

Traffic Measurement and Analysis of Wide Area Network (WAN) Link Usage: a case study of UiTM Perlis campus network

Abidah Hj Mat Taib*, Rafiza Ruslan

Faculty of Computer and Mathematical Sciences, UiTM Perlis, 02600 Arau, Perlis

Tel: +604-9882830 Email: abidah@perlis.uitm.edu.my

Abdul Hadi Kamel Abdullah*

Faculty of Computer and Mathematical Sciences, UiTM Johor, Km 12, Jalan Muar, 85000 Segamat, Johor

Tel: +607-9352160 Email: hadikamel@johor.uitm.edu.my

(This project is funded by Institute of Research, Development and Commercialisation, (IRDC), Universiti Teknologi MARA (600-IRDC/ST 5/3/912))

Abstract

Network users normally will start complaining when the response time for the network access is intolerable or their network access is intermittently interrupted due to unclear reasons. Knowing how the network bandwidth is being utilized by the users will give the administrator some ideas of possible reasons for the network access problems and how to alleviate the problems. Some users may just use the network for running non bandwidth intensive applications like email and telnet and some users may use the network for running bandwidth intensive application such as video streaming and downloading big files using ftp application. Many users may visit popular web sites that are not related to their work during office hours. The above information about bandwidth utilization on a network link can be acquired through measurement and analysis of the network traffic that pass through the link. This paper discuss real time passive measurement of the network traffic on the WAN link that connects UiTM Perlis campus network to the main campus of UiTM in Shah Alam. The analysis of the traffic helped us to know some aspects of the WAN link utilization. This includes the average bandwidth utilization, popular application layer protocols and popular websites among different user segments and also some anomalies found in the captured traces.

Keywords: *bandwidth, traffic analysis, WAN link usage*

1. Introduction

Nowadays it is common to use the Internet in our daily lives. Many private and public organizations offer their online Web services to their customers as extension to the existing “bricks and mortar” business. Students in universities and colleges consider the Internet as indispensable as additional source of information besides their text books. Users always expect from their Internet Service Provider to provide the guaranteed services, easy access and fast response time when they surf the Internet. However the users do not always completely get what they expect when they experience a slow response time and intermittent network downtimes. One of the vexing questions facing researchers interested in the World Wide Web is why users often experience long delays in document retrieval (Barford & Crovella, 1999).

This paper discuss a study to get a clearer picture of the bandwidth utilization of the WAN link that provide the Internet access to the users in UiTM Perlis. At the time of the study, there were around 1500 Local Area Network (LAN) end-user nodes in UiTM Perlis campus network. The campus network provides Internet access as well as the access to some selective Intranet applications in the UiTM Shah Alam to the administrative staff, lecturers and students. Among popular Internet applications and services used by all groups of users were WWW, e-mail and ftp. At the time of the study, users in UiTM Perlis depend on a single WAN link for their connections to the Intranet and to the Internet. The WAN link used the Frame Relay technology with a data rate of 2 Mbps connected between UiTM Perlis campus network and UiTM Shah Alam campus network. The same WAN link configuration

was also applied to all other UiTM branch campuses. All access to the Internet from branch campuses will be merged in the router of UiTM Shah Alam before going out through UiTM Shah Alam WAN link to the respective end computers on the Internet. As a result, the performance of Internet access in UiTM Perlis was not only determined by the UiTM Perlis users' access behaviour and the bandwidth available on the UiTM Perlis WAN link but also by the other users' access behaviour for other campuses because they were sharing the same link to the Internet Service Provider (ISP) in the UiTM Shah Alam network.

Our random observation showed that, users occasionally experience slowness in response time and intermittent access problems in their Web access for unclear reasons. We suggested the network administrator to know the actual causes of those problems first before deciding to increase the capacity of the WAN link to alleviate the problems. We anticipated the above problems might be due to:

- i) Congestion on the WAN link due to legitimate high bandwidth usage on the link by different group of users in UiTM Perlis.
- ii) Congestion on the WAN link due to anomalous traffics due to worms or security attacks.
- iii) Delay caused by forwarding rate of intermediate devices located outside UiTM Perlis network.

Thus, to provide better services, network operators must be aware of the characteristics of the traffic and the network behaviour. So, measurement and analysis of real traffic is important to gain knowledge about the characteristics of the traffic (Abrahamsson, 1999). Moreover, in depth information about network requirements such as the availability, performance and capacity of the LAN and WAN sides of the network can only be obtained through data gathering and analysis of network performance trend (Nametka, 1999). As such, the result obtained from the study of the network usage and characteristics may help the network designer in deploying or expanding a similar network (Stern, 2001). Additionally, understanding the nature of Internet and web traffic flows is vital and is a main criterion in designing an optimal network configuration and policy on how the network should be used (Ziviani, 2006).

This paper presents some results regarding the bandwidth utilization of the WAN link connecting the UiTM Perlis campus network and UiTM Shah Alam campus network. Our study utilized a passive measurement of the traffic that passed through the port of core switch that is connected to the router that has a connection to the UiTM Perlis main WAN link. The measurement of traffic was done by performing port mirroring on the core network switch to capture in-bound and out-bound packets that pass through between the port of core switch and the router that is connected to the WAN link. The port mirroring was done by enabling the roving analysis feature on the core network switch. This paper describes the detailed of passive measurement of the traffic used for collecting the measurement of the WAN link as well as some findings from our study based on the captured access traces on the WAN link.

The remainder of this paper is organized as follows: Section 2 provides background of traffic measurement and other related works. Section 3 describes measurement scenario and Section 4 presents our approach towards WAN link analysis. Section 5 provides results and some discussions based on the data analysis. In the last section, we present a summary of our findings and conclusions as well as a brief discussion on plans for future work.

2. Background and Related Work

Several studies have been carried out in the past ten years were trying to have a better understanding of the networking issues as well as the characteristics of the network traffic. One of them was discussed by Epsilon, Ke & Williamson (1999) which presented low-level measurement of Internet traffic flowing to and from an Internet Service Provider over an ATM backbone network. Traffic analysis at the cell-level, frame-level and TCP connection-level showed that much of the traffic in the traces is Web-related and at the individual host-level. They noticed a significant traffic structure at fine-grain time scales due to TCP/IP and HTTP protocol behaviours. From the measurement, they also confirmed that internet users generate very demanding traffic for telecommunications networks to handle.

2.1 Traffic Measurements

Traffic measurement methods include *active* and *passive* approaches. Active measurement involves injection of traffic into the network in order to probe certain network devices or to measure network properties such as round-trip-times (RTT). The passive measurement is pure observation of network traffic that passes through a device. Special traffic acquisition hardware and software is used to collect traffic directly on the physical links or on network interfaces without any additional traffic injection. An example of passive measurement is by mirroring ports on the network device. One or more ports will act as monitor ports and another port will act as an analyzer port which mirrors the monitor ports. Measurement data is collected on the analyzer port.

Our work is similar to Coucoubetis and Siris (1999) which presented a platform for collecting and analyzing measurements of real network traffic on the link connecting between two university campuses. They had devised a low cost high performance PC-based monitor to collect the measurements of the high speed link. In contrast our work featured a roving analysis as demonstrated by Tamosoft (2005) that used passive monitoring tool to capture incoming and outgoing traffic at the edge router.

This work is also motivated by Murnan (2001) who discussed about strategies used to ensure acceptable WAN network bandwidth availability for all users in their organization. Equipped with a traffic-shaping tool and Simple Network Management Protocol (SNMP)-based bandwidth usage monitoring tool (MRTG), they succeeded in preventing a small group of top-talkers from monopolizing the large portion of the WAN link bandwidth. Fair distribution of bandwidth among users was achieved through prioritization of WAN link bandwidth usage based on their source addresses and time when the access was made. They also used a packet sniffer to pin-point the exact causes of slow response time in web access at certain time duration. After identifying which group of users had consumed a large portion of network bandwidth, the administrator will be able to adjust accordingly using the above prioritization scheme implementation in the traffic-shaping tool. We also applied a SNMP-based MRTG tool to observe incoming and outgoing traffic by producing daily and weekly graphs.

The collected data can be stored in most popular trace format, the *PCAP format*, the format of the BSD Packet Filter and TCPdump. For instance, Kotz and Essien (2002) had traced network activity in the campus by using Tcpcap to capture all the packet headers on a selection of the APs around campus-wide wireless network. Their research was similar to our study but the only difference is they focused on the entire Dartmouth campus-wide wireless network which looking for other factors like access points and mobile activities while ours concentrate on the main gateway usage of the wired network.

2.2 Traffic Classification

After collecting data, the services have to be classified. The classification has to use another tool or steps in order to allow observation on different services or protocols in the traffic. A study by Barford and Plonka (2001) used the FlowScan open source software to gather and analyze network flow data. FlowScan takes Netflow feeds from Cisco or other Lightweight Flow Accounting Protocol (LFAP) enabled routers. Netflow data includes source and destination AS/IP/port pairs, packet and byte counts, flow start and end times and protocol information. FlowScan maintains a set of counters based upon the attributes of each flow reported by a router. The attributes include IP protocol (ICMP, TCP, UDP), well-known service (such as FTP or HTTP) based on source/destination port, CIDR block of local IP address and source/destination Autonomous System (AS) number. Their analysis approach employed a variety of tools including simple statistics, time series analysis and wavelet analysis to characterize anomaly features. Alternatively, our analysis approach involved tools like TCP reduce (Paxson, 1995) as well as iNetmon Portable Plus (iNetmon) for data categorization.

In the next section, we describe our measurement setup and how we performed passive measurement of network traffic that pass through the WAN link using the roving analysis feature on the core network switch.

3. Measurement Scenario

The data collection serves two important parts of the concern in this study: the packet capturing through the WAN port on the edge router and the bandwidth usage of the incoming and outgoing live network traffics on the 2 Mbps Frame Relay WAN link.

3.1 Measurement setup

The data have to be captured on the edge router that acts as the main gateway of UiTM Arau network. The LAN port of the edge router was connected to a port of a 3Com 4007 switch and the WAN port of the router was connected to the CSU/DSU device that connect to Frame Relay network. The Simple Network Management Protocol (SNMP) agent was activated on the core network switch to collect data regarding network device status and traffic statistics.

3.1.1 Bandwidth usage measurement on the WAN link

We activated the SNMP agent on the core switch. As for the bandwidth usage data, we utilized a utility called *snmpget* which is part of the NET-SNMP, a SNMP applications suite to poll the core switch monitored port for every 1 minute for data regarding incoming and outgoing number of bits that pass through the WAN link.

3.1.2 Passive measurement of traffic on WAN link

To acquire data transmitting via the WAN link, data collection should be done directly by capturing the packets that pass through the WAN port on the edge router. However, due to some access constraints to the edge router, the data collection cannot be done directly on the WAN port that was directly connected to the 2 Mbps Frame Relay WAN link. We opted for collecting data regarding the incoming and outgoing traffic on the switch port that was directly connected to sole LAN port of the edge router using the roving analysis feature on the core switch as shown in Figure 1.

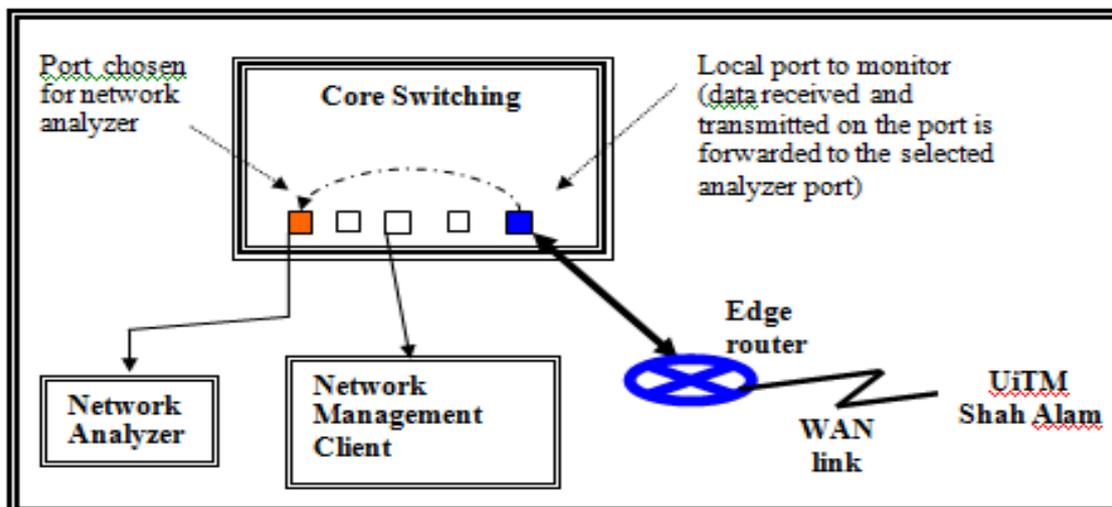


Figure 1 : Connection of devices for packet capturing

The feature allows us to mirror data in real-time basis from a switch port that is designated as monitored port to another switch port that was designated as analyzer port. The analyzer port had been connected to a computer which was installed with a packet capturing utility named *Ethereal 0.10.12*. Another good feature of the roving analysis feature is once a port is configured as an analyzer port, it is disabled from receiving or transmitting any other data except from the port that is designated as monitored port.

3.2 Trace Description

The data collection process was done started from July 26th, 2005, until August 1st, 2005 and from August 17th till August 30th, 2005. The gap in between was due to malfunctioned router as a result of heavy storm at UiTM Perlis. Packet capturing process using the *Ethereal* utility

had been running 24 hours a day throughout the above period. All the data were written to disk for later offline analysis.

3.3 Traffic Categorization

Categorizing and analysing the incoming and outgoing data packets from a WAN network tell us the different type of packets traversing through the WAN links of an organization. As data centres are consolidated to a central location, the WAN links almost become an extension of LAN and as a network is only as fast as its slowest link, it becomes essential to optimize the performance of a WAN link. So, understanding the current WAN link utilization may help us figure out the optimum bandwidth required to satisfy the user needs. Regardless of the capacity of bandwidth on the WAN link, there are always some non-critical applications that can choke the WAN link. If we can come up with a policy to filter these non-critical applications from accessing through the WAN link, the optimum usage of bandwidth can be achieved and possibly will make every network user happy.

There are some criteria that can be used to categorize the network traffic. They could be based on network application used, source TCP port address, protocol used, network layer address, data link address and destination TCP port addresses. In our study, we categorized the network traffic based on parameters like applications, protocol, subnet address and specific TCP destination port addresses.

The data analysis had been done using two important tools. We used the tool named TCP-Reduce (Paxson,1995) to analyse the data collected during passive WAN traffic measurement using Ethereal utility which is in the standard TCPdump format for data simplification and categorization. We also used a network monitoring tool called iNetmon Portable Plus for categorizing the network traffic. The next section describes how interpretations using tcp-reduce and tcp-summary utilities which are part of TCP-Reduce tool were done on the data in TCPdump format.

4. Our approach to WAN link usage analysis

In the beginning of our analysis, we did not have access to any network monitoring tool. Since the raw data we obtained was a pcap data (TCPdump) it has to undergo some form of data interpretation to help us with the analysis. To categorize the data we used TCP-reduce (Paxson, 1995) which produced summary of each connection with a single ASCII line. TCPdump data line contains 8 or 9 columns:

1. **timestamp** when the connection began (first SYN packet)
2. **duration** of the trace in seconds, or a ? if the trace did not show the connection terminating (no FIN or RST packets seen)
3. **protocol** used by the connection. This in general is derived from the port number used by the responder to the initial SYN packet. An exception is made for an initial SYN packet sent from TCP port 20, which corresponds to the well-known port for *ftp-data*. Unidentified ports are reported as *other-XXXX* if non-privileged (> 1024) and *priv-XXXX* if privileged. If the unidentified traffic is coincident with an *ftp* connection between the two hosts, then it is reported as *ftpdata-XXXX* instead.
4. **bytes sent by originator** of the connection, or ? if not available (due to connection not terminating, or terminating with RST).
5. **bytes sent by responder** to the connection, or ? if not available.
6. **local host** - the (possibly renumbered) *local* host that participated in the connection. See below for discussion of *local* and *remote* hosts.
7. **remote host** - the (possibly renumbered) *remote* host that participated in the connection.
8. **state** that the connection ended in. This can be one of the following:
 - **SF** normal SYN/FIN completion
 - **REJ** connection rejected - initial SYN elicited a RST in reply
 - **S0** state 0: initial SYN seen but no reply
 - **S1** state 1: connection established (SYN's exchanged), nothing further seen

The output of *tcp-summary* consists of six columns (plus a header to remind us of what is in each column):

1. **TCP protocol**
2. **Number of connections** made or attempted using that protocol
3. **Number of kilobytes** of user-level transferred in both directions
4. **Successful** - percentage of connections that terminated in state **SF**, indicating normal SYN/FIN completion
5. **Local** - percentage of connections initiated by *local* hosts
6. **Neighbor** - percentage of connections in which the *remote* site was from a *neighbor* network.

The output is sorted on the third column (kilobytes). *priv-XXXX*, *other-XXXX*, and *ftpdata-XXXX* connections are lumped together as three collective protocols. Protocols for which fewer than **insig_bytes** were transferred, or **insig_conn** connections were made, will not be reported. These variables are defined at the beginning of the script, and default to 500 KB and 100 connections, respectively. An excerpt of our summary data is shown in Table 1.

Table 1: Example of summary data

overall number of hosts = 1787					
26/7/2005					
no. of local stations					
539					
proto	# conn	KBytes	% SF	% loc	% ngh
www	551626	5453409	62	100	19
other-4662	30141	539755	71	100	0
other-xxx	222626	513143	NA	NA	NA
other-15871	27964	141444	96	100	100
smtp	3755	49837	18	86	18
other-1352	762	44148	91	29	30
priv-xxx	2026	41885	NA	NA	NA
ftpdata-XXXX	103	29370	37	100	2
priv-443	3796	21772	49	100	6
other-4661	1225	21443	69	100	0
other-5662	848	16776	83	100	0
other-5773	372	7225	74	100	0
other-3128	1260	7003	88	100	0
ftp-data	519	5941	87	11	0
other-1433	101	3630	99	97	97
other-8398	236	2584	80	100	0
other-8080	437	2026	36	100	0
other-5050	346	1534	23	100	0
other-5100	271	1487	17	100	0
other-6350	481	943	45	100	0
nntp	88	485	17	100	0
pop-3	171	238	33	100	5
telnet	72	145	11	100	6
printer	13	81	92	8	8
ftp	214	71	25	100	1
gopher	35	16	94	100	3
systat	14	7	93	100	7
auth	14	5	64	100	7
discard	5	2	60	100	20
whois	1	0	0	100	100
time	42	0	0	100	2
shell	1	0	0	100	100
nntp	1	0	0	100	100
netbios-ssn	106	0	0	8	1
netbios-ns	1	0	0	100	100
netbios-dgm	1	0	0	100	100
login	1	0	0	100	100
isode-dua	16	0	0	100	0
finger	2	0	0	100	50
exec	1	0	0	100	100
echo	1	0	0	100	100
domain	1217	0	0	100	9
daytime	1	0	0	100	100
aol	1	0	0	100	100
x11	26	0	8	100	8

Beside summary data provide by TCP-reduce tools, we also did offline analysis by observing the packets using packet analyzer Ethereal or its newer version Wireshark in order to have some ideas of the traffic that actually travels between UiTM Perlis campus network and UiTM Shah Alam campus network. From the summary data and the packet analyzer we attempted to study the how the campus network was being utilized. Later, with the help of network monitoring tool such as iNetmon Portable Plus, we were able to categorize in detail the network protocols used as well as figure out the existence of some worms in the network.

Our bandwidth analysis covers bandwidth usage levels per IP address, host or subnet; top few applications that are generating the bulk of the traffic; bandwidth wasted due to retransmissions, dropped packets, aborted TCP connections by the server and applications/users/servers that get the most traffic and send the most traffic. The next section discussed our findings based on questions raised in our problem statement.

5. Results of Traffic Analysis

Our traffic analysis using tools mentioned in the previous section provides several aspects of the WAN link bandwidth utilization and the present of some anomalies in the network traffic that passed through the UiTM Perlis WAN link.

5.1 Bandwidth usage on the WAN link

The capturing of bandwidth usage on the WAN link is done by capturing real traffic comprises of incoming and outgoing number of bits that pass through the core switch port at the main gateway. Based on the traces collected, we observe every 1 minute for about 14 days from 17th to 30th August 2005. Figure 2 and Figure 3 show the trend of average incoming and outgoing bandwidth usage during day-time and night-time periods respectively. On working days the average bandwidth usage on the WAN link for the incoming traffics was almost fully utilized to the maximum of 2 Mbps bandwidth while the outgoing traffics utilized less than 1 Mbps of the maximum 2 Mbps. On the other hand, on the weekends there is only a few activities occurring during the weekends with incoming and outgoing traffics utilized less than 500Kbps. As for night-time from 6.00 pm to 4.00 am, the utilization is less than 50 percent of the total 2Mbps bandwidth. We consider the low utilization during night-time a waste because the Frame Relay WAN link is an always-on link that is paid on monthly, quarterly or annually basis.

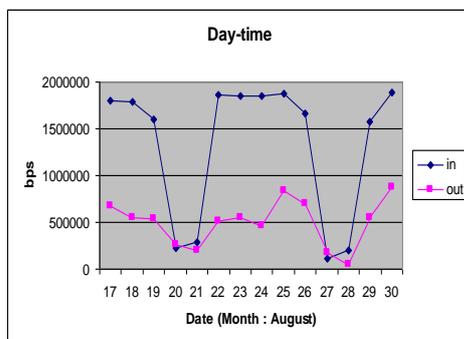


Figure 2 : Data captured from 8am-6pm

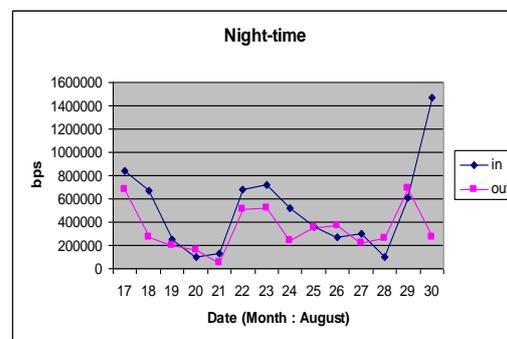


Figure 3 : Data captured from 6pm-4am

5.2 Popular application layer protocols

Among analysis that have been done is the list of protocols used by the user. The analysis focuses on common protocols like http, smtp, ftp and telnet. The daily trend of application layer protocol usage throughout the data collection period is more or less similar. Figure 4(a) shows the quantity of data transferred in Kbytes for the popular protocols. The most dominant protocol is http which recorded almost 56 percent of the number total access with 63 percent of data transferred. Figure 4(b) shows a snapshot of the overall application layer protocols. This is in line with the study conducted by Zhang et. al. (2005).

Noticed in that graph, the Web access is extremely high compared to other protocols but when comes to week end (30 and 31 July are Friday and Saturday respectively), the amount of Web traffic is dramatically decreased. Figure 2 and 3 also indicate the similar pattern in which less incoming and outgoing data activity during the weekends.

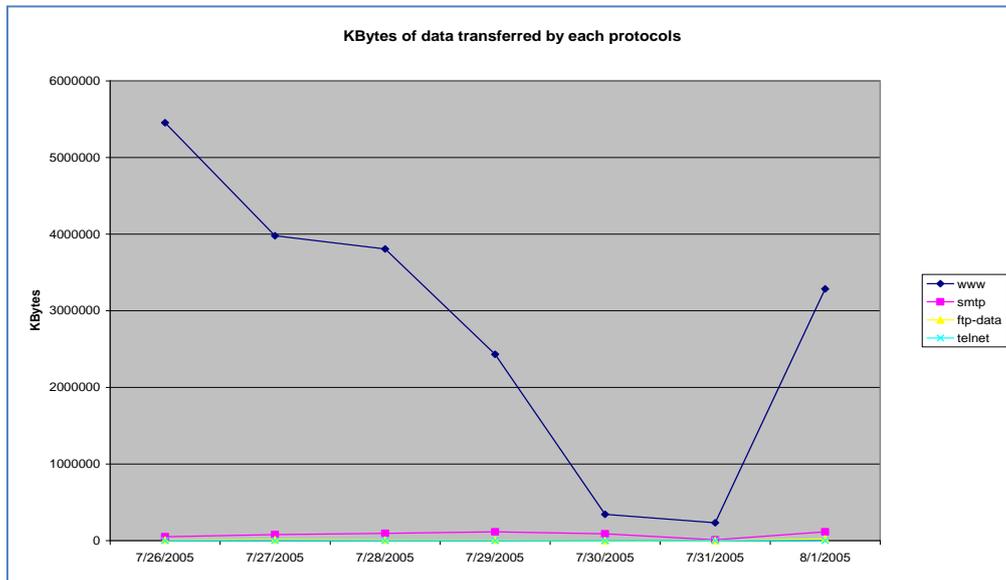


Figure 4(a) : Kbytes of data transferred by each protocols

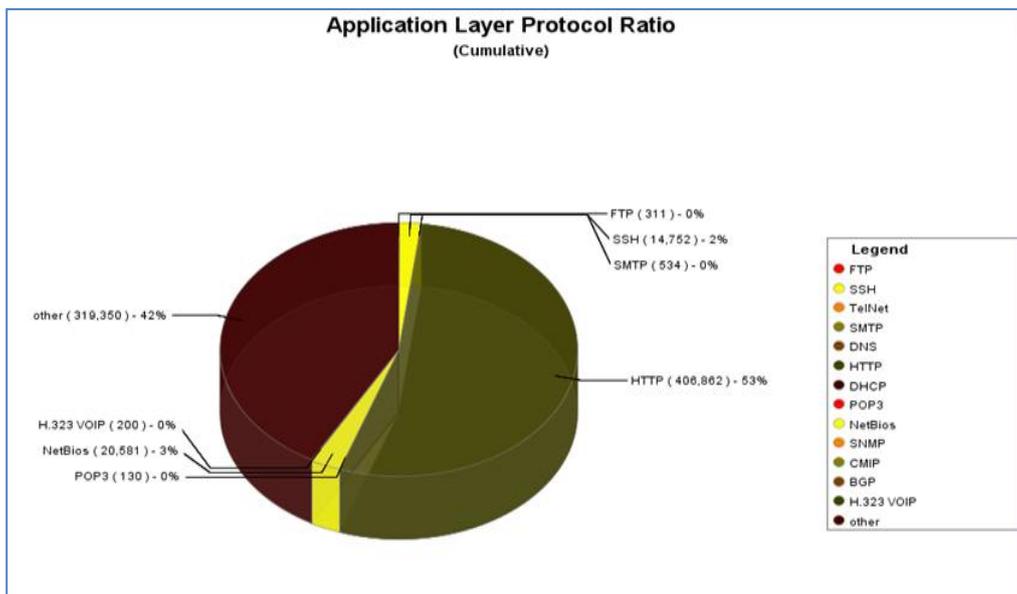


Figure 4(b) : A snapshot of the overall application layer protocol

Another interesting finding is regarding the significant number of access made by the peer-to-peer (P2P) application and also the responses to the users due to blocked Web sites. As for the peer-to-peer application usage, we found that e-Donkey P2P file-sharing programs (eDonkey) that uses TCP port 4662 generated almost 3.4 percent of the total number of access for files transfer to its client located in the UiTM Perlis network that accounted for 9.4 percent of the total data transferred across the WAN link. As for the number of responses from the Websense Web filtering utility which uses TCP port 1581 to block users Web access that is against the UiTM Web access policy, it took almost 2.7 percent of the total access.

5.3 Popular website among users

We also analyzed the captured data for popular web sites accessed by different segments of users. Based on network subnets, we categorized the users into three groups which are lecturers, administration staff and students. We produced site ranking list which lists out frequent sites visited by different group of user. The top five of the list is shown in Table 2. The top most visited site by every group is UiTM home page. This explains that UiTM homepage which integrates student portal and staff portal could be frequently visited for latest news and updates regarding academic and student affair as well as important announcements for staff.

Table 2 : Top-five popular website among users

STUDENTS	LECTURERS	ADMINISTRATION STAFF
www.uitm.edu.my	ns1.imagri.net.au	amadeus.uitm.edu.my
66.28.200.34	64.233.189.104	www.uitm.edu.my
oaktree.solidoak.com	www.uitm.edu.my	bs1.vip.scd.yahoo.com
66.94.234.72	203.106.50.10	online.isis.uitm.edu.my
www.friendster.com	202.186.153.6	johor.mschoosing.com

Some of the Web site did not have domain name and the access to the Web site is using the IP address of the Web server where we have to feed the IP address of the Web server as the URL on the Web browser for the purpose of the tracing the Web sites. Surprisingly, the second top site visited by students is Naruto homepage where public can register for membership and download Naruto multimedia product. The third top site is Cybersitter Web sites, where user can download internet filtering products, followed with Yahoo and Friendster Web sites. Yahoo could be visited for mailing purpose and everyday surfing while friendster could be a social networking site where students can find new friends with common interests as the friendster site provide various services which attract youth. The tenth top site is google search. If assumption were to make regarding the students behaviour in surfing the Web, they tend to access sites that is related to entertainment and social networking rather than for academic related sites.

Figure 5 is a snapshot using iNetmon that shows the website surfed by the students in one of the computer laboratory. From the figure we can see some websites such as www.jakionline.net and www.jiwang.org were being visited by students in the UiTM Perlis computer laboratory. Besides that the trace of e-Donkey, a P2P file sharing site also could be seen in the student laboratory. We also noticed that services such as Microsoft DirectX Gaming, Yahoo Messenger-messages and BitTorrent being utilized during the office hours which consumed bandwidth that contribute delay in accessing. Figure 6 shows an excerpt of the traffic that contains e-Donkey and Bittorent.

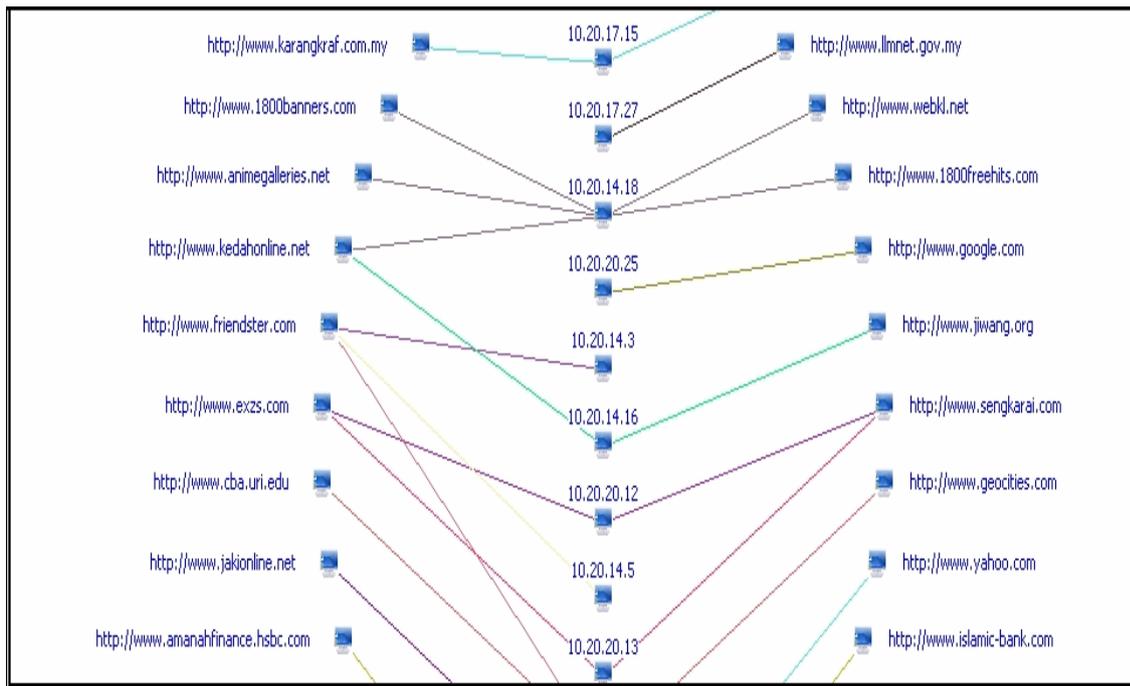


Figure 5: A snapshot of website surfed by students

As for administration staff, by ranking site visited by them, we identified the top 4 sites are UiTM sites which including the student information system ISIS service. This followed with yahoo, radio era and other miscellaneous sites. As for lecturers, other than UiTM site, sites that offer services like web site design, yahoo site, NETEASE, studiografic, audio and data music CD burner, FORTINET, and on-line newspapers like NSTP and Utusan are among the top twenty sites visited. These accesses have contributed to the usage of 2Mbps WAN link. This is a quite interesting finding as the above popular sites among lecturers are not highly related to teaching, learning or research that was assumed to be popular among academic staff.

5.4 Incomplete TCP Connections

Throughout the packet capturing period, a total of 5,030,285 TCP connections had been detected involving 30,974 Mbytes of data transfers. The most unexpected finding from our study is that the overall percentage of the successful network access is very low at only 39 percent of the total TCP connections with the percentage for successful Web access is only at 57 percent. We hypothesize that the low percentage of successful connection for Web access is caused by extremely slow response time of Web access due to network congestion during peak-hours which resulted into high number of aborted and failed connections. An in-depth study should be conducted to clarify and verify the actual reasons for such low percentage.

5.5 Anomaly in the captured packets

We noticed an anomaly in the data captured that concerned the number of active nodes found in the respective subnets. The number of nodes is counted based on the number of unique source IP addresses that originated from a subnet. For instance the number of computer in Computer Lab 2 is only 30 but the number of unique source addresses originated from the lab that pass through the edge router is 93. During the study period UiTM Perlis has not yet implemented dynamic IP address configuration for the computers in the student laboratory using a Dynamic Hosts Control Protocol (DHCP) server. Detailed finding is shown in Table 3.

We suspect that this may be caused by malicious programs such as worm or spyware that spoofed the source IP addresses of the infected computers as their propagation strategy to avoid detection. Besides, this scenario may rise due to various possible reasons such as malfunctioning network devices, network overload, malicious denial of service attacks and

network intrusion that disrupt the normal delivery of network services. This may require an in-depth study to learn more on the pattern characteristic of the network anomalies and possible network intrusions (Thottan & Ji, 2003).

Table 3 : Anomaly in the number of nodes

Subnet	Number of Nodes captured	Actual Nodes
A	93	30
B	50	30
C	80	30
D	77	30
E	83	30
F	62	30
G	85	30
H	83	30
I	74	30
J	89	30

After undertaking some further analysis on the captured data, we detected a number of worm definitions occurrence almost every day. Some of the source IP addresses are detected from the library and one of the lecturer's nodes. Most of the worms are generated from P2P file-sharing activities and through e-mails. Table 4 shows example of worms detected in the traffic.

Table 4: Example of worms detected in the traffic

Source IP	Destination IP	Attack Port	Severity	Worm Name
10.20.22.212	66.94.234.72	80	Warning	Email-Worm.Win32.Sober.e
10.0.28.12	10.20.1.10	25	Warning	Email-Worm.Win32.Sober.e
10.0.28.12	10.20.1.10	25	Alert	P2P-Worm.Win32.Bare.a
10.0.28.12	10.20.1.10	25	Warning	Email-Worm.Win32.Sober.e
10.0.28.12	10.20.1.10	25	Warning	Email-Worm.Win32.Sober.e
10.0.28.12	10.20.1.10	25	Warning	Email-Worm.Win32.Sober.e
10.0.28.12	10.20.1.10	25	Warning	Email-Worm.Win32.Sober.e
10.120.91.1	10.20.29.3	139	Alert	P2P-Worm.Win32.Bare.a
10.120.91.1	10.20.29.3	139	Alert	P2P-Worm.Win32.Bare.a
10.20.22.228	64.233.189.104	80	Alert	P2P-Worm.Win32.Bare.a
10.120.91.1	10.20.29.3	139	Alert	P2P-Worm.Win32.Bare.a
10.120.91.1	10.20.29.3	139	Alert	P2P-Worm.Win32.Bare.a
10.20.29.7	128.252.19.20	3128	Alert	P2P-Worm.Win32.Bare.a
10.120.91.1	10.20.29.3	139	Alert	P2P-Worm.Win32.Bare.a
10.20.29.7	128.252.19.20	3128	Alert	P2P-Worm.Win32.Bare.a
10.120.91.1	10.20.29.3	139	Alert	P2P-Worm.Win32.Bare.a

5.6 Other considerations

Another thing that worrying us was the existing of ICMPv6 messages as shown by the router solicitation and also router advertisement in Figure 6. Although the network administrator did not make any steps towards IPv6 transition, he might not realized that IPv6 packet is already travels in the network. This is because many operating system manufacturers have helped aid the migration to IPv6 by including it in their OSes by default. These dual stack operating systems perform DNS lookups and if they get back both an IPv4 and IPv6 address in their

query they will actually prefer communicate using IPv6 if they have the option. Most users do not aware of this capability is turned on by default and this preference exists. Additionally, operating systems like Microsoft XP, Vista and Windows 7 will help the user establish IPv6 Internet access by creating tunnels where the IPv6 traffic will be encapsulated in IPv4 packets to be pre-determined points on the Internet. If the administrator who takes care of the campus network was not aware of this scenario, users may have actually communicating with the Internet using IPv6 without realizing it. The most alarming is probably the administrator has not considered the security implications of this or taken any action to secure the communication in IPv6 in the same ways he secures IPv4 traffic. Consequently, all users may be exposed to IPv6 attack. Since those OSES have the IPv6 protocol ready to be used, the ICMPv6 message router advertisement might be sent by a layer 2 adjacent attacker pretending to be an IPv6 router. Thus, an attacker can enable IPv6 on those OSES, and if the computer is not protected by a firewall against attack over IPv6, and if the computer is vulnerable, than the attack will succeed.

No.	Time	Source	Destination	Protocol	Info
293582	476.085948	86.129.103.142	10.20.19.55	ICMP	oms > kjtsiteserver [ACK] Seq=891 Ack=271907 Win=83840 Len=0
293583	476.084347	10.20.21.61	10.0.0.1	TCP	workflow > ariganetfs [SYN] Seq=0 Win=16384 Len=0 MSS=1460
293584	476.086229	10.20.24.250	66.163.178.21	TCP	cas > http [SYN] Seq=0 Win=8192 Len=0 MSS=1460
293585	476.103911	202.157.188.189	10.20.23.103	TCP	[TCP segment of a reassembled PDU]
293586	476.109940	202.157.188.189	10.20.23.103	TCP	[TCP segment of a reassembled PDU]
293587	476.110879	fe80::5445:5245:444f	ff02::2	ICMPv6	Router solicitation
293588	476.111792	10.20.23.103	202.157.188.189	TCP	vpad > http [ACK] Seq=1439 Ack=57298 Win=17184 Len=0
293589	476.115994	202.157.188.189	10.20.23.103	TCP	[TCP segment of a reassembled PDU]
293590	476.119104	10.0.10.46	10.20.1.12	TCP	crprotocol > 10319 [PSH, ACK] Seq=784551 Ack=1 Win=64641 Len=710
293591	476.121245	10.1.4.235	10.20.27.51	UDP	Source port: 26776 Destination port: 6970
293592	476.127441	202.157.188.189	10.20.23.103	TCP	[TCP segment of a reassembled PDU]
293709	476.353716	10.20.1.12	10.0.10.46	TCP	4539 > 10319 [PSH, ACK] Seq=5676729 Ack=714103 Win=64905 Len=328
293710	476.354845	10.20.23.103	202.157.188.189	TCP	vpad > http [ACK] Seq=1439 Ack=79378 Win=17184 Len=0
293711	476.356112	10.20.17.26	64.70.34.135	TCP	codima-rtsp > http [SYN] Seq=0 Win=8192 Len=0 MSS=1460
293712	476.358690	fe80::8000:f227:bffb	fe80::5445:5245:444f	ICMPv6	Router advertisement
293713	476.363155	10.1.4.235	10.20.27.51	UDP	Source port: 26776 Destination port: 6970
293714	476.365576	10.1.4.235	10.20.27.51	UDP	Source port: 26776 Destination port: 6970
293715	476.367736	10.1.4.235	10.20.27.51	UDP	Source port: 26776 Destination port: 6970
293716	476.368727	10.20.20.30	10.233.12.141	TCP	nsft-gc-ssl > urdisnet [SYN] Seq=0 Win=16384 Len=0 MSS=1460
293717	476.372674	10.20.19.33	194.106.44.61	TCP	[TCP segment of a reassembled PDU]
293718	476.373389	Cisco:74:75:e0	Broadcast	ARP	who has 10.20.169.129? Tell 10.20.10.6
293719	476.373665	10.20.19.33	194.106.44.61	TCP	[TCP segment of a reassembled PDU]
294412	477.707580	10.20.23.103	202.157.188.189	TCP	[TCP Dup ACK 294389#1] vpad > http [ACK] Seq=1439 Ack=202195 Win=17184 Len=0 SLE=20
294413	477.711780	202.157.188.189	10.20.23.103	TCP	[TCP segment of a reassembled PDU]
294414	477.712237	85.137.163.28	10.20.19.33	eDonkey	eDonkey TCP: Hello Answer
294415	477.713591	10.20.22.103	202.157.188.189	TCP	[TCP Dup ACK 294389#4] vpad > http [ACK] Seq=1439 Ack=202195 Win=17184 Len=0 SLE=20
294416	477.714125	10.20.14.3	10.0.8.20	DNS	Standard query A teredo.ipv6.microsoft.com
294417	477.718493	210.213.244.87	10.20.25.29	BitTorrent	Continuation data
294418	477.719969	10.20.25.29	210.213.244.87	TCP	quaddb > 6880 [ACK] Seq=373681677 Ack=2520210871 Win=16560 Len=0 SLE=2520212251 SRE=
294419	477.723013	64.70.34.135	10.20.17.26	TCP	[TCP segment of a reassembled PDU]
294420	477.723044	211.121.98.116	10.20.19.33	TCP	oms > re-conn-prot [ACK] Seq=553 Ack=2240255 Win=55535 Len=0
294421	477.723655	10.20.22.225	67.49.46.221	TCP	webmachine > 37301 [SYN] Seq=0 Win=65535 Len=0 MSS=1460
294422	477.724706	10.20.27.145	10.2.8.10	SSH	Encrypted request packet len=36
792846	4933.213306	154.20.75.165	10.20.23.55	TCP	33284 > ncadg-ip-udp [ACK] Seq=2076415128 Ack=3877599199 Win=63023 Len=0
792847	4933.215531	10.20.23.55	154.20.75.165	TCP	[TCP segment of a reassembled PDU]
792848	4933.216693	10.20.23.55	154.20.75.165	BitTorrent	Continuation data
792849	4933.217862	10.20.23.55	154.20.75.165	BitTorrent	Continuation data
792850	4933.219027	10.20.23.55	154.20.75.165	BitTorrent	Continuation data
792851	4933.220197	10.20.23.55	154.20.75.165	BitTorrent	Have, Piece (Idx:0xd464d808)
792852	4933.221359	10.20.23.55	154.20.75.165	BitTorrent	Continuation data
792853	4933.222528	10.20.23.55	154.20.75.165	BitTorrent	Continuation data
792854	4933.223694	10.20.23.55	154.20.75.165	BitTorrent	Continuation data
792855	4933.224861	10.20.23.55	154.20.75.165	BitTorrent	Continuation data
792856	4933.226027	10.20.23.55	154.20.75.165	BitTorrent	Continuation data

Figure 6: Snapshots of selected portions of captured traffic to highlight that eDonkey, BitTorrent and ICMPv6 packets existed in the UiTM Perlis campus network.

6. Concluding remarks

The large number of incoming and outgoing network traffics occur during day-time (8am-6pm) compared to night-time (6pm – 4am) has introduced an unbalanced bandwidth usage. Heavy usage during day time affects the Web access performance which resulted in slow response time but during night-time the bandwidth is always under-utilized. As to improve the unbalance bandwidth utilization between day-time and night-time we proposed a short-term and long-term web prefetching as suggested by Crovella and Barford (1998) may be introduced by network administrators to utilize the night-time wasted bandwidth.

In terms of popular application layer protocol, HTTP is still a dominant protocol and it is in line with other previous studies by other researchers. The worrying parts of the finding are in terms of significant number of access and data transferred by Peer-to-Peer (P2P) applications. The large number of access to P2P networks to download big files will eat up the limited WAN link bandwidth leaving other users suffer from a very slow response time. Putting a bandwidth shaper application to limit the access to the P2P network may be able to ease the problem. Another worrying trend that will waste a WAN link bandwidth is the significant user continuous retry to access Web sites that is considered illegal from the UiTM Web access policy point of view that was being blocked by the UiTM proxy server.

The finding of the low percentage of successful connections and the finding of some traffic anomalies requires more in-depth further studies. We believe the overall percentage of successful connection at only 39 percent is considered very low and this means that a high amount of WAN link bandwidth is wasted to the unsuccessful access. The findings support our expectation that there are possibilities that the congestion on the WAN link are due to legitimate high bandwidth usage during working hours as well as the possibility of anomalous traffic. Our future work will be emphasized on thorough investigation on the unsuccessful connections, anomalous traffic and potential security attacks for both IPv4 as well as IPv6 packets in the campus network.

REFERENCES

- Barford, P. & Crovella, M. (1999). Measuring Web Performance in the Wide Area. Computer Science Department, Boston University.
- Barford, P., & David Plonka, D. (2001). Characteristics of network traffic flow anomalies. Proceedings of the First ACM SIGCOMM Workshop on Internet Measurement, pp. 69-73, San Francisco, California, USA
- Caceres, R., Duffield, N. G., Fieldman, A., Friedmann, J., Greenberg, A., Greer, R., Johnson, T., Kalmanek, C., Krishnamurthy, B., Lavelle, D., Mishra, P.P., Ramakrishnan, K.K., Rexford, J., True, F., and van der Merwe, J.E. (2000). Measurement and Analysis of IP Network Usage and Behavior. AT&T Labs, USA
- Courcoubetis, C, Siris, V. A. (1999). Measurement and Analysis of Real Network Traffic. Technical Report 252, ICS-FORTH, March 1999. Retrieved August 15, 2005, from http://www.ics.forth.gr/netlab/publications/1999.tr252_measurements.pdf.
- Crovella, M. and Barford, P. (1998). The Network Effects of Prefetching. In Proceedings of IEEE Infocom '98, San Francisco, CA.
- Fomenkov, M., Keys, K., Moore, D., & Claffy, K. (2004) . Longitudinal study of Internet traffic in 1998-2003. Proceedings of the winter international symposium on Information and communication technologies, pp.1-6, Cancun, Mexico
- iNetmon, iNetmon Portable Plus. Retrieved Jan 15, 2008, from <http://www.inetmon.com>
- Kotz, D. & Essien, K. (2002). Characterizing Usage of a Campus-wide Wireless Network. Dartmouth Computer Science Technical Report TR2002-423, Dartmouth College Hanover, NH, USA
- Murnan, C.A. (2001). Bridging the bandwidth gap: measures to maintain bandwidth availability. Proceedings of the 29th annual ACM SIGUCCS conference on User services, pp. 115-119, Portland, Oregon, USA
- Nametka, W. (1999). Network performance and capacity planning: Techniques for e-business world. Retrieved March 24, 2003, from <http://www-1.ibm.com/services/its/us/source/nametka.pdf>
- Paxson, V. (1995). TCP-Reduce, Retrieved Oct 30, 2005, from <http://ita.ee.lbl.gov/html/contrib/tcp-reduce.html>
- Stern, P. (2001). Measuring Early Usage of Dartmouth's Wireless Network. Dartmouth College Computer Science Technical Report TR2001-393.
- TamoSoft. (2005). Promiscuous Monitoring in Ethernet and Wi-Fi Networks. Retrieved Dec 13, 2005, from <http://www.tamos.com/docs/monitoring.pdf>

- Thottan, M. & Ji, C. (2003). Anomaly Detection in IP Networks *IEEE Transactions on Signal Processing*, vol.51, no.8.
- Zhang, J., Yang, J., An, A. & Wang, J. (2005). Traffic Measurement and Analysis of TUNET, Proceeding of the 2005 International Conference of Cyberworlds, Singapore, November 2005.
- Ziviani, A. (2006). An Overview of Internet Measurements: Fundamentals, Techniques, and Trends, *African Journal of Information and Communication Technology*, Vol. 2, No. 1, March 2006
- Setting Up the System for Roving Analysis. Retrieved 13 Dec, 2005, from http://support.3com.com/infodeli/tools/switches/s_stack2/2200/ssadmin/a09rmont.html