, Unique IP Address counter, host dynamic bandwidth allocator, bandwidth estimator and timer.
- The Report tool, which displays the report on the current status of bandwidth utilization in VLAN or in department level.
- Help tool which implements how to use the different tools or components of the Dynamic bandwidth allocator.

4.4 Bandwidth Allocation Key Features

- Incoming and outgoing traffic is managed, based on traffic type (telnet, FTP, e-mail, NFS, etc.), end-user source or destination address, or organization source or destination address.
- Bandwidth management rules can be configured to map organization, systems, or geographical layouts.
- The system will use database to store bandwidth management policies and dynamically retrieved by it. In the future we will integrate directory service with it that stores the bandwidth policies and important bandwidth criteria.

- Quality of service policies are dynamically allocated to remote users as they access the network, regardless of their dynamic IP address.
- Any type of UDP/TCP-based traffic can be managed.
- The system will run on top of WAN links as well as LAN links such as Ethernet.
- The system will provide a full set of reporting utilities to track how bandwidth is used, and how efficiently the traffic provisioning rules are working.

4.5 How Bandwidth Management Works

The Dynamic Bandwidth Allocator enables network administrators to manage the bandwidth used by IP traffic by:

- Allocating traffic to a class based on the application type, source and destination addresses, or a combination, then assigning individual limits for each class. For example, “traffic to engineering must have at least 50% of the link”, or “HTTP traffic cannot exceed 10% of the link”.
- Prioritizing traffic. Some types of traffic, for example, interactive traffic generated when using telnet or rlogin, need a quick response time. The system can assign a higher priority to that traffic. Traffic that does not require a quick response time, such as a file transfer using FTP, can be assigned a lower priority.

By balancing the bandwidth allocated to different types of network traffic and the relative priorities, network performance can be optimized.

4.6 Bandwidth Management Provisioning Rules

The provisioning rules used by the Dynamic Bandwidth Allocator to classify traffic are based on the following filters:

- Traffic type, corresponding to TCP/IP or UDP ports or services (HTTP, FTP, e-mail, news, telnet, NFS, etc.)
- Source IP address: Facilitates discrimination between traffic coming from different machines

Filters can be combined using hierarchies. For instance, a specific quality of service can be assigned to all the traffic originating from building X within the organization. Within this traffic, for example, a subset of the specified quality of service is reserved for e-mail.

Dynamic Bandwidth allocation based simply on filtering by protocol is not sufficient to meet bandwidth management needs. One of the key issues in this area is the extensive and increasing use of HTML/HTTP systems for e-commerce. A fine level of granularity is needed for bandwidth management to take into account more than just the protocol when assessing the relative importance of network traffic. Bandwidth management must base allocation not only on protocol type, but also on the network traffic and users involved.

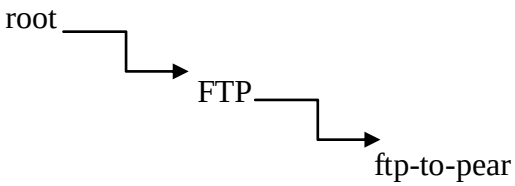
4.7 Bandwidth Allocation Provisioning Examples

Some examples of the specifications which can be defined using provisioning include:

- Ensure that telnet traffic from workstation A, B, or C has a minimum allocated bandwidth of 5% over a leased line.
- Reserve 50% of a WAN link to traffic that originates from offices A, B, and C. Then within this 50%, reserve 10% for NFS traffic.
- Ensure that FTP transfer from machine B in office Y will not use more than 30% of the capacity of the line between offices Y and X.

4.8 Network Traffic Classification

Class definitions are hierarchical, and every class has a parent. For example, if a class for FTP traffic and a class for FTP traffic to a host called pear is defined, the classes are connected in a hierarchy as shown below.



In this example, the FTP class is a child of the root class and is the parent of the ftp-to-pear class.
The Dynamic Bandwidth Allocator will specify the set of known classes for a network node defined in terms of some or all of these factors. It also allocates a percentage of bandwidth and a priority to each class.

4.9 Bandwidth Allocation policy

To allocate bandwidth in campus network it is the first thing to define policy of the bandwidth sharing. During defining the bandwidth allocation policy we consider the following factors.

Number of users in a certain domain: in this case, we are assuming that a network that implements Active Directory. The Active Directory uses the security principals (user, group and computer accounts) in the directory to secure their resources. Active Directory thus acts as an identity store, providing a single trusted list of Who's Who in the domain. Hence, Active Directory itself is more than just a database. It is a collection of supporting files that includes transaction logs and the system volume, or sysvol, that contains logon scripts and group policy information [10]. From the above we can count the number of users who are logon to the server. This helps to allocate the bandwidth based on number of users in a specific building, department or VLAN.

How can you identify whether a number of users are from same building or not? In this case the Active Directory keeps records of computer information. Hence, we can identify based on the source address. In the case of VLAN we identify the computer in which VLAN is based on the VLAN number. Having known the number of users in a specific building, department or VLAN, the network administrators can allocate the available bandwidth in comparison to the total number of users.

Allocate the available bandwidth based on the traffic type. During network traffic analysis in case of Mekele University we found that IP data summed up to be 116.7 Mbytes. From the ntop protocol distribution analysis, it is shown that TCP data covers the highest percentage which amounts to 87.3 Mbytes which is about 74.8%. The summarized ntop protocol distribution of Mekele University intercampus network is given as follows. For further details refer network analysis using ntop the previous chapter of this thesis.

Table 4.1: Summarized ntop Protocol Distribution for MU

Protocol	Data	Percentage (%)
TCP	87.3MB	74.8%
UDP	20.9 MB	17.9%
ICMP	8.2 MB	7.1%
HTTP	3.5 MB	3%
DNS	62.1KB	0%

Moreover, ntop protocol distribution analysis in case of Mekele University shows the TCP flow rate.

Table 4.2: MU Historical view of ntop Protocol Distribution

Protocol/class	TCP flow rate(Bytes/sec)	
	Max	Average
DNS	10.6	1.9
FTP	1.6	796.7n
Mail	9.1	4.6
DHCP-BOOTP	20.2	447.0n
SNMP	597.8	570.0
Proxy	1.6	696.0n
SSH	26.7	17.3
HTTP	388.3	114.1
NBiosIP	86.0	20.7

From the table we can notice that HTTP and SNMP are protocols/ traffic classes that account higher TCP flow rate. Hence, the network administrator should allocate better bandwidth of the respective link in the respective building.

As ntop protocol distribution analysis shows there are a number of protocols, in the case of Mekele university intercampus network, which are with “nan” TCP flows as shown in Figure 3.2. It indicates that there is no TCP flow. Hence, the network administrator should allocate none or less bandwidth for such type of protocols or traffic classes.

Generally, traffic classifier reads the traffic passing through the network and classifies it to any of the traffic class type and assigns the bandwidth amount according to the defined policy. We used the above traffic classes for developing Dynamic Bandwidth Allocator conceptual framework.

During defining policy for bandwidth management we identified the traffic type classes; then, we should prioritize the traffic class types. In this work, the prioritization of traffic class is based on the response time assumption. This prioritization can vary from department to department, from VLAN to VLAN or from building to building. In some department let’s say ICT department the “telnet” or “rlogin” can be given higher priority than other traffic as they are interactive traffic classes. Hence, the Dynamic Bandwidth Allocator can assign higher priority traffic. For example “telnet” traffic cannot be less than

20% of the link. In other case, for example economics department, the “HTTP” traffic class types can be allocated with higher bandwidth; hence, the Dynamic Bandwidth Allocator should allocate higher bandwidth of the corresponding link. On the other hand, traffic that does not need quick response time such as file transferring using FTP can be assigned a lower priority. By balancing the bandwidth allocated to different types of network traffic and the relative priorities, network performance can be optimized.

During developing the Dynamic Bandwidth Allocator conceptual framework we used the following provisioning rule to classify traffic.

- Traffic type, corresponding to TCP/IP or UDP or service types such as HTTP. FTP, Email, Telnet, NFS, DNS, etc.
- Source IP address: facilitates discrimination between traffic coming from different machines.
- Network or sub-network source or destination IP address: enables discrimination between specific organizations or based on network topology such as department, building or VLAN.

Remarks:

- In the above lists of provision rules traffic classification can be combined using hierarchies. For example a specific quality of service can be assigned to all the traffic originating from department, building or VLAN x, within this traffic, for example, a subset of specified quality of service is reserved for email.
- Bandwidth allocation based on simply on filtering by protocol or class type is not sufficient to meet bandwidth management needs. Hence, a fine granularity is needed for bandwidth management to take into account more than just the protocol when assigning the relative importance of network traffic. Bandwidth management must base allocation not only protocol type, but also the available link traffic and users involved.

4.10 Dynamic Bandwidth Allocator

Dynamic Bandwidth Allocator will contain the following major components:

- The administration tool, which provides a graphical interface for configuring bandwidth allocation.
- The policy agent implements the configuration and handles communication with the kernel module.
- The classifier and the scheduler.
    - The classifier allocates packets to a class queue according to information in the IP
    - The scheduler sends the queued packets according to the pre-defined priority and transfer rate for each class. It can also mark the type of service (TOS) fields of IP packets to a specific value.

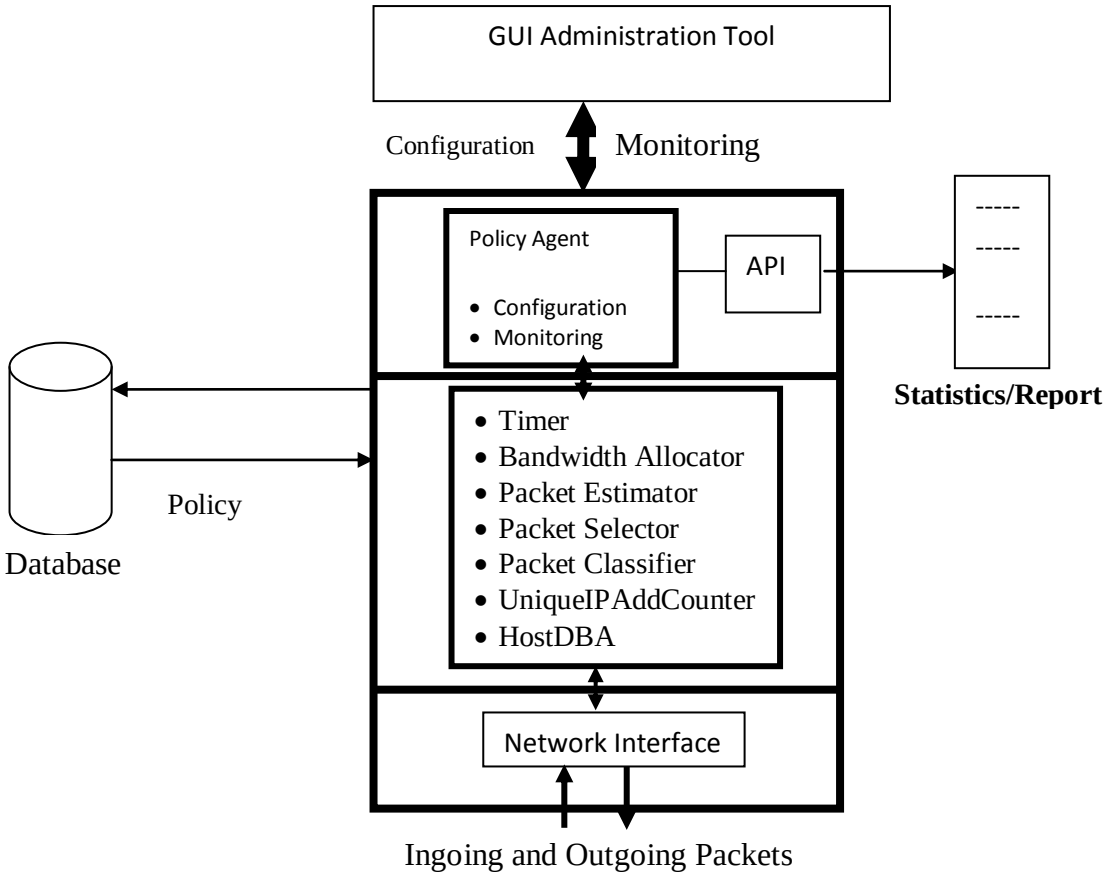


Figure 4.1: Dynamic Bandwidth Allocator Architecture

Administration Tool: The administration tool can be used to configure Solaris Bandwidth Manager. The administration tool communicates with the kernel module through the policy agent, and has two modes of operation:

- In online mode, the configuration currently being used by the kernel module can be changed. This is useful if an immediate temporary change is required due to a problem in the network. The current configuration may also be saved to preserve any changes. With online mode, the consequences of a particular configuration can be observed before being saved.
- In offline mode, a configuration can be changed without disturbing the current behavior of the kernel module. This is useful for making changes in the configuration without disrupting users, and having the changes implemented the next time the product is restarted.

The Policy Agent: The policy agent is the communications hub of the bandwidth allocator. It controls the information sent to and from all other components, as well as the policies that they operate.

The Packet Classifier: The packet classifier collects packets from the IP layer. It applies the filters defined by the provisioning rules to assign each packet to a class. The class of a packet is defined by IP source address, Protocol (TCP or UDP) or By service type (HTTP. FTP, DNS…)

In this paper we deployed service type packet classification in hierarchical link sharing mechanism to have fair bandwidth allocation amongst different buildings and users of intercampus network.

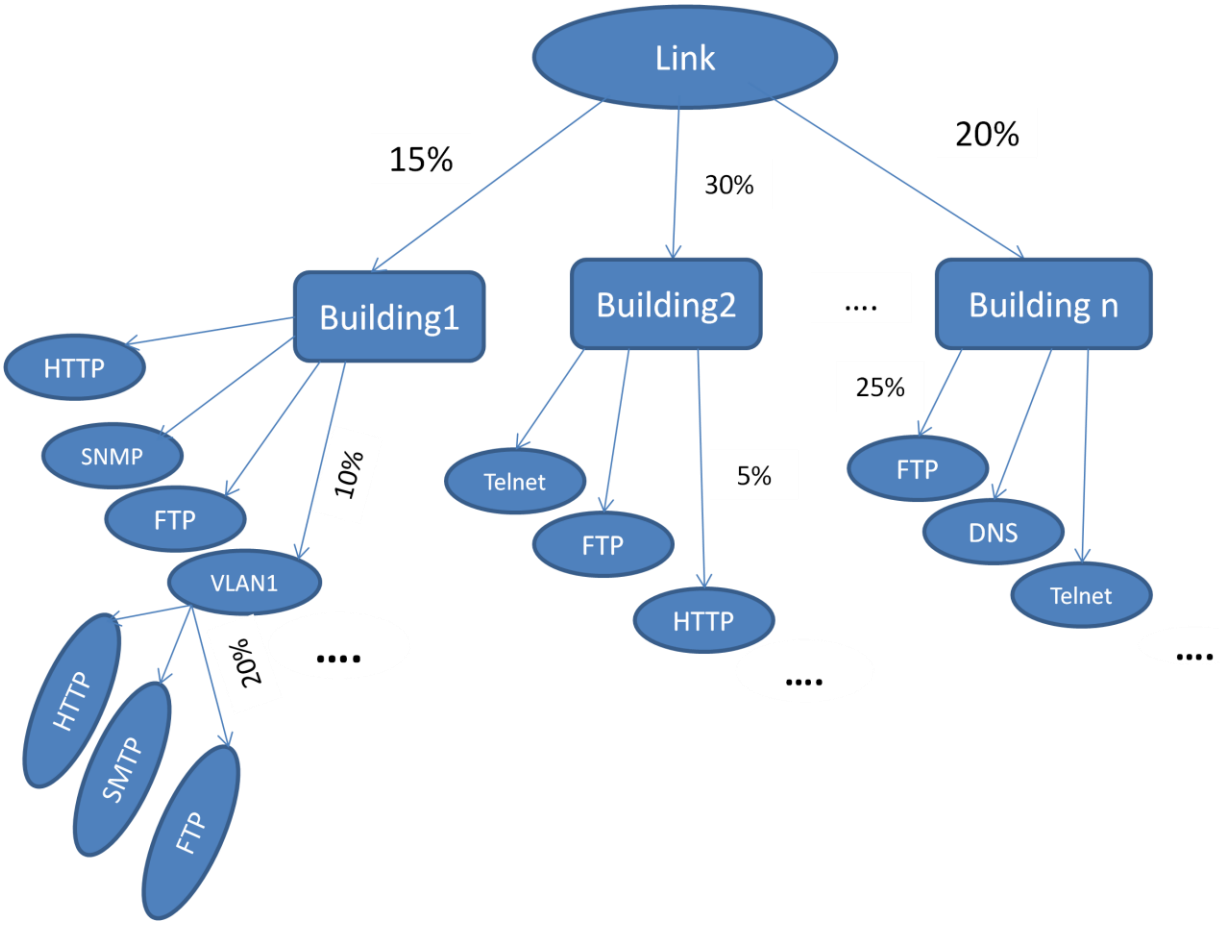


Figure 4.2: A Hierarchical Link-sharing Structure

The above figure shows the hierarchical bandwidth link sharing link in the intercampus network based on service type. As it is shown in the Hierarchical Link-sharing Structure each of the service types measured their current percent of traffic load in the respective link. From example if ftp traffic is originated from VLAN one in building 1, then its traffic load is measured at link that connects different nodes at VLAN one.

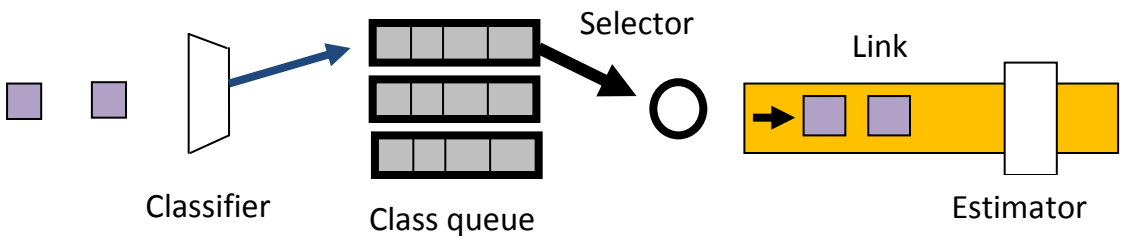


Figure 4.3: Class based queuing and bandwidth allocation.

According to [8] the purpose of class based bandwidth allocation is to let packets with higher priorities get the channel easier than those with lower priorities. Thus, one can set a higher priority to the real-time traffic to satisfy its Quality of Service requirements in throughput or delay bound etc. However, the priority setting may results in the starvation problem for users with lower priorities. Therefore, we propose a controlled link-sharing mechanism, which is similar to class based queuing proposed in [8], to guarantee that each service can get necessary channel allocation and to prevent traffic with lower priority from bandwidth starvation.

In this work we propose the dynamic bandwidth sharing algorithm based on different criteria. These criteria include number of users in respective geography, link bandwidth and queue of the traffic. Having these parameters the dynamism of the algorithm basically depends on the timer which calls the bandwidth allocation algorithm every five minutes. To fulfill the goal of Dynamic Bandwidth Allocator for intercampus network some modules such as classifier, Timer, estimator, selector, Unique IP address counter, and Host dynamic bandwidth allocation are required as shown in Fig. 4.3.

- A classifier classifies packets, assigns them class-IDs, and stores them into the class queue.
- The function of an estimator is to figure out the traffic flow in a certain period for each class. Then the information is used to estimate whether enough channels or not have been assigned.
- Using the information from the estimator, a selector has to decide which class queue is allowed to send packets so that each class satisfies its bandwidth requirement.
- Unique IP Address counter module is concerned with counting the unique IP addresses of outgoing packet traffic passing a network link in the campus network.
- Host dynamic bandwidth allocator module allocates the traffic of respective link among connected users based on the total user traffic measured at network interface.

- Timer counts the time which is obtained from the system and change its starts counting again with new initialization after five minutes. Moreover, the timer calls the Dynamic Bandwidth Allocator module every five minute and makes the allocation to the network traffic.

4. 11 Developing Dynamic Allocation Algorithms

Classifier: A classifier is used to classify arriving packets to appropriate classes which will be added to the header of every packet. According to the class-ID, packets are put into each class queue. The pseudo code of a classifier is shown below.

```
Packet Classifier
BEGIN
    GET a packet P passing network link
    SET i = P.classid
    INSERT P INTO traffic_class_queue[i]
END
```

Selector: A selector is used to determine which packet can be sent from one of the class queues. Thus, a selector can use the randomized packet_wait_timer to avoid collision. That is to say, each class queues have their own independent packet_wait_ timers. Once a packet_wait_timer of the class queue expires, then the packet in that queue can be sent. Moreover, the selector is used to call bandwidth estimator module to allocate the available bandwidth for the selected class queue. The pseudo code for the selector is shown below.

```
BEGIN
    FOR i = 0 TO count(class_queue)
        IF class_queue[i].packet_wait_timer is expired THEN
            IF class_queue[i].length > 0 THEN
                IF class_queue[i].status is not blocked THEN
                    CALL hand_shake_procedure();
                    CALL BandwidthEstimator(class_queue[i])
                    SEND packet_class_queue[i].
                ENDIF
            END IF
        END IF
    NEXT i
END
```

Bandwidth Estimator: is used to estimate the bandwidth utilization for each class or service type. We can add the class-ID of the packet to class queue and record the packet wait time. Thus, we can exchange the bandwidth usage information from network link terminal based on the allotted bandwidth amongst the different departments or VLAN.

Bandwidth Estimator

```
Function: BandwidthEstimator(Data type: P)
BEGIN
    READ allocated bandwidth of respective VLAN/Department allocBW
    GET p.currentPercent_ traffic_class_queue[id]
    p.Bandwidth= p.currentPercent_ traffic_class_queue[id]*allocBW
END
```

Timer: counts the time which is obtained from the system till there is five minutes time difference between starting time and system time obtained after five minutes; and, starts counting again with new initialization after five minutes. Moreover, the timer calls the Dynamic Bandwidth Allocator module every five minute and makes the allocation to the network traffic. Its algorithm is given below.

```
BEGIN
    LABEL Start:
    INIT T1=currentSystemTime();
    FLAG time=1;
    WHILE(time) DO
        READ NextCurrentTime as T2
        IF(T2-T1==5 minutes) THEN
            CALL BandwidthAllocator()
            FLAG time=0;//false
            GOTO Start;
        ENDIF
        NEXT WHILE
    ENDWHILE;
END
```

Algorithm for counting unique IP address

```
Function: uniqueIPAddressCounter()
BEGIN
    DECLARE: Global variables
        IPAddPanel[]; // store unique IP addresses with respect building or VLAN
        cntIPAddress=0; //counter for unique IP addresses
    FOR(each outgoing SYN packet at link)
        READ srcIPAddress;
        IF (IPAddressPanel is EMPTY) THEN
            IPAddPanel[0] = srcIPAddress;
            INCREMENT cntIPAddress;
        ELSE
            FOR (i=0 TO sizeOf(IPAddPanel[]))
                IF (IPAddPanel is EOF) THEN
                    IPAddPanel[sizeOf(IPAddPanel[])]= srcIPAddress;
                    INCREMENT cntIPAddress;
                ELSE
                    IF (IPAddPanel[i] <> srcIPAddress|| IPAddPanel[i] ==srcIPAddress) THEN
                        NEXT  i;
                    ENDIF
                ENDIF
            ENDFOR
        ENDIF
        NEXT packet
    ENDFOR
RETURN cntIPAddress;
END
```

Dynamic Bandwidth Allocation per Hosts or Users in the campus Buildings

Function: hostDynamicBandwidthAllocator(linkTrafficSize[], linkBandwidth[], linkNum)

BEGIN

DECLARE: Global Variable

numUser= uniqueIPAddressCounter(); // Calling function

FOR (cntUser = 0 TO numUser)

READ totalTraffic at IPAddressPanel[cntUser] as

totalTraffic[cntUser]= $HTTP_{traffic}$ + $DNS_{traffic}$ + $FTP_{traffic}$ + $SNMP_{traffic}$+ …

//READ totalTrafficSize at a link as

//trafficSize[cntLink]= $\sum_{i=0}^{numUSER} totalUserTraffic[i]$

userBandwidth[cntUser]=$\frac{totalUserTraffic\ \ [cntUser\ ]}{linkTrafficSize\ \ [linkNum\ ]}$ * linkBandwidth[linkNum]

CALL packetCalssifier();

CALL packetSelector();

NEXT cntUser

ENDFOR

END

Bandwidth Allocation Algorithm for campus Links

BEGIN

GET total campus allotted bandwidth B

GET totalTraffic passing at all Links of campus as totalTraffic= $\sum_{i=0}^{count\ (links\ )} trafficSize[i]$

FOR cntLink=0 TO count(links)

GET trafficSize[cntLink] passing at a Link.

linkBandwidth[cntLink]= $\frac{\text{tarffi cSize [cntLink]}}{\text{totalTraffic}}$ * B

checkVLAN= Is there VLAN at cntLink? (Yes/no)

IF(checkVLAN==Yes) THEN

FOR cntVlan=0 TO count(VLANs)

GET vlanTrafficSize passing at VLAN link.

vlanBandwidth[cntVlan]= $\frac{\text{vlanTrafficSize [cntVlan]}}{\text{trafficSize [cntLink]}}$ * linkBandwidth[cntLink]

CALL packetCalssifier()

CALL packetSelector()

hostDBA(vlanTrafficSize[cntVlan], linkBandwidth[cntVlan], cntVlan)

NEXT cntVlan

ENDFOR

ELSE

CALL packetCalssifier()

CALL packetSelector()

hostDBA(trafficSize[cntLink], linkBandwidth[cntLink], cntLink)

ENDIF

NEXT cntLink

ENDFOR

END

As it is stated in the above the dynamic bandwidth allocation consists of different components such as bandwidth administration tool as shown in the figure below.

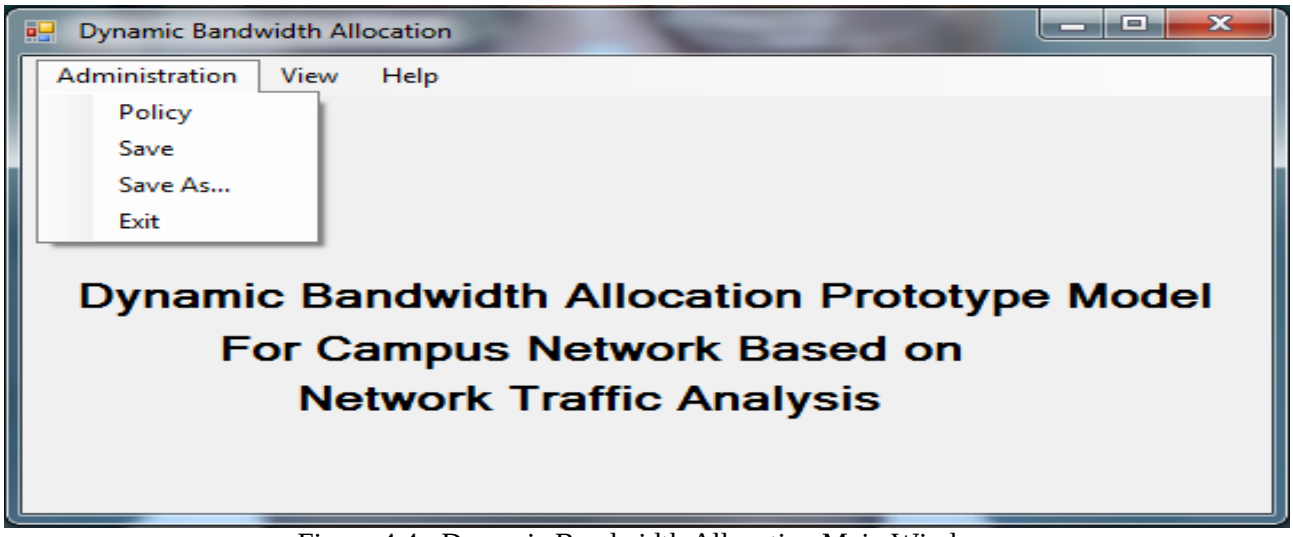


Figure 4.4 : Dynamic Bandwidth Allocation Main Window

As it is stated in the literature bandwidth is private good of an organization and it needs series management. The above window is developed to define bandwidth allocation policy for the campus network and it helps us to see the different statistics of VLAN, specific building and the campus statistics as a whole and it serves as Main form for displaying other windows such as bandwidth allocation policy window.

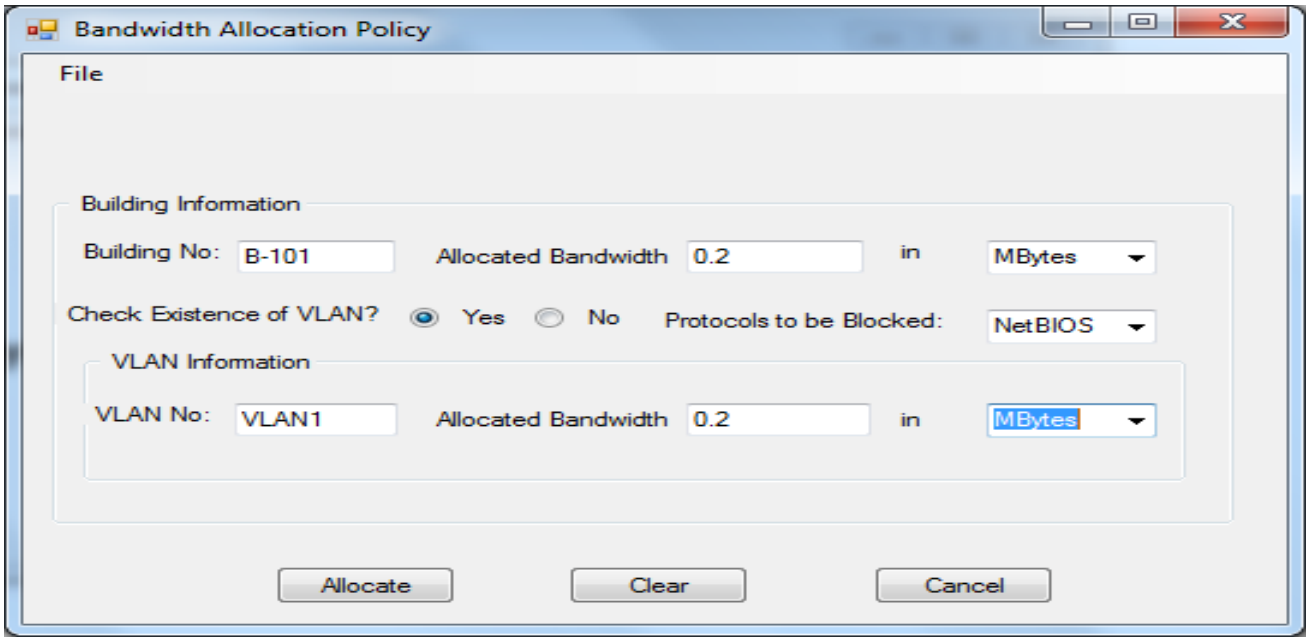


Figure 4.5: Bandwidth Allocation Policy Window

Bandwidth allocation policy window helps the network administrator to define the bandwidth for each building, to block unnecessary protocols and to allocate VLANs with appropriate bandwidth. During defining the bandwidth for the campus network, the administrator needs to fill the building number, the allocated bandwidth for each building, specifying the necessary protocols, specific VLAN number and allocated bandwidth.

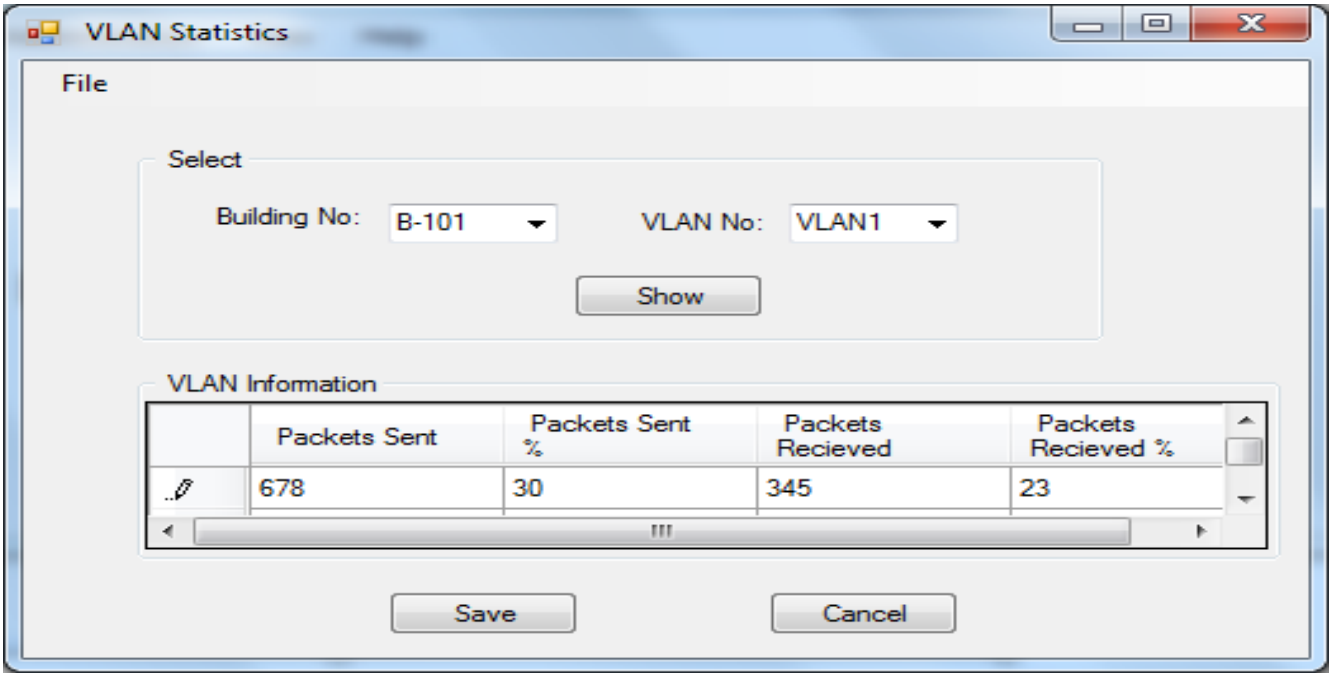


Figure 4.6: VLAN Statistics Window

VLAN statistics window displays the VLAN information in terms of packet sent, packet sent percent, packets received and packets received percent. Moreover, the window allows the network administrator to specify the building number and VLAN number to show the specific information; and it allows saving the VLAN information for report.

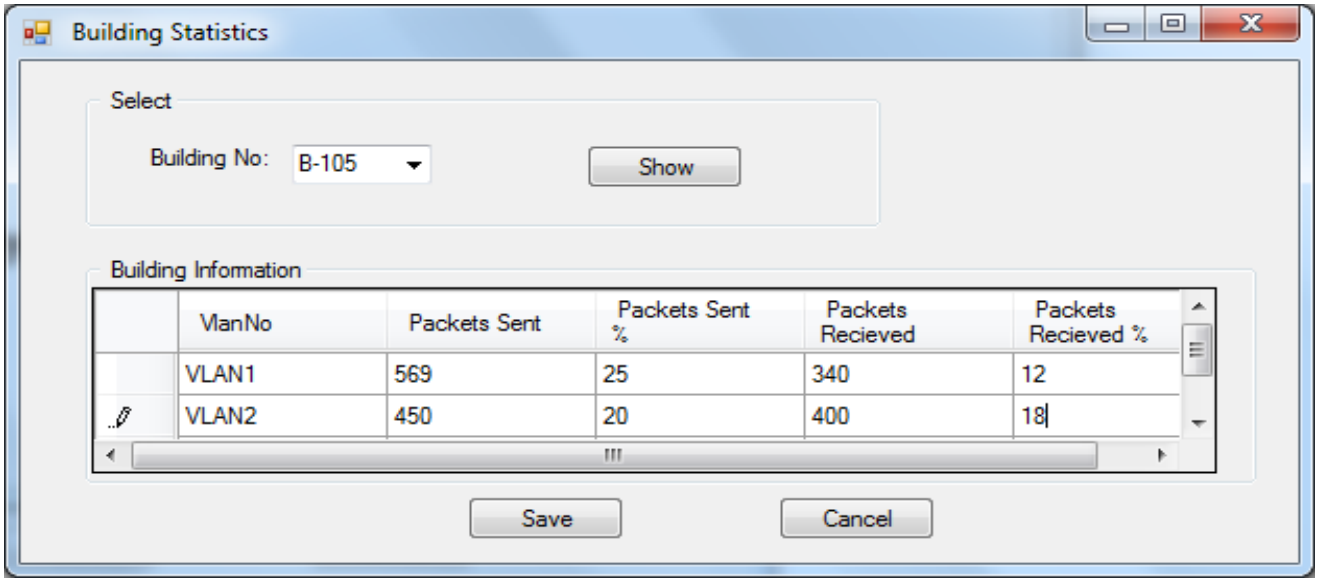


Figure 4.7 Building Statistics Window

Building statistics window displays the building information in terms of VLAN number, packet sent, packet sent percent, packets received and packets received percent. Moreover, the window allows the network administrator to specify the building number to show the specific information; and it allows saving the building information for report.

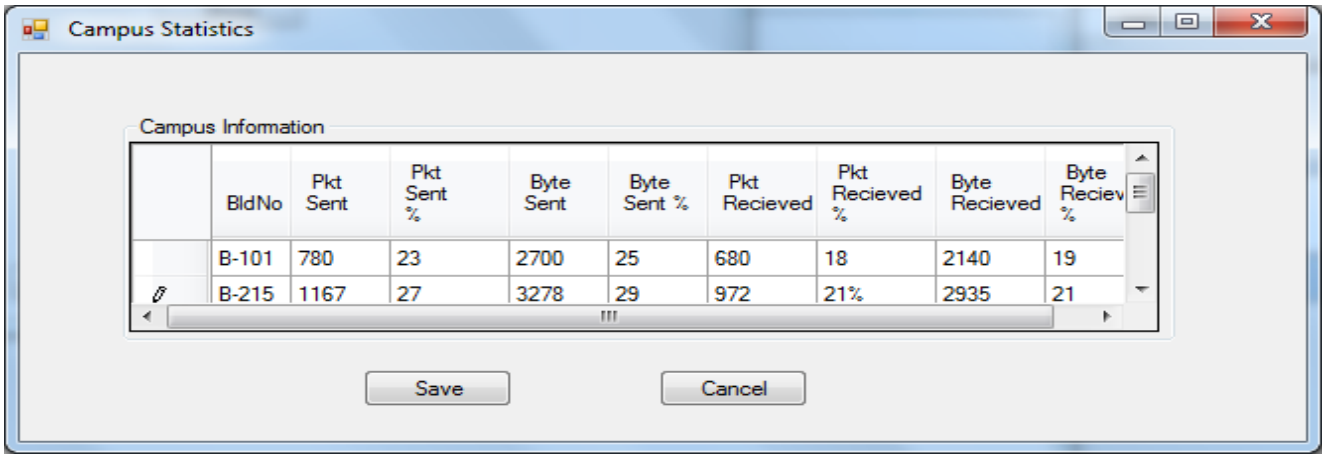


Figure 4.8: Campus Statistics Window

Campus statistics window displays the overall campus information in terms of building number, packet sent, packet sent percent, bytes sent, bytes sent percent, packets received and packets received percent, bytes received, bytes received percent. Moreover, the window allows the network administrator to save campus information for report.

Chapter Five

Conclusions and Recommendations

5.1 Conclusion

This thesis focuses on developing dynamic bandwidth management conceptual framework for intercampus network based on network traffic analysis. In this work we took Mekele University intercampus network as case for the network the network traffic analysis. The results of network traffic analysis and the proposed bandwidth management model are summarized as follows.

The MRTG is online traffic analysis tool that shows the ingoing and outgoing traffic in terms of daily, weekly, monthly and yearly in the intercampus network. Accordingly, from the MRTG daily graph for MU there is a steady rise in traffic by 14 GMT, when the break or lunch time is over and users are back in their offices and then a steady drop when users finally leave at about 18:00 GMT. Some academic staff and students typically work till midnight. Network usage is at its lowest from 0:00 GMT midnights local time till 8:00 GMT at the dawn local time and starts its cycle from 8:00 GMT once again. The MRTG daily graph for MU shows that there is maximum ingoing traffic of 29.2 Mbps which is recorded during 16 GMT and maximum outgoing traffic of 9.281 Mbps which is recorded by 18GMT. The MRTG weekly graph for MU shows that the weekly traffic load of the router link to Endayesus Campus measured at 16 September 2010 reaches its peak by Wednesday. The MRTG weekly graph for MU shows that there is maximum ingoing traffic of 22.0 Mbps which is recorded on Wednesday and maximum outgoing traffic of 5.1113 Mbps which is recorded on Wednesday. Moreover, MRTG shows the monthly ingoing and outgoing traffic load of MU campus network. Accordingly, MRTG monthly graph for MU shows that there is maximum ingoing traffic of 25.4 Mbps which is recorded during week 34 and maximum outgoing traffic of 4.7732 Mbps which is recorded by the end of week 34.

MRTG also shows the network traffic load the firewall. Accordingly we collected the most useful log entries from the firewall access log and we found, in the main log file, “Built”, “Deny”, and “Teardown” to show the accessed entries, denied entries and terminated calls respectively. These entries are especially useful for seeing port scans, host sweeps, and general probing.

MRTG generates the firewall daily outside and inside interface traffic. Accordingly, the daily MRTG graph for MU firewall outside interface clearly depicts that the daily maximum input traffic amounts to 10.4 MB/S and the daily maximum out traffic amounts to 738.8 KB/S. MRTG generate balanced traffic in the firewall inside interface . It depicts the daily maximum input traffic which amounts to 739.8KB/S and daily maximum out traffic which amounts to 10.4 MB/S which are the vice versa of the outside interface of the firewall.

MRTG shows the weekly traffic load of the MU firewall measured at 16 September 2010 reaches its peak by Wednesday; Thursday of the week with weekly average input traffic of 20.15MB/s and weekly average output traffic of 104.5kB/s. MRTG generate balanced traffic in the firewall inside interface. The MRTG clearly depicts the weekly average input traffic which amounts to 105.4 KB/S and weekly average output traffic which amounts to 1.9962 MB/S which are the vice versa of the outside interface of the firewall.

Ntop implemented in MU intercampus network to generate traffic statistics for IP protocols. The traffic statistics report general information about the observed traffic, historical view of the traffic statistics and traffic statistics regarding the individual protocols. Accordingly, the data collected by ntop shows that TCP and UDP are the highest bandwidth consuming protocols currently present in the MU network. Together they account for 92.7% of the network usage. Moreover, The data collected by ntop shows that SNMP and HTTP are the highest bandwidth consuming TCP flows currently present in the MU network. Together they account for 980.8 bytes/sec of the maximum network usage.

Generally, the ntop collects data about each and every specific protocol in the network. In this case it collects data about DNS with maximum of TCP flow 10.6 bytes/sec, FTP with maximum of TCP flow 1.6.bytes/sec, DHCP-BOOTP with maximum of TCP flow 20.2 bytes/sec, Proxy with maximum of TCP flowb1.6 bytes/sec, SSH with maximum of TCP flow 26.7 bytes/sec and NBIOS-IP with maximum of TCP flow 86.0 bytes/sec.

Weblog analysis consists in measuring the usage of relevant traffic activities. Weblog expert tracks web server log file, generating a series of statistics for each host, for operating system, for each browser, for each visitor and soon in the MU inter campus network as a whole. The needed information is collected by the Web server running specialized network server operating system by simply observing the traffic on the network. Web log expert can help the network administrators reveal important statistics about website usage: general statistics, activity of visitors, access statistics, activity statistics, Referrers, Website errors, paths through the site, visitors' browsers, click overlay and much more. According to the analysis the web log expert generates the total hits of 440,902, visitor hits of 428,702, spider hits of 12,200, Average Hits per Day of 73,483, Average Hits per Visitor of 35.80, Cached Requests of 24,475, and Failed Requests of 293,248, the Total Visitors of 11,974, Average Visitors per Day of 1,995, and Total Unique IPs of 2,313, the Total Page Views of 67,673, Average Page Views per Day of 11,278, Average Page Views per Visitor of 5.65. Moreover, it generates Total Bandwidth of 35.78 GB, Visitor Bandwidth of 35.57 GB, Spider Bandwidth of 213.32 MB, Average Bandwidth per Day of 5.96 GB, Average Bandwidth per Hit of 85.08 KB, and Average Bandwidth per Visitor of 3.04 MB.

Based on network traffic analysis conducted at different links it is proposed to develop dynamic bandwidth allocation conceptual framework with detailed description of the proposed model. Dynamic Bandwidth Allocator works by sorting traffic into classes based on the application type, source and destination addresses within specified period according to the specified bandwidth policy in the intercampus network. It then schedules the traffic according to the bandwidth defined for each class.

The dynamic bandwidth allocator works based on different promising rules which are feed as policy by network administrator which are exemplified as follows.

- Ensure that telnet traffic from workstation A, B, or C has a minimum allocated bandwidth of 5% over a leased line.
- Reserve 50% of a WAN link to traffic that originates from offices A, B, and C. Then within this 50%, reserve 10% for NFS traffic.
- Ensure that FTP transfer from machine B in office Y will not use more than 30% of the capacity of the line between offices Y and X.

- Dynamic Bandwidth Allocator for intercampus network some modules such as classifier, Timer, estimator, selector, Unique IP address counter, and Host dynamic bandwidth allocation.

Generally, the bandwidth allocator considers the available buildings in the campus, the number of implemented VLANs in a specific building, the number of users and the traffic flow type for developing its algorithm.

5.2 Recommendations

This thesis focuses on developing dynamic bandwidth allocator based on network traffic analysis in case of Mekele University intercampus network. Network traffic is a complex; hence, the thesis covers only network protocol distribution, network traffic analysis at different links and web server log file analysis. Based on these analyses we proposed the dynamic bandwidth manager which takes in to consideration the geographic locations, available number of VLANs and traffic load of each of the TCP/IP protocols/classes.

In this thesis it is not believed that full investigation is done on the campus. Hence future work is needed to investigate further analysis on the network traffic and generate usage patterns. Further work is required to build a better understanding of network traffic analysis in Ethiopian universities intercampus network. During development of dynamic bandwidth management we consider the factor like the geographic locations in the campus, the number of VLANS in a specific building, number of users and traffic types passing through the link.

Its real implementation needs a real network interface interaction and we propose Java platforms for its implementation.

There is also a great deal of work that could and should be done in web log analysis using web utilization mining and generating network traffic utilization patterns. During web log analysis the data set is analyzed using web log expert which cannot generate the pattern usage. The web log expert only shows us the website traffic usage.

References

[1]. Flickenger R. How To Accelerate Your Internet. A practical guide to Bandwidth Management and Optimization using Open Source Software. INASP/ICTP. October 2006

[2].http://en.wikipedia.org/wiki/Bandwidth_management. Bandwidth management Accessed on February 2011.

[3].Wiryaman, Santa. Bandwidth management algorithm. Barracuda Networks Inc (Campbell, CA, US). 2009

[4]. African Tertiary Institutions Connectivity Survey (ATICS): Full Report; Gakio; 2006.

[5]. Bandwidth management position paper. Aptivate, June 2007.

[6]. Optimizing Internet Bandwidth in Developing Country Higher Education

[7]. Mintesinot Behailu (PhD). ICT Infrastructure at Mekelle University, Ethiopia. Mekelle University. August 2005

[8]. S. Floyd and V. Jacobson, Link-sharing and resource management models for packet networks, IEEE Trans. Networking, vol. 3, 1995.

[9]. Chyouhwa Chen, Huei-Wen Ferng, Jun-Chuan Chen, Hao-Lun Chin, and David Shiung. A Class-Based Queueing Service for IEEE 802.11e Wireless Networks National Taiwan University of Science and Technology Taipei 106, Taiwan

[10].Dan Holme and Orin Thomas. Managing and Maintaining a Microsoft Windows Server 2003 Environment. Microsoft Press. 2006.

[11].Solaris™ Bandwidth Manager 1.5. Bandwidth Management for IP Networks, Inc.

[12]. I-Shyan Hwang, Zen-Der Shyu, Chun-Che Chang. A Novel Dynamic Wavelength and Bandwidth Allocation Scheme for WDM-EPON with Survivable Network Architectures.

[13]. http://www.cmu.edu/computing/doc/network/faq-bandwidth.html. FAQ Bandwidth Usage

[14].Laurent Bernaille, Renata Teixeira, Kav´e Salamatian, Early Application Identification, Universit ´e Pierre et Marie Curie LIP6, CNRS Paris, France

[15]. Abdulrahman Hijazi, Hajime Inoue, Ashraf Matrawy, P. C. van Oorschot, and Anil Somayaji, Towards Understanding Network Traffic through Whole Packet Analysis, Carleton University School of Computer Science TR-07-06 , 2007

[16].TCP/IP Networking Protocols

[17]. J. Postel and J. Reynolds. File Transfer Protocol. RFC 959 (Standard), October 1985.

[18]. H. Alvestrand. A Mission Statement for the IETF. RFC 3935 (Best Current Practice), Oct 2004.

[19]. J. Postel and J.K. Reynolds. Telnet Protocol Specification. RFC 854 (Standard), May 1983.

[20]. Douglas R. Mauro & Kevin J. Schmidt. (2001). Essential SNMP (1st ed.). O'Reilly & Associates.

[21]. RFC 3411 — An Architecture for Describing Simple Network Management Protocol (SNMP) Management Frameworks

[22]. Masinter, L., "Hyper Text Coffee Pot Control Protocol (HTCPCP/1.0)", RFC 2324, 1 April 1998.

[23]. J. Postel. Simple Mail Transfer Protocol. RFC 821 (Standard), August 1982.

[24]. T. Berners-Lee, R. Fielding, and H. Frystyk. Hypertext Transfer Protocol HTTP/1.0. RFC 1945 (Informational), May 1996.

[25]. R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach, and T. Berners-Lee. Hypertext Transfer Protocol – HTTP/1.1. RFC 2616 (Draft Standard), June

[26]. J. Case, M. Fedor, M. Schoffstall, and J. Davin . Simple Network Management Protocol (SNMP). RFC 1157, May 1990

[27]. _____ Transmission Control Protocol. RFC 793, September 1981.

[28]. www.wikipedia.com. User Datagram Protocol. Accessed on October 23, 2010.

[29]. J. Postel, User Datagram Protocol. RFC 768, August 1980

[30]. www.wikipedia.com. Secure Shell. Accessed on October 23, 2010.

[31]. T. Ylonen, C. Lonvick. The Secure Shell (SSH) Transport Layer Protocol (Standards Track). RFC4253, January 2006.

[32]. www.wikipedia.com. Telnet. Accessed on October 23, 2010.

[33]. www.wikipedia.com . NetBIOS. . Accessed on October 23, 2010.

[34].Wikipedia, the free encyclopedia http://en.wikipedia.org/wiki/Multi_Router_Traffic_Grapher , accessed on 07-Nov-10

[35]. Ling Xiao, John Gerth and Pat Hanrahan. Enhancing Visual Analysis of Network Traffic Using Knowledge Representation. Stanford University,

[36]. Remco van de Meent, Aiko Pras Michel Mandjes, Hans van den Berg, and Lambert Nieuwenhuis. Traffic Measurements for Link Dimensioning. University of Twente, 2003

[37].A scalable architecture for network traffic monitoring and analysis using free open source software.

[38]. T. Oetiker, "Monitoring your IT gear: The MRTG story," IEEE IT Profesionals, Vol. 3, No. 6, December 2001.

[39]. G. Robert Malan and Farnam Jahanian, "An extensible probe for network protocol performance measurement," in Proceedings SIGCOMM'98, September 1998.

[40]. L. Deri and S. Suin, "Effective traffic measurement using ntop," IEEE Communication Magazine, Vol. 38, No. 5, May 2000.

[41]. L. Deri, R. Carbone and S. Suin, "Monitoring networks using ntop," Proceedings of IEEE/IFIP International Sympo-sium on Integrated Network Management, May 2001.

[42]. L. Deri and S. Suin, "Practical network security experiences with ntop," Computer Networks, Vol. 34, 2000.

[43].A. Hussain, G. Bartlett, Y. Pryadkin, J. Heidemann, C. Papadopoulos and J. Bannister, "Experiences with a continuous network tracing infrastructure," in Proceedings of ACMSIGCOMM Workshop on Mining Network Data, August 2005.

[44]. Using MRTG to Monitor and Graph Traffic Loads

[45]. http://www.firewall.cx/firewall.php. Cisco Technical Knowledgebase. Accessed on Nov.28, 2010

[46]. Assigning Port Numbers. http://www.iana.org/assignments/port-numbers. Accessed on Oct 2010.

[47]. Mining Web Server Logs

[48]. http://www.weblogexpert.com/. Web Log Expert Help, Accessed on November 23, 2010.

[49]. http://goldmark.org/netrants/webstats/. Why web usage statistics are (worse than) meaningless. Accessed on November 24, 2010.

[50].Oliver A. McBryan. GENVL. WWWW: Tools for Taming the Web. First International Conference on the World Wide Web. CERN, Geneva (Switzerland), 1994.

[51]. Wilson, Tim. "Web Traffic Analysis Turns Management Data to Business Data." TechWeb, April 2, 1999

[52]._____. Assessing Web Site Usability from Server Log Files. · December 1999 ·

[53]. http://en.wikipedia.org/wiki/Server_log , Server log, accessed on 26-Oct-10

[54].obert Cooley, Bamshad Mobasher, and Jaideep Srivastava. Data Preparation for Mining World Wide Web Browsing Patterns. Department of Computer Science and Engineering , University of Minnesota, 4-192 EECS Bldg., 200 Union St. SE, Minneapolis, MN 55455, USA.

[55].Mathias G´ ery, Hatem Haddad. Evaluation of Web Usage Mining Approaches for User's Next Request Prediction. Information Technology Department , VTT Technical Research Centre of Finland, Espoo, Finland

[56].Jaideep Srivastava , Robert Cooley, Mukund Deshpande, Pang-Ning Tan, Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data

[57]. http://web.media.mit.edu/~lieber/Lieberary/Letizia/Letizia-Intro.html

[58]. http://en.wikipedia.org/wiki/Web_crawler

[59].Aniket Dash, Liju Robin. Web Usage Mining: An Implementation. National Institute of Technology, Rourkela Rourkela, Oriss, India, 2010

[60]. Allen, Cliff, Deborah Kania, and Beth Yaeckel. Internet World Guide to One-to-one Web Marketing. New York, NY: John Wiley and Sons, Inc., 1998.

[61] Stout, Rick. Web Site Stats: Tracking Hits and Analyzing Traffic. Berkeley, CA: Osborne McGraw-Hill, 1997.

Appendix I: Ntop Installation in Debian

Installing ntop

Download the latest ntop version

Type the following commands, enter:

$ sudo apt-get update

$ sudo apt-get install ntop

Set ntop admin user password

Type the following command to set password, enter:

/usr/sbin/ntop –A OR $ sudo /usr/sbin/ntop –A

Restart ntop service

Type the following command, enter:

/etc/init.d/ntop restart

Verify ntop is working, enter:

netstat -tulpn | grep :3000

viewing network usage stats

Type the url:

http://localhost:3000/ OR http://server-ip:3000/

Appendix II: MRTG Installation in Debian

Installing MRTG in Debian

#apt-get install mrtg snmpd

The installation will create an mrtg subdirectory where the Apache Web pages reside. On your Debian system the path of this subdirectory is:

/var/www/mrtg

Now you need to edit the mrtg configuration file to edit the some of the settings. File is located at /etc/mrtg.cfg you need to change the global settings as follows

Global Settings

cat /etc/cron.d/mrtg

0-55/5 * * * * root if [-x /usr/bin/mrtg] && [-r /etc/mrtg.cfg]; then env LANG=C /usr/bin/mrtg /etc/mrtg.cfg >> /var/log/mrtg/mrtg.log 2>&1; fi

Now we need to assign the snmp community name in snmp configration file etc/snmp/snmpd.conf

sec.name source community

com2sec paranoid default public

com2sec readonly default public

#com2sec readwrite default private

Now you need to restart the snmp service

#/etc/init.d/snmpd restrat

The configuration file creating using

#cfgmaker public@localhost > /etc/mrtg.cfg

Creating a configuration file for a device using

#cfgmaker public@192.168.0.1 >> /etc/mrtg.cfg

With the configuration file created correctly there's only one other thing you have to do and that's to use the indexmaker utility to create the summary home page. Since you have to re-run this command every time you make certain changes to the /etc/mrtg.cfg configuration file, Creating index file for the webserver using

#indexmaker /etc/mrtg.cfg > /var/www/mrtg/index.html

Now you need to reboot your system wait for five minutes or so and then take a look at your summary home page. If your Debian system's IP address is 172.16.0.20 then you'd type in the following in the address bar of a browser running on a system on the same network: http://172.16.0.20/mrtg/