Traffic growth analysis over three years in enterprise networks

Conference Paper · November 2009

DOI: 10.1109/APCC.2009.5375520 · Source: IEEE Xplore

5 authors, including:

Seong-Cheol Hong
Pohang University of Science and Technology
9 PUBLICATIONS **113** CITATIONS
SEE PROFILE

Byungchul Park
University of Toronto
28 PUBLICATIONS **180** CITATIONS
SEE PROFILE

Traffic Growth Analysis over Three Years in Enterprise Networks

Seong-Cheol Hong, Jin Kim*, Byungchul Park, Young J. Won, and James W. Hong

Dept. of Computer Science & Engineering, POSTECH, Korea
*Graduate School of Information Technology, POSTECH, Korea
{pluto80, *aroma, fates, yjwon, jwkhong}@postech.ac.kr

***Abstract* – The increasing popularity of P2P and multimedia applications around the world has resulted in significant global increases in network traffic. This paper investigates the traffic changes in our enterprise network by analyzing the day-long traffic traces from 2007 to 2009. We observe that there has been a substantial increase in traffic in the past few years; however, no single application type is a dominant cause. This paper analyzes the following categories: Traffic volume, flow characteristics and protocol usage, and the use of IP address space.**

***Abstract* – Passive monitoring, Internet traffic, traffic analysis**

I. Introduction

There is no doubt that popular application services on the Internet have driven a vigorous traffic volume increase. Peer-to-peer (P2P) and multimedia traffic applications are believed to be generating a significant amount of traffic in both enterprise and residential networks. In order to effectively observe and determine the traffic growth patterns, it is necessary to undertake traffic trend analyses over long time scales (e.g., yearly).

Kim et al. [1] show that P2P applications generate a substantial volume of Internet traffic in enterprise networks. In 2008, a study by Japanese ISP [2] observed that a significant portion of P2P applications is replaced by multimedia and Web traffic. This phenomenon is due to the increased popularity of video streaming services such as YouTube. Yearly comparison analysis of the collected packet traces are presented in order to show the residential traffic patterns in Japan. According to their study results, the traffic growth is clear; however, the growth pattern may differ according to regionally popular applications.

This paper investigates the traffic growth details reflecting Internet user behavior. We provide a trend analysis of our campus network, based on day-long network snapshots from the previous three years.

The rest of the paper is organized as follows. Section II provides an overview of the traffic collection location and data sets. The measurement categories and analysis results are described in Section III. Finally, concluding remarks are given, and possible future work discussed, in Section IV.

II. Data Set

Since 2007, we have obtained day-long traffic traces for incorporation into the yearly DITL [5] traffic collection. We collected the traffic traces from one of the two Internet junctions at POSTECH, a university with a user population of about 3500 people. To avoid any possible packet loss, the monitoring probe equipped with Endace DAG 4.3GE [6] was used to monitor 1 Gbps Ethernet link. The collected traces present the bidirectional packet traces of the network. Table 1 shows the date and captured trace size of each collection trial. Three consecutive years worth of day-long sample traces allow us to compare snapshots of the network, and observe any changes in the behavior of internet users.

Table 1. Traffic traces

Data set	*Date (GMT)*	*Captured size*
’07	Jan. 9-10, 2007	0.8 TB
‘08	Mar. 18-19, 2008	0.9 TB
‘09	Mar. 30-Apr. 1, 2009	1.2 TB

The meta information for each data set (except for ‘09) is available in DatCat [7]. The first 98 bytes of packet were captured. A single hour trace is about 20~30 GB. In this study, we extracted hour traces from each data set to represent the specific time of the day: 5:00-6:00, 13:00-14:00, and 20:00-21:00. Each time interval reflects light traffic with fewer users, traffic load during typical office hours, and heavy user traffic at night, respectively. Overall, three hour traces from each data set were sampled and analysed to enable yearly comparisons of traffic patterns.

III. Traffic Analysis

This section provides an empirical analysis of our enterprise traffic. The following analysis categories are investigated accordingly: Traffic volume, flow characteristics and protocol usage, and use of IP address space.

A. Traffic volume

Fig. 1 illustrates the change in bandwidth consumption over the three years. The average bandwidth increased continuously, and there is clear growth throughout the entire monitoring period. Table 2 shows the growth ratios in each hour interval. As can be seen, the bandwidth during the peak times (13:00 and 20:00) consistently increased by 20-30% each year. This demonstrates that our campus traffic follows a typical time-of-day trail where traffic increases during office hours and at night. This is typical user behavior in commercial residential networks.

Table 3 illustrates that the most of the bandwidth is occupied by TCP and UDP packets. Although the UDP traffic volume increases significantly every year, TCP is still the dominant protocol, as shown in Table 4.

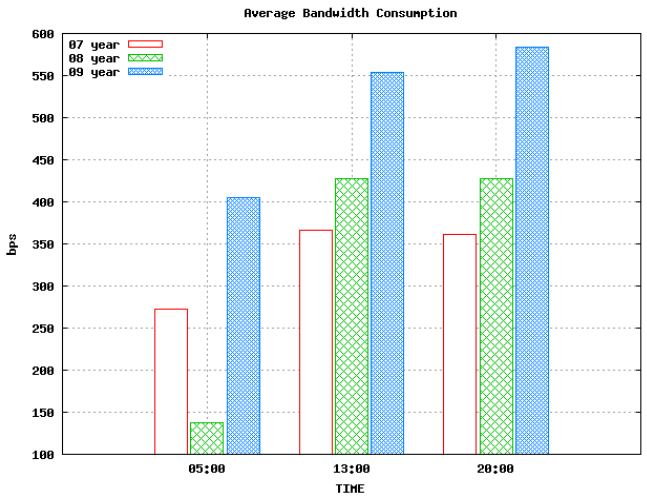

Fig. 1. Average bandwidth of sample time interval

Table 2. Average bandwidth (Mbps) and growth ratio

Time	*'07*	*'08*	*'09*
5:00~6:00	272	136(-50%)	405(+197.79%)
13:00~14:00	366	427(+16.67%)	554(+29.74%)
20:00~21:00	361	428(+18.56%)	584(+36.45%)

'+': increase, '-': decrease (against the previous year's value)

Table 3. Total vs. TCP+UDP bytes (Mbytes)

Time	*'07*	*'08*	*'09*
5:00~6:00	981 / 980	492 / 491	1457 / 1445
13:00~14:00	1319/1317	1538/1537	1994/1992
20:00~21:00	1298/1296	1540/1539	2103/2096

Table 4. TCP vs. UDP bytes (Mbytes)

Time	*'07*	*'08*	*'09*
5:00~6:00	970 / 8	482 / 8	1425 / 19
13:00~14:00	1286 / 30	1503 / 34	1956 / 36
20:00~21:00	1278 / 17	1512 / 26	2025 / 70

B. Flow Characteristics & Protocol Usage

In this section, we analyzed characteristics of Internet traffic from the perspective of flows. Flow is a unidirectional sequence of packets, typically sharing the same 5-tuple meta information: Source IP, destination IP, source port, destination port, and protocol. Flow analysis reflects more general traffic characteristics than packet-level analysis. We focused on flow size (byte) and duration distribution over the monitoring periods.

We analyzed the proportions of Web traffic (http/https) to confirm the growth of multimedia service traffic based on port numbers. We assumed that the recent growth of Web traffic is attributed to the increasing popularity of multimedia services such as YouTube. Table 5 illustrates protocol usage in traffic volume. It implies the traffic portion of each application in our network. The Web traffic ratio decreased from 11% in 2007 to 10% in 2009. Meanwhile, the volume of Web traffic increased proportionally to the total traffic in terms of total flow counts and bytes. Kenjiro et al. [2] report that the growth of HTTP traffic takes the lead in recent traffic; however, our results do not show such a pattern. Instead, we observe that the occupancy ratios of each application remain relatively steady, as in Table 5, which implies that no single application type is the dominant cause of traffic growth in our network. This is due to regional variances in the popularity of Internet applications. The volume ratios for the other protocols remain relatively constant each year.

Fig. 2 and 3 show the CDFs of flow size and duration, respectively. While the numbers of flows and total bytes increase each year, the characteristics of each do not change significantly. In the case of flow size distribution, 90% of flows are less than 10,000 bytes in size, regardless of the time and year. Similarly, more than 95% of flows have a duration less than 100 seconds, although there are small variations depending on the time interval.

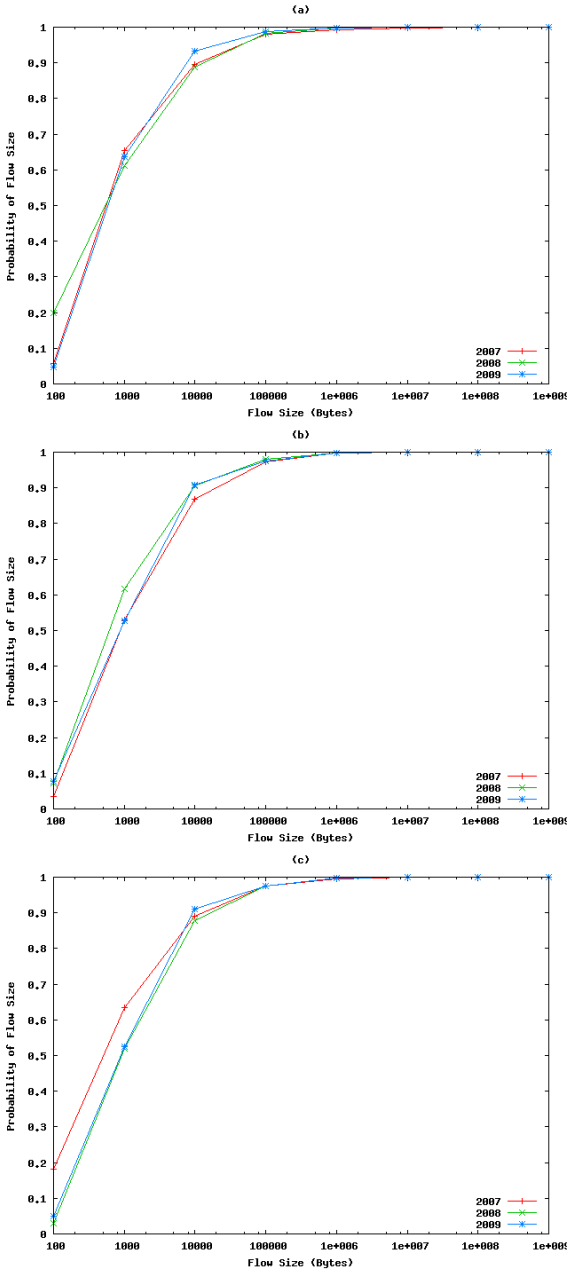

Fig. 2. CDF - flow size (bytes) distribution; (a) 5:00-6:00; (b) 13:00-14:00; (c) 20:00-21:00

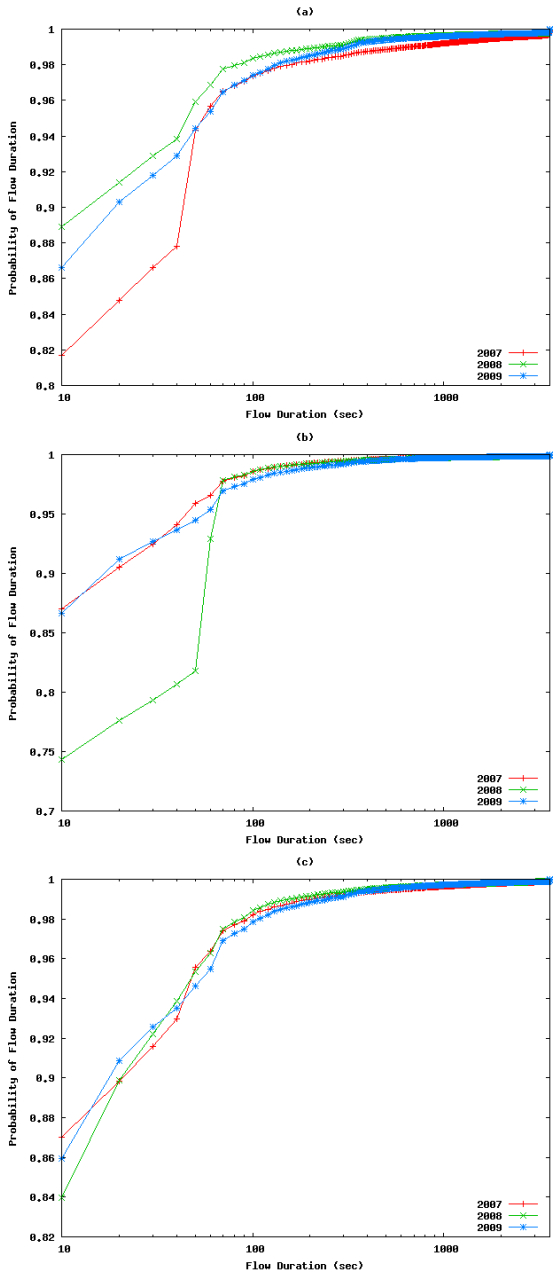

Fig. 3. CDF – flow duration distribution; (a) 5:00-6:00; (b) 13:00-14:00; (c) 20:00-21:00

C. Use of IP address space

IP address management in enterprise networks is important from the perspective of network security [3]. Detecting security incidents such as scanning or DDoS activities is essential. One way to detect abnormal activities is to analyze the distribution of unused IP address space. Such traffic strongly indicates malicious activity such as DDoS backscatter, port scanning, and probing activity from various worms.

We found an abnormal phenomenon in relation to used IP addresses, which are internal IP addresses that appear at least once in the collected traces. Table 6 represents the number of entire used IPs in the POSTECH network. We observed that light traffic with relatively fewer users occurs at 5:00, and that the number of used IPs at this time is not much different than in other time slots. We divided *used IP* into two groups according to where the given IP appears in the IP header: *used source IP,* and *used destination IP*. At 5:00, the number of *used source IP* is 4,888 and the number of *used destination IP* is 32,699. This infers that there are many packets which originated from outside of the enterprise network to unused IP addresses. We suspect scanning activity at the time of measurement.

To demonstrate the differences between the usage patterns of *used source IP* and *used destination IP*, we present their distributions in Fig. 4. The POSTECH network is allocated with B class address space. We set the third octet value in the IP address as the radian angle, and the fourth octet value as the radius. In Fig. 4, the distribution of *used destination IP* is much denser than *used source IP*. Most packets in *used destination IP* have a small byte size (<60 bytes). Thus, we speculate that most packets at night are malicious traffic.

Table 5. Protocol usage in volume

Protocol	***Port***	***'07***	***'08***	***'09***
	Total	**93.90**	**94.11**	**93.75**
TCP	(<1024)	15.93	13.27	10.98
	80 (http)	11.31	12.12	10.07
	544 (rtsp)	-	0.00	-
	443 (https)	0.06	0.28	0.22
	20 (ftp-data)	0.09	0.05	0.09
	110 (pop3)	0.01	0.01	0.01
	22 (ssh)	0.00	0.00	0.00
	25 (smtp)	0.14	0.18	0.12
	others	4.31	0.63	0.46
	(>=1024)	77.97	80.84	82.77
	6346 (gnutella)	0.00	0.00	0.00
	6699 (winmx)	0.00	0.00	0.00
	7743 (winny)	0.00	0.00	0.00
	1935 (rtmp)	0.01	0.01	0.05
	6881 (BT)	0.01	0.04	0.22
	4662 (edonkey)	0.01	0.00	0.00
	others	77.93	80.79	82.48
UDP	**Total**	**5.91**	**5.53**	**5.84**
	53 (dns)	0.04	0.05	0.06
	others	5.87	5.47	5.78
ICMP		**0.17**	**0.36**	**0.35**
Others		**0.03**	**0.00**	**0.06**

Table 6. Used IP address counts

	5:00~6:00	***13:00~14:00***	***20:00~21:00***
'07	8471	10924	9920
'08	32647	32738	31256
'09	26255	32754	28005

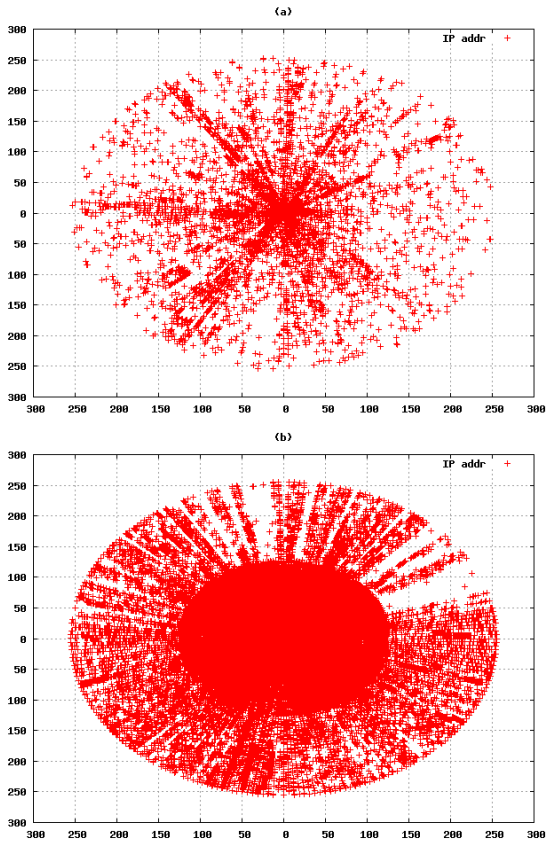

Fig. 4. Distribution of (a) used source IP, (b) used destination IP - 5:00-6:00, 2008

IV. Conclusion

In this paper, we have presented a measurement study and trend analysis of the traffic traces from our campus network. Through the comparison of yearly network snapshots, we observed a clear pattern of traffic volume growth. We summarize the analysis results as follows:

- There is a clear sign of average bandwidth increase. This increase occurs irrespective of the introduction of new applications.
- There is no significant change in flow characteristics. The flow size and duration analysis showed no significant changes over the years.
- According to the port-based traffic analysis, no single application type was responsible for the traffic growth. This finding is in contrast to previous traffic studies in other regions. The relative proportions of each application remain similar for each year.
- There are frequent accesses the unused IP address space in our network.

For future work, we plan to focus more on insightful application traffic analysis information of the collected traces. We will also continue to obtain daily network snapshots to provide long-term traffic trend analyses.

Acknowledgements

This work was partly supported by the IT R&D program of MKE/IITA [2008-F-016-02, CASFI] and WCU program through the KSEF of MEST, Korea [R31-2008-000-10100-0].

References

[1] Myung-Sup Kim, Young J. Won, and James W. Hong, "Characteristic analysis of internet traffic from perspective flows", Computer Communications, Volume 29, Issue 10, June 19 2006, pp. 1639-1652.

[2] Pierre Borgnat, Guillaume Dewaele, Kensuke Fukuda, Patrice Abry, and Kenjiro Cho, "Seven Years and One Day: Sketching the Evolution of Internet Traffic," INFOCOM 2009, Rio de Janeiro, Brazil. April 2009.

[3] David Moore, Geoffery M. Voelker, and Stefan Savage, "Inferring Internet Denial-of-Service Activity," 10th USENIX Security Symposium, Washington, D.C., USA, August 13-17 2001, pp. 9-22.

[4] Khadija H. Ramah, Hichem Ayari, and Farouk Kamoun, "Traffic Anomaly Detection and Characterization in the Tunisian National University Network," IFIP Networking 2006, Coimbra, Portugal, May 15-19 2006, pp. 136-147.

[5] A Day in the Life of the Internet (DITL), http://www.caida.org/projects/ditl/.

[6] Endace, DAG 4.3GE, http://www.endace.com/.

[7] Internet Measurement Data Catalog (DatCat), http://www.datcat.org/.