2009 ISECS International Colloquium on Computing, Communication, Control, and Management

Observations of IPv6 Traffic

Wenchao Shen, Yanjiao Chen, Qianli Zhang, Yang Chen, Beixing Deng, Xing Li
Department of Electronic Engineering
Tsinghua University
Beijing, P.R.China
e-mail: shenwenchao@gmail.com

Guohan Lv
Wireless & Networking Group
Microsoft Research Asia
Beijing, P.R.China
e-mail: lguohan@microsoft.com

***Abstract*—Our understanding of IPv6 usage is quite limited. In this paper, we carry out a detailed analysis of Netflow records collected from a China Tier-1 ISP. Many aspects related to traffic attributes and deployment issues are investigated: application types, transitional technologies, interface identifier assignment schemes, etc. Dominating of video streaming and P2P traffic as well as the widely use of random identifiers for privacy are most peculiar characteristics in China's IPv6 networks. To have a better understanding of how IPv6 network is used, we studied users' behavior from several perspectives, such as the traffic volume produced by each user, number of applications he uses and the frequency he uses IPv6 networks. Results show that IPv6 is only effectively used by a small fraction of the users. On the whole, our study suggests that from our vantage point, China's IPv6 network is in the transitional phase from the initial experimental period to practical application.**

Keywords - IPv6; traffic measurement; user behaviour; Netflow

I. INTRODUCTION

With IPv4 addresses becoming scarcer and scarcer, the current Internet protocol is showing its age. People are looking forward to IPv6 as the solution. Although IPv6 is beginning to see larger deployments, our understanding of it is surprisingly limited. Questions such as, ``what the applications being used are'', ``how users connect to IPv6 network'', and ``whether IPv6 is frequently used or just used for a try'' remain largely unanswered.

To answer these questions, we analyze traffic from a link connecting a China's major ISP with the whole country's PoP. China, with its huge population and disproportionally limited IPv4 addresses, is quite active in IPv6 deployment. The CNGI (China Next Generation Internet) project promoted by the government has built a nationwide IPv6 network several years ago. Since China was in the leading position of IPv6 deployment, exploring its special characteristics would be quite meaningful.

There are very few traffic analysis results regarding to IPv6 networks. [10], [11] analyzed traffic seen at public 6to4 relays. Arbor Networks [9] studied how much IPv6 traffic is on the internet and found that IPv6 traffic is growing at the same rate as IPv4 traffic. A recent study by Google [3], studied IPv6 penetration in the world, and found IPv6 penetration was still less than 1% of internet traffic in any country. Their methodology is to trace some of the users visiting Google and study their IPv6 connectivity.

David Malone's work [1] on IPv6 addresses analyzed transitional technologies and the assignment of IPv6 address to machines. While we use some of the same techniques in a part of our study, we have quite different data sets. Instead of ftp site log, name server log and traceroute data, we use Netflow [4] records collected from real networks. Also we have much broader view instead of focusing on addresses. In a most recent study [2], E. Karpilovsky etc. reported traffic analysis results in a US tire-1 ISP. While we have similar datasets, our observations are quite different in application types and address analysis. We also perform many additional analyses and offer a new perspective in user behaviour tracking.

The paper is organized as follows. Section II introduces the data source we use in the paper. Application types used in IPv6 networks are analyzed in Section III. The results of address analysis, namely observations of transitional technologies and interface identifier assignment schemes, are presented in Section IV. We investigate users' behavior in Section V and conclude with Section VI.

II. DATA SET

The data is from a link connecting a China's major ISP with the country's PoP. The traffic rate of each direction is on an average of 550 Mbps in the daytime and becomes very low in the night. Fig. 1 shows the ISP's incoming traffic rate from 2009-3-14 to 2009-3-21.

We build a high-speed flow aggregation system which is able to process all the packets without sampling and generate Netflow records in Netflow V9 format. The concurrent flow number of the link is about 10,000 in the daytime.

We collect Netflow records of both directions from 2009-2-21 to 2009-3-21. Time frame is four weeks long.

III. APPLICATION MIX

Straightforward statistics show TCP makes up 91% the total traffic, UDP makes up another 9%, leaving ICMP around 0.01%. To further analyze the application types, we use the signature *(source port, destination port, protocol)* to map it to an application name. This analysis yields quite interesting results as presented in TABLE I. Traffic from

978-1-4244-4246-1/09/$25.00 ©2009 IEEE CCCM 2009

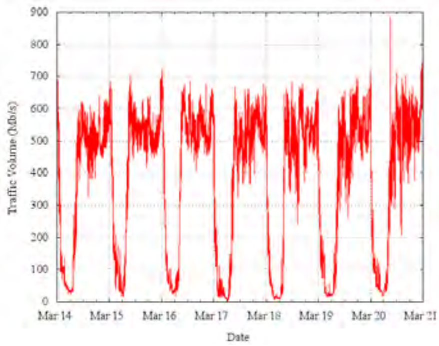

Figure 1. The ISP's incoming traffic

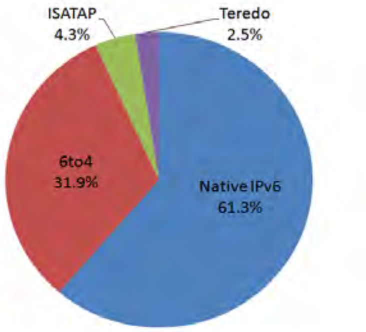

Figure 2. Transitional technologies

traditional applications such as Web, FTP and Mail is surprisingly low, even the sum of them is still less than 1%. Over 85% of total traffic is from unknown ports. As random port numbers are characteristic of P2P traffic, deep investigation reveals that the majority of the unidentified traffic is P2P. There is also a higher ratio in RTP/RSTP traffic. Manual check shows that most of this traffic comes from a few sites who offer live streaming of television programs.

Results observed here are quite different from that in [2]. In their observation, DNS queries make up 85% of total traffic and ICMP probes make up 8%, with little realistic traffic. They believe that, the high ratio of probing traffic demonstrates, at least in their observed tire-1 ISP, customers view IPv6 as experimental. Though we have a different observation time window, we believe the sharp contrast mainly comes from the different vantage points. In China, IPv6 has been deployed and promoted by the government for several years. The application mix observed here indicates that, in some way, IPv6 in China has stepped over the experimental period and entered into practical phase.

Most applications here are bandwidth consuming ones. It seems that people are attracted by the higher available bandwidth here, as IPv4 network is quite crowded. Survey of the IPv6 users confirms that higher download speed is one of the main reasons.

IV. IPv6 Address Analysis

Unlike IPv4 addresses, IPv6's extra expressiveness can tell us how they are being used. As is known to all, a 128-bits long IPv6 address can be divided into two parts: the upper 64-bits network prefix which gives valuable information about transitional technology used, and the lower 64-bits interface identifier which is reserved entirely for host machines. Through analyzing IPv6 addresses, two questions regarding to IPv6 deployment can be answered here: how people are transitioning their IPv4 networks to IPv6, how IPv6 addresses are actually assigned to individual machines.

A. *Transitional Technologies*

The migration from IPv4 to IPv6 will be a long process. Along the way, there are several transitional technologies designed for the co-existence of IPv4 and IPv6. We identify three most widely used technologies here: 6to4, ISATAP [7] and Teredo [5]. From address prefix, 6to4 (2002::/16) and Teredo (2001:0000::/32) can be easily identified. With the first 32 bits of the interface identifier to be 0000:5efe or 0200:5efe, ISATAP can also be identified. The other transitional technologies are rare in practice.

The results are presented in Fig. 2. Compared with results presented in [2], we have very similar observation of 6to4 and Teredo technologies, but quite different observation of ISATAP. Note that, here, there is a significant amount of ISATAP. But in [2], there is hardly any ISATAP observed. It's easily seen that, with 6to4 as the dominating transitional technology like in other countries, ISATAP plays an important role in IPv6 networks of China.

B. *Assigning Addresses to Machines*

The low order 64 bits of the IP address, namely the interface identifier, are reserved entirely for host machines. We can use this to see how organizations assign addresses to machines. The methodology used here is similar as presented in [1]. The assignment types are: Auto-configured (generated using EUI-64 algorithm based on MAC address), Teredo (Teredo encodes options and routing information in the identifier), Low (only last 16 bits are used), IPv4 Based (identifier contains IPv4 address), Privacy (temporary identifier generated according to the IPv6 privacy specification [6]), and Unidentified.

TABLE I. Application Mix in Byets

ICMP	*DNS*	*RTP/RSTP*	*Web*	*FTP*	*Mail*	*Others*
0.013%	0.020%	12.430%	0.348%	0.298%	0.000%	86.891%

TABLE II. Assignment Schemes Seen

Auto-configured	*ISATAP*	*Teredo*	*Low*	*IPv4 Based*	*Privacy*	*Unidentified*
2.2%	4.3%	2.5%	0.5%	19.9%	51.8%	18.8%

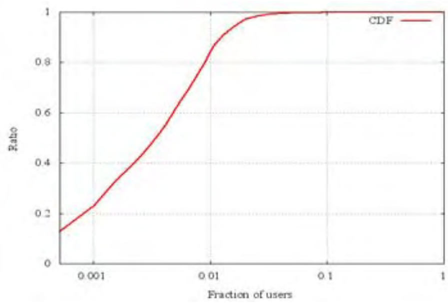

Figure 3. Per user traffic volume CDF

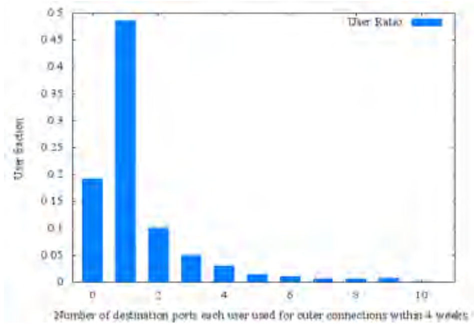

Figure 4. Number of applications each user use

Ratios of various host configurations are shown in TABLE II. Note that, as reported in [1], using the same criteria, we can only identify 73.35% of privacy addresses. From TABLE II. 51.8% of addresses are identified as privacy addresses. However, the actual figure should be adjusted to 0.518/0.7335 or 70.6%. This is quite surprising, as both figures in [1] and [2] are no more than 6%. They reported that privacy extensions remain relatively unused, but our observation is just the opposite. Windows Vista uses privacy identifiers by default [8]. From our own experience, Windows XP SP2 and XP SP3 use privacy identifiers by default too, but in Linux and Mac OS, auto-configuration is the default. Considering the different observing time-frames and also the prevalence of Windows in China, these could be the reason.

V. User Behaviour Tracking

It's quite interesting to see how the IPv6 network is being used. Do the users generate similar amount of traffic volume? Do they use IPv6 frequently for everyday use or just occasionally for a try? How many applications each user use? In this section, we will try to answer these questions. This will help us have a better understanding of the network.

As we mentioned above, about 70% of addresses are private ones. Privacy address users will renew their addresses periodically. That's to say, for most IP addresses, they cannot be used to trace the users. However, there are addresses which can be used to identify the user with no ambiguity: the auto-configured ones, as each of them is constructed from the MAC address which is unique in the world. In this section, we will use these addresses from the customer ISP to study the users' behavior.

As we use the MAC address as the identifier of each user, it is interesting to know how many network prefixes each user (or exactly NIC) has. Our observation shows that about 95% of the users have only one network prefix, another 4% have two prefixes. It's quite rare for a single user to have three or more prefixes.

TABLE III. Usage percentage among users

ICMP	***DNS***	***RTP/RSTP***	***Web***	***FTP***	***Mail***	***Others***
39.4%	3.6%	5.6%	50.3%	7.9%	0.6%	46.5%

A. *Minority of Users Generate Majority of Traffic*

First, we examine how much traffic each user generates. We sort the users by their traffic volume in descending order, and draw CDF in Fig. 3. It's easily seen, top 1% of the users take up almost 85% of total traffic volume, while lower 97% of users only take up 1% of traffic. This surprising figure shows that most of current IPv6 network traffic is generated by a very small fraction of users.

B. *Application Usage*

From previous results, it seems that IPv6 is not effectively used by most users. For further investigation, we will examine the application usage among users.

In Section III, we presented results of *what percentage of total traffic* each application takes up. However, here, we will take a look at the application mix from the user's perspective: given an application, *what percentage of users* that have used it?

The results are presented in TABLE III. Note that, the sum of the percentage is not 1, as a user may use several applications. It's easily seen that Web is used by *50% of the users*. From TABLE I. , web only makes up *less than 1% of total traffic*. It's similar as to FTP and Mail. These traditional applications are used by a larger fraction of users, although they seem negligible in the total traffic.

Then, let's look at the number of applications each user use. In former analysis, we use the label "Others" to include all applications which are ignored or unidentified, such as Telnet, SSH and P2P applications. However, while counting the number of applications, this trick cannot be used, and it is almost impossible to enumerate all applications. For feasibility and simplicity, we turn to *destination port number* as the indication.

For each user in the customer ISP, we count how many destination ports he used when connecting to the outside internet. As we focus on applications used, ICMP probes and DNS queries are excluded. That is to say, users who only generates ICMP probes or DNS queries but not realistic traffic are labeled with 0 destination port.

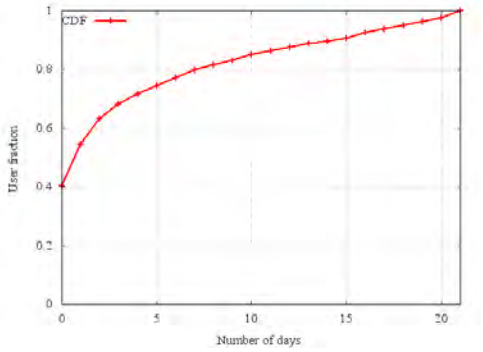

Figure 5. Usage Frequency

Results are presented in Fig. 4. Note that, users with more than 10 ports are not shown in the figure and they make up about 9% of total users. Manual check shows their high port numbers mostly come from P2P applications. Though this simple scheme is flawed, it can still be a good indication to see how many applications each user use. In the four-week long time frame, almost 20% of users only generate ICMP probes or DNS queries, another 48% of users only use one application, over 80% percent of users use no more than two applications. Deeper investigation shows that, among the users who only use only one application, 70% of them use Web.

Along with former results, it's easily seen, the applications in the IPv6 network are still not rich enough. Most people are just here for one or two applications. Survey and further investigation show that, users are mainly attracted for two reasons: one is the relatively not crowded link, which is most suitable for bandwidth consuming applications such as video streaming and p2p sharing, the other is to avoid the fees in IPv4 networks: many users are charged by the traffic volume they used in IPv4 networks while free of charge in IPv6. The network is effectively used only by a very small fraction of users.

As we mentioned in Section 3, ICMP probes and DNS queries make up over 90% of total traffic in a US tire-1 ISP. And here in our targeting ISP, nearly 20% of the users only use ICMP or DNS. Apart from the sharp contrast, there is also something in common. The reasons of so much ICMP and DNS are as demonstrated in [2].

C. Usage Frequency

Another interesting issue is how often the users use IPv6. We will take a look at this problem in this paragraph. The target is still the users in the customer ISP with MAC based addresses. The four-week long observation is divided into separate days. First, we list the users who appear in the first week, then we watch these users and count how many days they "appear" in the following three weeks. As we just mentioned, 20% of the users only produce ICMP probes or DNS queries not realistic traffic, so the definition of "appear" exclude such situations. A user won't "appear" unless he produces realistic traffic other than ICMP probes or DNS queries.

The results are shown in Fig. 5. About 40% of the users never appear in the following three weeks, and 25% of users appear in more than five days. We can roughly draw the conclusion that a quarter of the users use IPv6 quite frequently, while nearly a half of the users use IPv6 occasionally. This indicates many users still treat IPv6 network as experimental.

VI. CONCLUSIONS

IPv6 is beginning to see larger deployments. For a better understanding of its current usage, we performed a detailed analysis of traffic seen in China's IPv6 network. Unlike networks in other places which are dominated by ICMP probes and DNS queries, IPv6 is moderately in real use in China, with most of its traffic being P2P sharing or video streaming. With 6to4 as the main transitional technology, ISATAP also plays a nonnegligible role. Most addresses have temporary interface identifier for privacy. Although China's IPv6 networks are quite different from those experimental ones, further investigation shows: most traffic is produced by a very small fraction of users, most users use no more than two applications and many users just use IPv6 occasionally. Above all, we can see, IPv6 in China is still in the transitional phase from initial experimental period to practical application.

As future work, we will collect more Netflow data and study the trend of these characteristics. Moreover, it would be interesting to perform a detailed analysis of the seen P2P traffic. Knowing how they perform and why they choose IPv6 would help us in search of "the killer application".

ACKNOWLEDGMENT

This work is supported by Natural Science Foundation of China (No.60850003), National Basic Research Program of China (973 Project, No.2007CB310806) and National High Technology Development Program of China (No.2007AA010306).

REFERENCES

[1] D. Malone, "Observation of IPv6 Addresses," Claypool, M., Uhlig, S. (eds.) PAM 2008. LNCS, vol. 4979, pp. 21–30. Springer, Heidelberg (2008)

[2] E. Karpilovsky, A. Gerber, D. Pei, J. Rexford and A. Shaikh, "Quantifying the Extent of IPv6 Deployment", Proc. Passive and Active Network Measurement (PAM 2009), in press.

[3] Steinar H. Gunderson, "Global IPv6 Statistics - Measuring the current state of IPv6 for ordinary users" November 2008, http://www.ietf.org/proceedings/08nov/slides/v6ops-4.pdf

[4] Cisco NetFlow, http://www.cisco.com/warp/public/732/netflow/.

[5] C. Huitema, "Teredo: Tunneling IPv6 over UDP through NAT," RFC 4380, 2006.

[6] T. Narten, R. Draves, S. Krishnan, "Privacy Extensions for Stateless Address Autoconfiguration in IPv6," RFC 4941, 2007.

[7] F. Templin, F. Templin, D. Thaler, ``Intra-Site Automatic Tunnel Addressing Protocol (ISATAP)," RFC 5214, 2008

[8] J. Davies, "IPv6 Autoconfiguration in Windows Vista,", August 2008, http://technet.microsoft.com/en-us/magazine/2007.08.cableguy.aspx

[9] S. Iekel-Johnson, C. Labovitz, D. McPherson, H. Ringberg, "Tracking the IPv6 Migration," 2008, http://www.arbornetworks.com/IPv6research

[10] P. Savola, "Observations of IPv6 Traffic on a 6to4 Relay," ACM SIGCOMM Computer Communication Review, vol. 35, January 2005, pp. 23-28.

[11] Y. Hei, K. Yamazaki, "Traffic analysis and worldwide operation of open 6to4 relays for IPv6 deployment," Proc. Symposium on Applications and the Internet, 2004