On Characterizing Peer-to-Peer Streaming Traffic

Citation:

Jie Yang, Lun Yuan, Chao Dong, Gang Cheng, Nirwan Ansari and Nei Kato, "On Characterizing Peer-to-Peer Streaming Traffic," IEEE Journal on Selected Areas in Communications, vol. 31, no. 9, pp. 175-188, Sep. 2013.

URL:

http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6559965

On Characterizing Peer-to-Peer Streaming Traffic

Jie Yang, Lun Yuan, Chao Dong, Gang Cheng, Nirwan Ansari, Nei Kato

Abstract—**Extensive studies have shown that the peer-to-peer (P2P) traffic has already become the dominant traffic in the current Internet. The current P2P streaming user base is still undergoing stunning growth in China although its user scale already reached 158 million in 2010, 68% of Chinese web users. Hence, a comprehensive understanding of the P2P streaming network traffic characterization is essential to Internet Service Providers (ISPs) in terms of network planning and resource allocation. In this paper, based on the massive data collected with a passive network monitoring equipment placed in the Internet backbone, we provide an in-depth view of the current P2P streaming traffic in the current Internet of China. In particular, we statistically study the P2P streaming traffic in both wired (ADSL in this paper) and wireless (CDMA) networks, and characterize the traffic from both flow-level and packet-level aspects. Our study uncovers the significant impact of the P2P streaming traffic on the underlying network due to its unique characteristics and the bandwidth intensive nature of the corresponding applications. In addition, the result reveals the significant difference between the characterizations of the P2P streaming traffic in wired and wireless network due to their respective intrinsic environmental characteristics.**

Index Terms—**Peer-to-Peer Streaming; Packet; Flow; Traffic Characterization**

I. Introduction

WHILE Internet connectivity has reached a significant part of the world's population, the usage of peer-to-peer (P2P) applications is growing dramatically, particularly for sharing video and audio files. Different from the traditional media streaming using the model of client/server (C/S) communications, P2P streaming can utilize bandwidth resource of host nodes more adequately and provide service to peer nodes without changing the current Internet deployment [1]. Statistics showed that China already had 158 million P2P streaming users in 2010, 68% of its internet users, while this number is still growing rapidly. Typical P2P streaming applications such as PPlive, PPstream, UUSee, and QQlive become very popular, and new applications are also introduced at a rapid pace. With such a large user population, tremendous amount of traffic generated by bandwidth intensive P2P streaming applications is inevitably impacting the performance of the underlying network. It is, therefore, important to understand and characterize the traffic in terms of the end-system behavior and network impact in order to develop workload models and to provide insights into network traffic engineering, security, and capacity planning.

In the past, owing to various reasons such as security concerns and technical difficulties of collecting large volumes of real-time traffic data from high-speed Internet backbone of China, not many works, if any, have been conducted to characterizing the P2P streaming, and therefore the corresponding traffic model of P2P traffic cannot be identified. Toward this goal, we have developed hardware based network traffic probe equipment, Traffic Monitoring System (TMS), which can operate at 10Gbps of throughput, and we have been allowed to place such equipment inside the Internet backbone of a major ISP of China, from which we have collected a huge amount of data to carry out our study. More specifically, two sets of the probing equipment were placed at different locations of the Internet backbone monitoring the P2P streams originated from and terminated in the wired (ADSL) and wireless (CDMA) access network. The deployment will be elaborated in Section III. We have performed a systematic characterization of P2P streaming traffic in both wired and wireless portions of the network, and studied the impact of such traffic on the underlying network. Our unique monitored traces provide a broad view of wired and wireless network traffic, enabling more comprehensive and detailed characterizations than was possible in previous works. Detailed statistical information, e.g., packet length, packet inter-arrival time, correlation between the number of flows and peers, correlation between the number of flows and traffic volume, concurrent features of the traffic, and the number of peers per user, is presented. We believe the work is of great interest to service providers, especially those in China, because the P2P streaming traffic has increased dramatically during the last couple of years and constituted a significant portion of the total traffic observed in Internet backbones. The characterization will enable service providers to better handle such traffic through well selected traffic engineering measures, enhance network robustness, and provide better service to the customers.

The rest of the paper is organized as follows. Section II discusses related works. The data collection infrastructure and collected data are described in Section III. Traffic characteristics are investigated in Section IV. Finally, concluding remarks are presented in Section V.

II. RELATED WORKS

Traffic characterization has been extensively studied. Mandelbrot [2] introduced the concepts of self-similarity and fractional Brownian noise to analyze the communication system. Self-similarity refers to the property of an object to maintain certain characteristics when observed at different scales. These concepts have been widely adopted to analyzing and modeling network traffic, upon which many traffic models [3]-[7] have been proposed. These models well capture the long-term dependency nature of the network traffic.

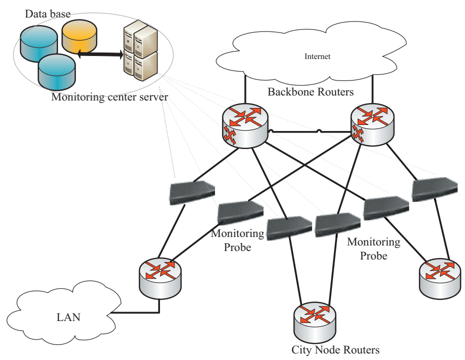
A number of works on characterizing traffics of various network environments have been reported in recent years. These prior works considered in detail the traffic volume in campus networks [8]-[10], enterprise networks [11], [12], core networks [13], and hotspot networks [14], [15]. Owing to the availability of broadband "last mile" connections, many applications such as P2P services have become prevalent, and hence there have been recent effort on characterizing traffic of such services [16] - [18].

It was found that the frequent occurrence of flash flows highly affects the performance of existing flow-based traffic monitoring systems [19]. Roughan *et al.* [20] introduced a metric for measuring backbone traffic variability based on simple but powerful traffic theory by using widely available SNMP traffic measurements. Owezarski and Larrieu [21] showed the highly oscillating nature of Internet traffic, thus explaining why it is almost impossible nowadays to guarantee a stable QoS in the Internet. Thompson *et al.* [22] revealed the characteristics of the traffic in terms of the packet size, flow duration, volume, and percentage composition by protocol and application, and patterns seen over two time scales. Maier *et al.* [23] studied a broad range of dominant characteristics of residential traffic including DSL session characteristics, network and transport-level features, prominent applications, and network path dynamics. In addition, this research group conducted a study on mobile hand-held (MHD) device usage from a network perspective [24], and found that MHD traffic is dominated by multi-media content and downloads of mobile applications.

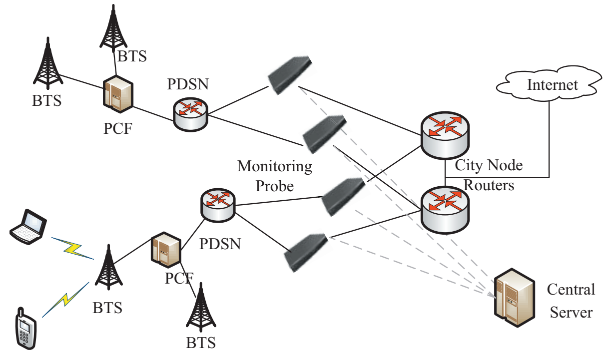
There have also been many researches on network modeling. Frequency domain techniques, e.g., wavelets coding and spectral analysis, have been applied to model network traffic [25]. Garcia-Dorado *et al.* [26] showed that while the occurrence of IP addresses and port numbers follows a Zipf distribution (as expected), the parameters of the distribution vary greatly in a spatial dimension (i.e., across individual university networks). Measurement traces of aggregated traffic from an ADSL access network were evaluated on multiple time scales in [27], and an unexpected smooth profile with less relevance of long range correlation than experienced for traffic from Ethernet LANs was observed. Zhang and Vernon [28] investigated a new technique called Bayesian-Block-Analysis (BBA) to analyze the time varying rate of events, and showed that BBA is highly accurate in identifying the rate changes in traces with exponential interevent times and known rate changes, and reasonably accurate in traces with heavier-tailed interevent times. Olivier and N. Benameur [29] found that Gamma and Weibull distributions provide excellent fits to the empirical flow inter-arrival time distributions.

With the wide spread of media streaming, characterization of both stored and live media streaming received considerable attention in the past several years. Sripanidkulchai *et al*. [30] studied the live media streams collected from a large CDN (Content Distribution Network) and found that media popularity follows a 2-mode Zipf distribution. Chesire *et al.* [31] in 2000 observed that most media streams viewed in their campus were encoded at low bit rates suitable for streaming for dial-up users, e.g., typically less than 1MB in size. Wu *et al.* [1] utilized more than 230GB of traces collected from UUSee to characterize the achievable bandwidth of streaming flows among peers in a large scale real-world P2P live streaming session. Zink *et al.* [32] characterized the nature of YouTube traffic in a large university campus network. McCreary and claffy [33] presented trends in application usage seen at the NASA Ames Internet Exchange for a period of over 10 months, and showed changes in the fraction of traffic from streaming media and online gaming, as well as an increase in traffic from new applications such as Napster and IPSEC tunneling. Brownlee and claffy [34] pointed out that streams can be classified not only by lifetime (dragonflies and tortoises) but also by size (mice and elephants), and noted that stream size and lifetime are independent dimensions.

As reviewed above, there has been a rich literature on modeling and characterizing traffics in different



(a)



(b)

Fig. 1. Deployment of TMS in (a) the ADSL access network, and (b) the CDMA mobile network.

networks. However, to our best knowledge, none has analyzed and characterized the P2P streaming traffic in China at such a large scale as will be reported next. The increasing P2P streaming traffic load on the Internet calls for an in-depth investigation of the traffic nature. Currently, the ability to managing P2P streaming effectively is becoming a critical networking issue, and thus tremendous efforts are focusing on making P2P streaming more effective. Therefore, it is essential to study the characteristics of the P2P streaming traffic in China in a realistic environment in order to serve the ever increasing size of the user base and facilitate emerging P2P streaming applications.

III. Collection and Classification of Traffic Data

A. Data collection

As mentioned above, the datasets used in this study were collected by using our self-developed Traffic Monitoring System (TMS) (This device has been placed in the production networks by several ISPs for traffic monitoring purpose.). Two sets of mirrored packets have been collected from ADSL and CDMA networks from a large Chinese ISP. Overall, the ISP has roughly 4.5 million ADSL users, and more than 3.02 million CDMA users including about 0.3 million 3G users. For the ADSL data, a monitoring system connecting to the 16x10G POS export ADSL links between the backbone routers and the core routers of a large metropolitan area network in Southern China is deployed, as shown in Fig. 1(a). Similarly, for the CDMA data, a monitoring system is deployed between the Packet Data Serving Nodes (PDSNs) and the backbone routers in the same city, as shown in Fig. 1(b). The access bandwidth of a monitored link varies between 1024 and 512Kbps for the downstream, and between 4096 and 512Kbps for the upstream, while in the CDMA network, the CDMA 1X can actually support 150Kbps and the 3G users can support 3100Kbps.

In order to characterize the P2P streaming traffic accurately, it is crucial to identify the P2P streaming traffic from the aggregate traffic. For this purpose, we first classify and capture the raw packets via TMS based Deep Packet Inspection (DPI) probes [35], which are proven to be capable of achieving high classification accuracy. Furthermore, to ensure that the classification result is reliable, we use a self-designed Network Traffic Analysis and Classification System (TACS) to further identify and classify the mirrored packets which are already labeled by TMS. TACS combines automatically mining signatures technology [36], attribute selection technology [37], and machine learning [38], and adopts both DPI [35] and Deep Flow Inspection (DFI) [39] to identify and classify the offline traces. A high classification accuracy can be achieved by synergizing the hardware speed of DPI probes and the intelligent software of TACS. In our previous work [40], the traffic classification accuracy can be higher than 96.88%. We shall next characterize the P2P streaming traffic classified by and extracted from TACS.

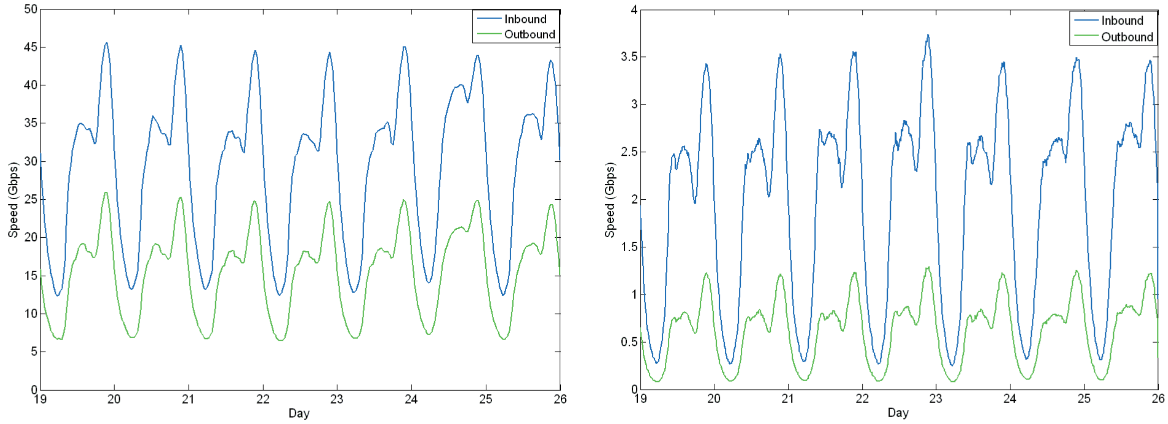
Note that in this paper, we group the packets into different flows by the 5-tuple {IP source address, IP destination address, source port number, destination port number, transport protocol}, i.e., a 5-tuple flow is a sequence of packets that share the same 5-tuple [26] during a certain period (e.g., 64s).

B. Data Reduction

Owing to the extremely large volume of collected data and limited computational resources, e.g., memory and CPU, we have to select a subset of the data for analysis.

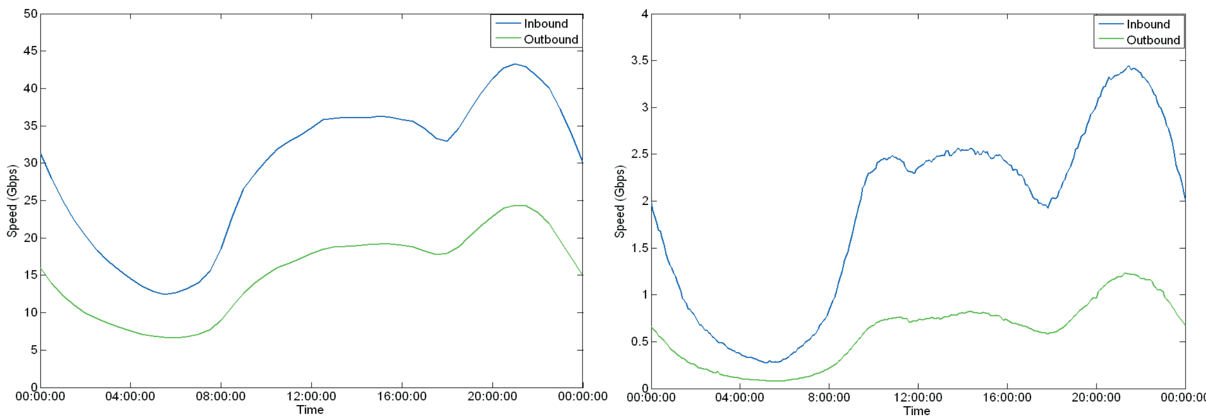
Weekly Traffic Pattern

Our monitoring system has one key advantage—it can sniff packets generated from thousands of ADSL lines and CDMA users from different urban areas, connected



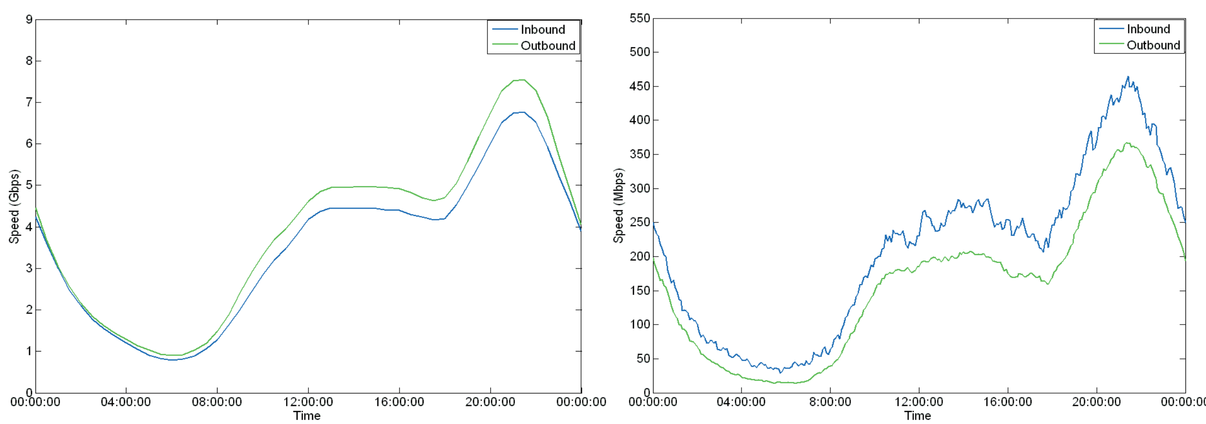
(a) (b)

Fig. 2. Inbound and outbound traffic in a week: (a) ADSL data, and (b) CDMA data.



(a) (b)

Fig. 3. Inbound and outbound traffic on September 21, 2011: (a) ADSL data, and (b) CDMA data.



(a) (b)

Fig. 4. P2P streaming media traffic collected on Sep 21, 2011: (a) ADSL data, and (b) CDMA data.

to city node routers. Fig. 2 provides an overview of the ADSL and CDMA data traces covering the period of a week (Sep. 19, 2011 to Sep .26, 2011). Fig. 2(a) and 2(b) show the inbound and outbound traffics of ADSL and CDMA networks, respectively. Note that the network traffic consistently follows a clear daily pattern in both networks, and the daily traffic patterns are similar to each other. As such, we pick the data collected on Wednesday (September 21, 2011) for our analysis.

TABLE I
DESCRIPTION OF MIRRORED PACKETS TRACES

Mirrored Traces	Collection Date	Duration	Direction	Users	Flows	Traffic Volume
ADSL trace 1(AT1)	2011-9-21 21:00	60min	Bi-dir	256	789,518	27.12GB
ADSL trace 2(AT2)	2011-9-21 21:00	60min	Bi-dir	256	615,916	26.91GB
ADSL trace 3(AT3)	2011-9-21 21:00	60min	Bi-dir	256	698,326	28.34GB
ADSL trace 4(AT4)	2011-9-21 21:00	60min	Bi-dir	256	781,264	28.83GB
CDMA trace 1(CT1)	2011-9-21 21:00	60min	Bi-dir	14,703	893,066	8.51GB
CDMA trace 2(CT2)	2011-9-21 21:00	60min	Bi-dir	13,547	938,002	8.30GB
CDMA trace 3(CT3)	2011-9-21 21:00	60min	Bi-dir	13,540	887,358	8.51GB
CDMA trace 4(CT4)	2011-9-21 21:00	60min	Bi-dir	10,312	659,249	8.16GB

TABLE II
TRAFFIC VOLUME AND NUMBER OF FLOWS PROPORTION OF ADSL DATA FOR VARIOUS APPLICATIONS

Application	Traffic Volume (%)				Number of Flows (%)			
	AT1	*AT2*	*AT3*	*AT4*	*AT1*	*AT2*	*AT3*	*AT4*
P2PDownload	31.38%	34.56%	44.01%	28.48%	32.42%	30.62%	34.45%	35.14%
P2PStream	24.68%	27.03%	18.81%	27.59%	8.92%	10.28%	8.32%	9.53%
VideoStream	12.82%	12.31%	13.73%	14.21%	3.47%	4.31%	4.00%	3.11%
Web	15.89%	11.51%	11.61%	15.66%	20.58%	22.09%	19.54%	21.14%
IM	0.39%	0.29%	0.55%	0.80%	1.54%	1.41%	1.45%	1.82%
Other	8.88%	8.54%	6.69%	7.80%	1.91%	2.18%	3.43%	2.02%
Unknown	5.96%	5.76%	4.60%	5.46%	31.16%	29.11%	28.81%	27.24%

TABLE III
TRAFFIC VOLUME AND NUMBER OF FLOWS PROPORTION OF CDMA DATA FOR VARIOUS APPLICATIONS

Application	Traffic Volume (%)				Number of Flows (%)			
	CT1	*CT2*	*CT3*	*CT4*	*CT1*	*CT2*	*CT3*	*CT4*
P2PDownload	22.27%	15.01%	26.24%	24.71%	4.44%	2.87%	4.08%	4.21%
P2PStream	9.17%	18.74%	10.49%	14.79%	1.52%	11.06%	6.40%	3.08%
VideoStream	9.82%	8.94%	10.57%	16.25%	1.59%	1.64%	2.10%	2.56%
Web	44.06%	44.68%	37.11%	32.89%	62.02%	54.88%	56.92%	57.80%
IM	2.12%	1.83%	2.55%	1.80%	10.15%	8.17%	9.76%	9.94%
Other	7.11%	5.37%	7.57%	4.99%	9.08%	10.64%	7.29%	8.84%
Unknown	5.45%	5.43%	5.47%	4.57%	11.20%	10.74%	13.45%	13.57%

Daily Traffic Pattern and Data Selection

Fig. 3 provides a detailed view of the inbound and outbound aggregate traffic volume of two networks on September 21, 2011. The green and blue curves in Fig. 3(a) show that in one day, the ADSL traffic volume naturally reaches its peak between 9:00 PM and 10:00 PM, and bottoms between 5:00 AM and 6:00 AM. Then, it climbs up to the peak around noon probably because of the noon break. When people begin to get off work around 6:00 PM, it bottoms again. From the curves in Fig. 3(b) we can see that the shape of the CDMA traffic volume is similar to that of the ADSL traffic volume in one day. There are still some differences between them. The CDMA traffic volume also climbs to the high peak around 10 AM, but there is a valley point around noon when many people are taking the lunch break. Hence, we can conclude that there is strong correlation between traffic pattern and user daily activities.

Since our purpose is to study the P2P streaming traffic, we also segregate the P2P streaming traffic as described in Section III(A) and study its traffic pattern. Fig. 4 shows the P2P streaming media traffic volume on Sep 21, 2011. Note that like the aggregate traffic, the P2P streaming media traffic patterns in both networks are similar to each other.

The purpose of this study is to gain insight into the P2P streaming traffic and understand its impact on the underlying network so that the network can be optimized. Toward this end, we choose to study the traffic data collected from 9 p.m. to 10 p.m. as this is the most

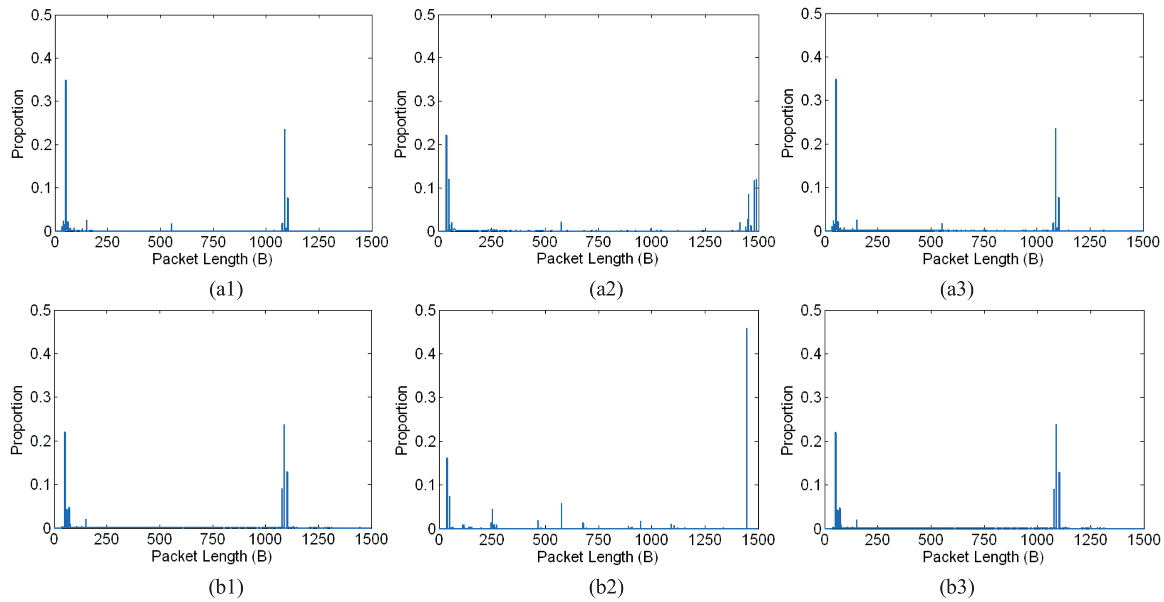


Fig. 5. The length distribution of packets in ADSL and CDMA networks.

congested period for the network in a day and also presents the biggest impact of the P2P streaming traffic considering its traffic volume. Since we have placed multiple network traffic monitoring devices inside the network, we will have multiple sets of data between 9 p.m. and 10 p.m. for investigation. In this paper, we choose four sets of data from the respective ADSL and CDMA networks.

C. *Detailed information of traces*

The detailed information of representative traces collected on Sep 21, 2011 for our analysis is shown in Table I. In addition, we have collected four packet traces in both ADSL access network and CDMA mobile network. Each ADSL packets trace was collected from a C class network of the same Chinese carrier network. These four ADSL packets traces cover 1024 users in total. The CDMA packets traces were collected by four network monitoring probes deployed in different nodes in the same city. The total number of users covered by these four CDMA packets traces is over fifty thousand. In the rest of the paper, we denote the ADSL traces as AT1 to AT4, and CDMA traces as CT1 to CT4.

Based on the identified results from TMS and TACS, we classify the ADSL data and CDMA data into several categories. Table II and Table III illustrate the traffic distribution of five major traffic contributors (applications), which are P2PDownload, P2PStream, VideoStream, Web and Instant Message (IM), and the 'other' category which includes other services such as FTP, Email, Game, VoIP, Telnet, etc. In addition, the unknown traffic, mainly VPN and encrypted traffic, is also occupying a portion, around 5% of the total traffic. Further analyzing the unknown traffic, over 80% of TCP flows contain less than 7 packets (i.e., they hardly represent complete sessions), and nearly 90% of UDP flows containing less than 16 packets cannot be identified and classified appropriately.

According to Table II, P2P applications contribute the majority of the traffic in the ADSL access network. Moreover, the P2P streaming traffic volume makes up around 25%. In CDMA mobile network, as shown in Table III, web and IM are the major players, mainly because the majority of users use handsets accessing the CDMA network to browse the website and chat with friends. P2P applications also account for a certain percentage of the traffic volume for users accessing the network via 3G network cards in their PCs.

Since P2P streaming applications exhibit similar characteristics among themselves (i.e., traffic generated by one P2P streaming application is similar to that by another), we focus on characterizing the aggregate P2P streaming traffic in the following context. Since the traces, collected from both the ADSL access network and CDMA mobile network, are rather recent and voluminous, the conclusions drawn from our analysis of these recently acquired data can reliably reveal the traffic characteristics of the current network.

IV. Detailed Traffic Analysis

In this section, we will characterize the P2P streaming traffic in terms of packet length distribution, packet inter-arrival time, correlation coefficient distribution,

TABLE IV

PROPORTION OF TOP FIVE PACKET LENGTHS

TCP				UDP			
ADSL		*CDMA*		*ADSL*		*CDMA*	
Packet length (B)	***Proportion (%)***	***Packet length (B)***	***Proportion (%)***	***Packet length (B)***	***Proportion (%)***	***Packet length (B)***	***Proportion (%)***
40	22.84%	1448	45.85%	53	35.01%	1089	23.86%
1480	12.54%	40	16.19%	1089	23.89%	53	22.14%
52	11.97%	52	7.38%	1104	7.46%	1104	12.92%
1492	10.66%	576	5.83%	41	2.41%	1080	9.08%
1454	6.75%	253	4.53%	151	2.37%	1103	5.99%
other	35.24%	other	20.22%	other	28.86%	other	26.01%

TABLE V

PROPORTIONS OF TCP PACKETS AND UDP PACKETS

Protocol	**Proportion (%)**				**Proportion (%)**			
	AT1	***AT2***	***AT3***	***AT4***	***CT1***	***CT2***	***CT3***	***CT4***
TCP	0.18%	0.05%	0.32%	0.31%	0.36%	0.57%	0.24%	0.09%
UDP	99.82%	99.95%	99.68%	99.69%	99.64%	99.43%	99.76%	99.91%

average concurrent connections, and peer analysis.

A. Distribution of packet length

The packet length of an IP packet is defined as the length of the sum of bytes of the IP header, the TCP or UDP header, and the payload of the packet. The distribution of packet length is shown in Fig. 5.

Fig. 5, (a1), (a2), and (a3) present the length distribution of ADSL data while (b1), (b2), and (b3) present the length distribution of CDMA data. Moreover, (a1) and (b1) are the length distribution of all packets, (a2) and (b2) are the length distribution of TCP packets, and (a3) and (b3) present the length distribution of UDP packets, respectively.

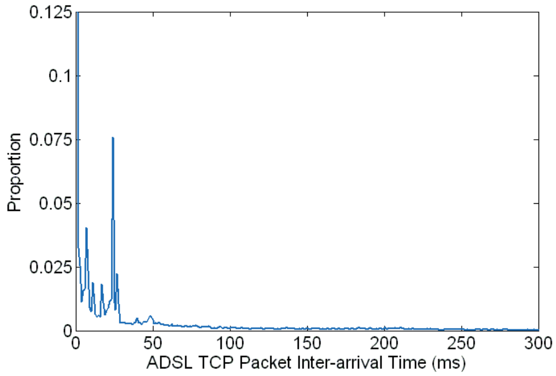
It can be observed that the majority of TCP packets are either less than 100 bytes or larger than 1400 bytes. The UDP packet length distribution is quite similar to that of TCP packets, i.e., multi-modal, implying that only certain packet lengths are adopted by P2P streaming services for data transmission and communications. Table IV lists the top five (percentile wise) of the packet length observed from the collected data. So, we may conclude that P2P streaming applications use specified packet lengths for data transmission. This feature is very valuable in identifying and classifying P2P streaming traffic.

Most importantly, Fig. 5 shows that the length distribution of all packets and the length distribution of UDP packets are almost the same. This phenomenon is owing to the fact that P2P streaming applications only use UDP protocol for real-time data transmission. In some P2P streaming applications, clients use TCP to login and download advertising and program lists, and thus this traffic is also classified as P2P streaming media traffic by TMS. The proportions of TCP packets and UDP packets are shown in Table V. Apparently, most traffic, constituted by the payload, is carried by UDP.

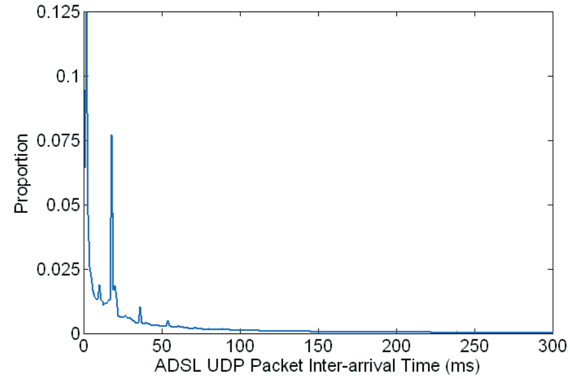
B. Distribution of packet inter-arrival time

In this section, we focus on the distribution of packet inter-arrival time of individual flows, in which the packet inter-arrival time is defined as the difference between the arrival time of the current packet and that of the previous packet.

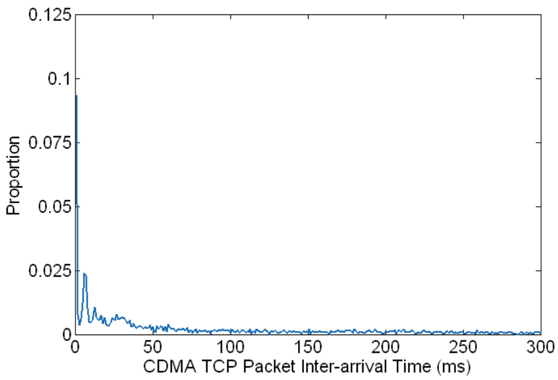
A five-tuple flow [26] is defined as a sequence of packets that share the same 5-tuple during a certain period. If the period is defined as 64 seconds, the upper bound of the packet inter-arrival time is also 64 seconds. Using milliseconds as the unit of packet inter-arrival time, the x-axis is ranged from 0 to 64000. The analysis of the acquired data indicates that the portion of packets whose inter-arrival time greater than one second is very small. The probability of having packet inter-arrival time less than 300 milliseconds is more than 80%. Hence, we only focus on the 0~300 milliseconds range. Fig. 6 illustrates the distribution of packet inter-arrival time in both ADSL access network and CDMA mobile network, respectively, in which the x-axis is the inter-arrival time in millisecond. The distribution of ADSL TCP packet inter-arrival time is shown in Fig. 6(a), that of ADSL UDP packet inter-arrival time in Fig. 6(b), that of CDMA TCP packet inter-arrival time in Fig. 6(c), and that of CDMA UDP packet inter-arrival time in Fig. 6(d).



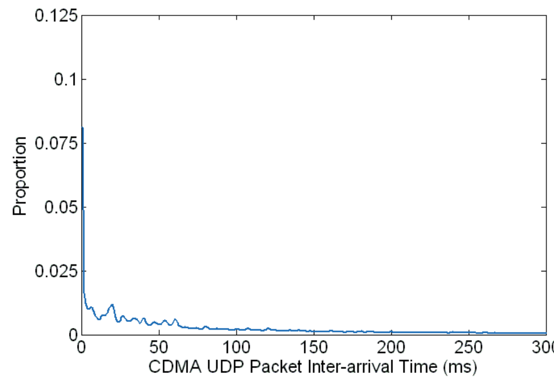
(a)



(b)



(c)



(d)

Fig. 6. Distribution of the proportion of packet inter-arrival time.

The average packet inter-arrival times of TCP and UDP packets in ADSL traces are 374.24 and 318.06, respectively, while those in CDMA traces are 796.62 and 677.66, respectively, which are twice those of the packet inter-arrival times in ADSL traces. Furthermore, the standard deviations of the packet inter-arrival times of TCP packets and UDP packets in ADSL trace are 2274.36 and 1867.77, respectively, and those in CDMA traces are 2669.01 and 3278.41, respectively. These figures show that the fluctuation of the inter-arrival time of packets in the ADSL traces is smaller than that in the CDMA traces. The inter-arrival time of UDP packets in the ADSL traces fluctuates less than that of TCP packets, while the phenomenon is just the opposite in the CDMA traces.

Fig. 6 also shows that the distribution of the CDMA packet inter-arrival time is more dispersive than that of the ADSL packet inter-arrival time in the sense that the probability distribution of ADSL packet inter-arrival time exhibits several distinct peaks before 50 milliseconds, and they occupy a large portion. For TCP packet inter-arrival time, there is a peak at 24 milliseconds which is the median value of the TCP packet inter-arrival time, while the curve of UDP packet inter-arrival time reaches its peak at 18 milliseconds, which is the median value. These characteristics imply that on average a P2P streaming service through CDMA has only one half of the bandwidth of that through ADSL.

C. *Correlations analysis*

In general, an elephant flow refers to a large flow while a mice flow a small flow; yet, there is no concrete agreement on the size of a flow to be considered a large flow, and vice versa [41]. In this paper, we define an elephant flow as the flow which contains more than 64 packets, and otherwise it is a mice flow. Extensive studies have shown that an elephant flow contributes more traffic than a mice flow does [42].

As described in Section III, we have collected four packet traces in both ADSL access network and CDMA mobile network, each of which covered one hour. Hence, we define N intervals, each of which covers $3600/N$ seconds. The number of flows, peers, elephant flows, and mice flows of P2P streaming service and traffic volume are calculated for each interval. In this paper, we set N as 10, and therefore we have 360 sets of statistics. For ADSL data, since we have four traces, we define a vector $\{x_i\}$, i=1, 2, 3, and 4, in each interval as the number of P2P streaming flows in the trace. We also

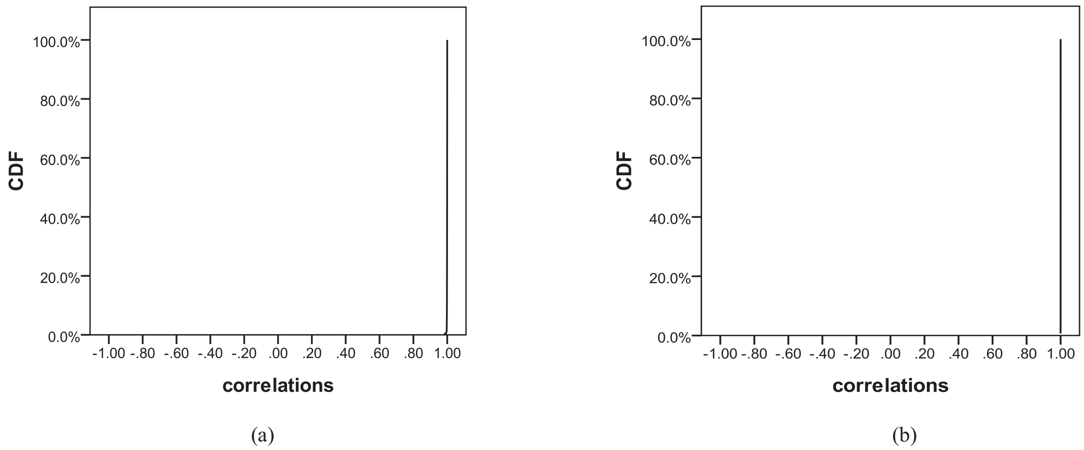


Fig. 7. The distribution of the correlation coefficient between the number of P2P streaming flows and the number of peers of (a) ADSL data, and (b) CDMA data, respectively.

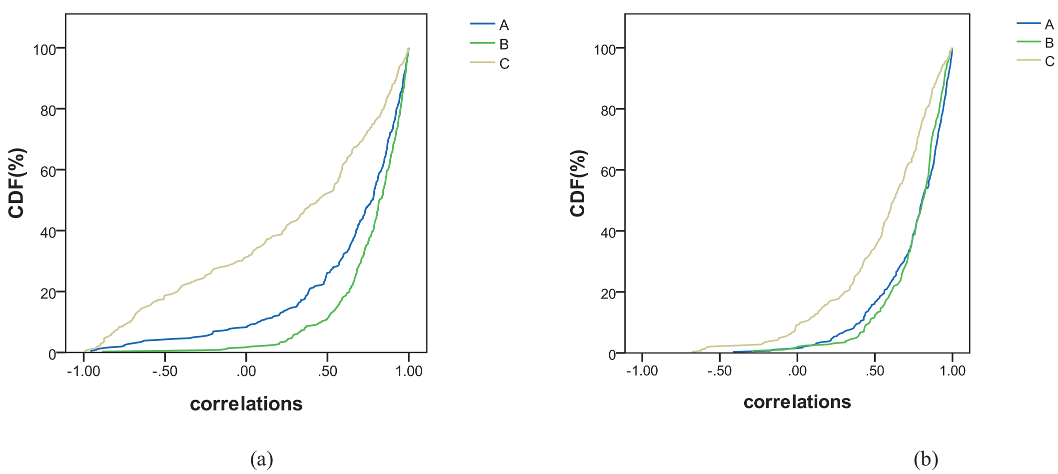


Fig. 8. Distribution of correlation between the number of flows and traffic volume of (a) ADSL data, and (b) CDMA data, respectively.

define the number of P2P streaming service peers as vector $\{y_i\}$, i=1, 2, 3, and 4, in each interval in the trace. Hence, there are 360 sets of X and Y vectors in total.

The correlation coefficient between the number of flows and the number of peers is calculated as:

$$\rho_{x,y} = \frac{\sum_{i=1,2,3,4}\left((x_i - \bar{X})(y_i - \bar{Y})\right)}{\sqrt{\sum_{i=1,2,3,4}(x_i - \bar{X})^2 \sum_{i=1,2,3,4}(y_i - \bar{Y})^2}}$$

where $\bar{X}$ is the mean of X and $\bar{Y}$ is the mean of Y. The correlation coefficients for all X and Y in each interval are shown in Fig. 7(a). Similarly, the cumulative probability of the correlation coefficient of CDMA data is illustrated in Fig. 7(b).

It can be seen from Fig. 7 that the number of flows is highly correlated to the number of peers; this is intuitive as the destination IP address is part of the five-tuple, and thus the more peers, the more number of flows. For P2P applications, it is very likely that more than one connection is set up between two peers. Hence, it is interesting to find out the distribution of the number of flows between an IP pair, which will be presented later in this paper.

In order to study the relationship between the number of flows and traffic volume, we calculate the correlation coefficient using three different definitions of X and Y, and also draw three cumulative probability distributions in the same figure. Below are the three definitions:

A) X is defined as the number of P2P streaming flows in each interval.

B) X is defined as the number of P2P streaming elephant flows in each interval.

C) X is defined as the number of P2P streaming mice flows in each interval.

Meanwhile, Y is defined as the traffic volume of P2P streaming service in each interval. Hence, we can calculate the correlation coefficient between X and Y. Fig. 8 illustrates the three cumulative probability curves.

From Fig. 8, it can be seen that the numbers of elephant flows are much more correlated to the traffic

TABLE VI
PROPORTION OF ELEPHANT AND MICE FLOWS IN P2P STREAMING MEDIA

	Flow (%)							
	AT1	*AT2*	*AT3*	*AT4*	*CT1*	*CT2*	*CT3*	*CT4*
Elephant flows	10.08%	12.14%	11.38%	13.43%	19.20%	4.87%	4.94%	18.28%
Mice flows	89.92%	87.86%	88.62%	86.57%	80.80%	95.13%	95.06%	81.72%

TABLE VII
STATISTIC OF CONNECTIONS AND PEERS

Statistics	Formula
Average Concurrent Connections	Connections/number of users
Average Concurrent Connections of Elephant Flows	Connections of elephant flows/number of users
Average Concurrent Connections of Mice Flows	Connections of mice flows/number of users
Average Peers	Number of peers/number of users

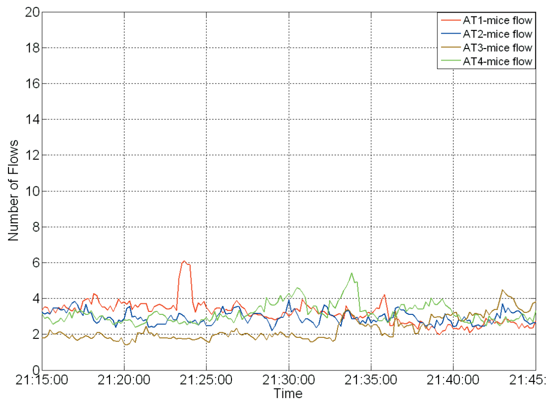


Fig. 9. Distribution of average number of concurrent connections of ADSL mice flows

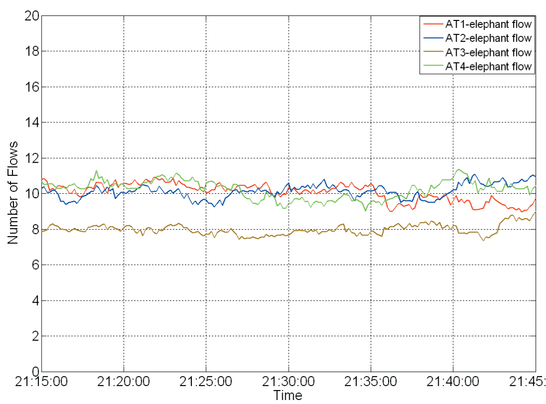


Fig. 10. Distribution of average number of concurrent connections of ADSL elephant flows

volumes than those of mice flows do (curve B is under curve C). This suggests that elephant flows contribute the majority of the traffic volume. In contrast, the numbers of mice flows show almost no correlation with the traffic volumes in the ADSL network because the corresponding CDF curve (C) is nearly linear over the range of -1 to 1, suggesting mice flows are mainly control flows whose numbers in average are not correlated to the traffic volumes. However, the numbers of mice flows do show positive correlation with the traffic volumes in the CDMA network because the corresponding CDF curve (C) mainly covers the range of 0 to 1. This is mainly owing to the fact that the CDMA network environment is relatively unstable, and therefore it is conceivable that the portion of mice flows in the CDMA network is much larger than that in the ADSL network. Mice flows in the CDMA network also contribute a significant portion of traffic volume even though the number of packets in each mice flow is small. The proportion of elephant flows and mice flows in both ADSL access network and CDMA mobile network is illustrated in Table VI.

D. Concurrent feature analysis

Each P2P streaming media client keeps an online peer list for the selected channel, and then tries to connect to the peers on the list [43]. Hence, one peer may connect many other peers at the same time, and then try to find active peers to transfer media data. In this section, we analyze this concurrent feature of P2P streaming media service. Taking 10 seconds as an interval, we count the number of users, concurrent connections, concurrent connections of elephant flows and mice flows, and numbers of peers of all the users in every interval. By doing so, we are able to calculate the average concurrent connections, average concurrent connections of elephant flows and mice flows, and average number of peers to which a user is connected, as illustrated in Table VII.

To avoid the impact of the outliers of the collected data on the analysis, we focus on analyzing the data in the middle half of the hour from 21:15 to 21:45. Fig. 9 shows that the average number of mice flows in the ADSL network in ten seconds is between 2 and 4, while the average number of elephant flows, as shown in

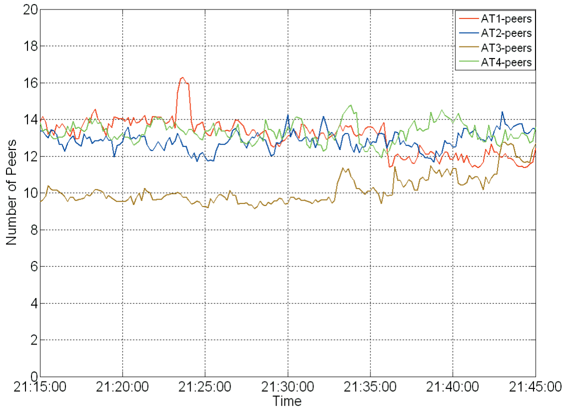


Fig. 11. Distribution of average number of peers per user in ADSL

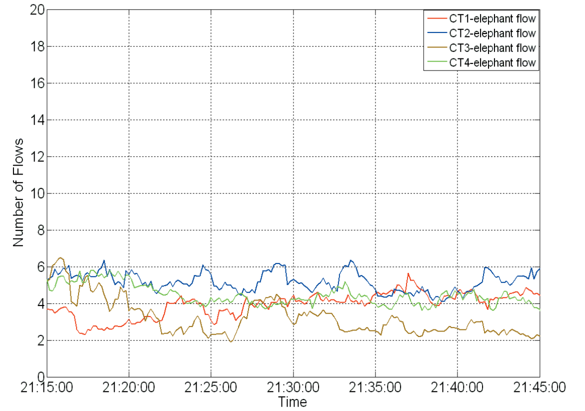


Fig. 13. Distribution of average number of concurrent connections of CDMA elephant flows

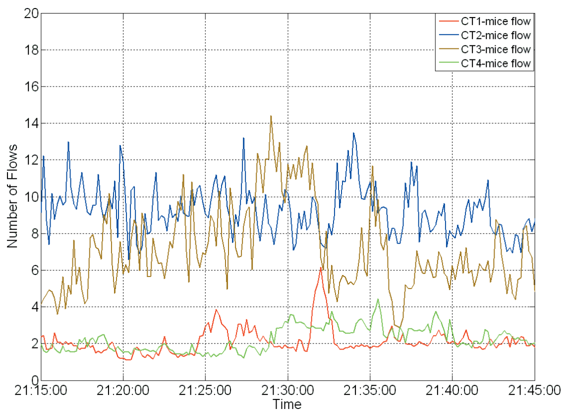


Fig. 12. Distribution of average number of concurrent connections of CDMA mice flows

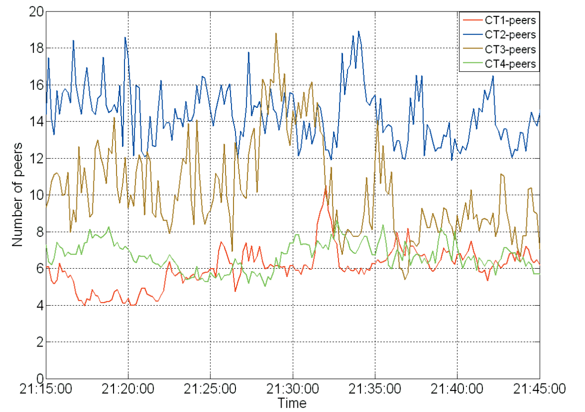


Fig. 14. Distribution of the average number of peers in CDMA.

Fig. 10, is around 10, except ADSL Trace 3 (AT3), which is around 8. From these two figures, we readily see that there are more elephant flows than mice flows for each user in each interval. It is, however, contrary to the overall distribution of the elephant and mice flows for the entire ADSL traces, shown in Table VI. This contrast suggests that an elephant flow exists across many more intervals than a mice flow does.

Therefore, although the total number of the elephant flows is much less than the total number of the mice flows, the numbers of existing connections of the two types of flows in the short intervals may exhibit an inverse phenomenon.

Fig. 11 shows the average number of peers per user during the half-hour period. Note the similarity between Fig. 11 and Fig. 9. This implies that the number of mice flows generated by a user is closely related to the number of peers the user holds. The average number of peers per user for most ADSL traces (except AT3 which is approximately 10) is approximately 13.

Fig. 12 to Fig. 14 present the analysis of the CDMA data. Unlike the ADSL data, differences among the four traces are more pronounced. The numbers of mice flows among the four traces shown in Fig. 9 do not deviate much as compared to those in Fig. 12. The average number of concurrent elephant flows in Fig. 10 are higher than 8, while the average number of concurrent elephant flows are lower than 6 in Fig. 13. In comparing Fig. 12 and Fig. 13, one finds that the numbers of elephant flows are no longer much larger than those of mice flows as in the situation in ADSL. In some traces, the numbers of elephant flows are even lower than those of mice flows. This is probably attributed to the error-prone wireless channel of the CDMA mobile network that results in a large number of retransmissions or probe flows.

As shown in Fig. 14, the average number of peers per user differs largely among the different traces, ranging from 5 in CT1 to 15 in CT2.

From the concurrent feature analysis, as compared to the ADSL access, the CDMA network is much more dynamic in terms of the average number of concurrent connections and average number of peers per user. In addition, more mice flows exist in the CDMA network,

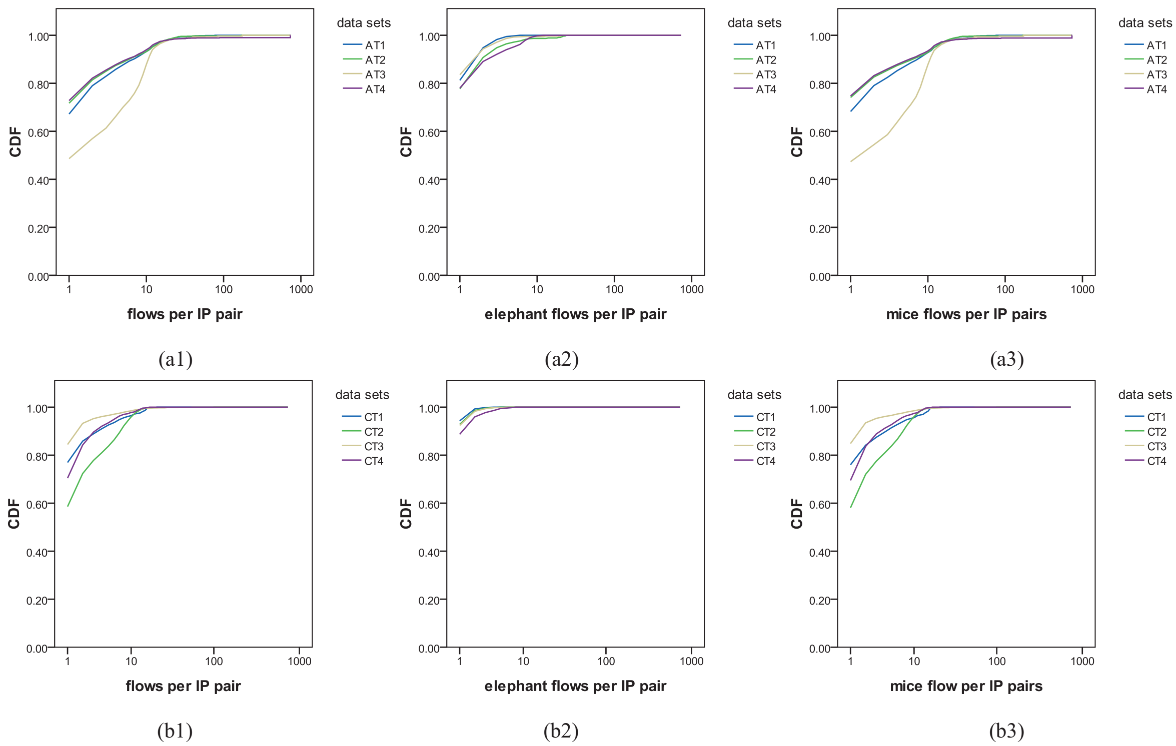


Fig. 15. Distribution of the number of flows (all flows, elephant flows, mice flows) between two peers.

which might be caused by the instability of the mobile channels.

E. Peer analysis

In this section, we analyze the number of flows between a pair of peers. We know from Section IV(C) that the number of flows is highly related to the number of peers. However, this does not mean that the number of flows equals the number of peers, and there may be more than one flow between two peers. Hence, we analyze the distribution of the number of flows between every pair of peers.

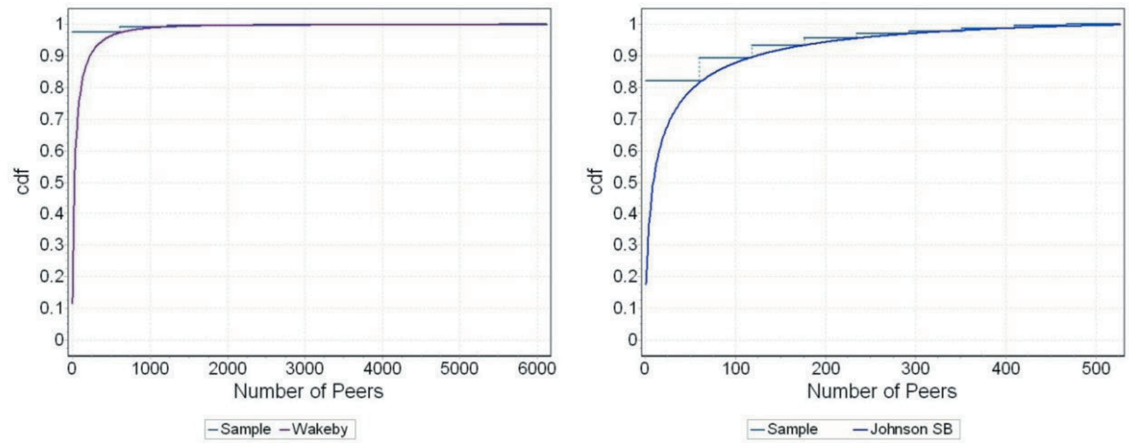
First, we extract P2P streaming flows from all traces, and then group them by the source IP address and destination IP address; the number of flows in each group is the number of flows between two peers. Moreover, we categorize the flows between two peers into elephant flows and mice flows. Fig. 15(a1), 15(a2), and 15(a3) respectively show the distribution of the number of flows, the distribution of the number of elephant flows, and the distribution of the number of mice flows between all possible pairs of peers in ADSL traces. Similarly, Fig. 15(b1), 15(b2), and 15(b3) respectively illustrate the distributions of the number of flows, the number of elephant flows, and the number of mice flows between all possible pairs of peers in CDMA traces .

It can be seen from Fig. 15(a1) and 15(b1) that over 50% pairs of peers have only one flow. The proportion of pairs of peers which have less than 15 flows is great than 95%. For IP pairs with elephant flows, more than 80% pairs from ADSL traces have only one elephant flow, while it is about 90% in CDMA data. Fig. 15(a3) and 15(b3) show the number of mice flows between two peers, and over 95% pairs of peers create less than 15 mice flows.

The distribution of the number of peers from the perspective of the user is also analyzed. First of all, we extract and identify the elephant flows from the traces. According to each user, we enlist the elephant flows into different groups. Hence, each user has a group of elephant flows. Finally, we calculate the number of distinct destination IP addresses as the number of peers in each group. Hence, the distribution of the number of peers per user is obtained. Fig. 16(a) shows the cumulative distribution of the number of peers per user in the ADSL access network, and Fig. 16(b) that in the CDMA mobile network. Note that a peer is one with the destination host with which the elephant flow has established. Moreover, the CDF (cumulative distribution function) of the number of peers per user of ADSL data follows a Wakeby distribution, while the CDF of the number of peers per user of CDMA data follows a Johnson SB distribution.

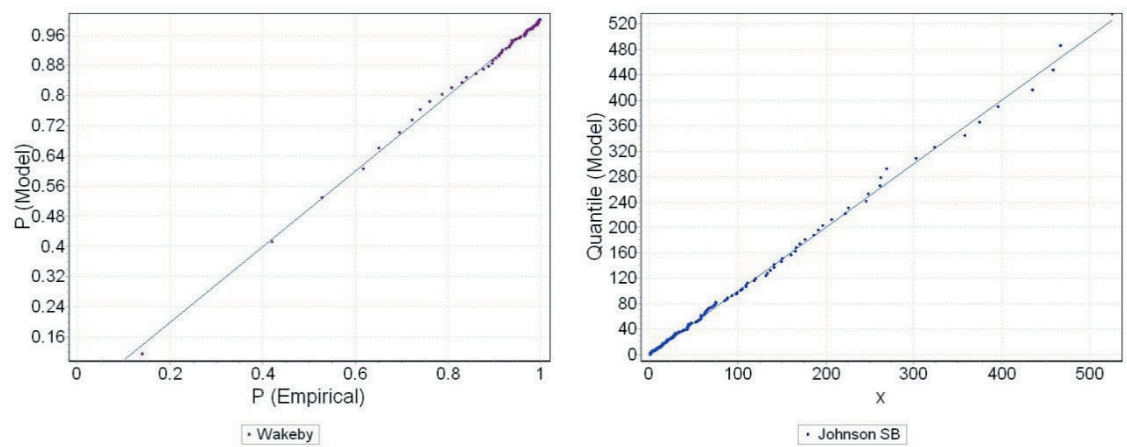
The Wakeby distribution is defined by the quantile function:

$$x(F) = \xi + \frac{\alpha}{\beta}\left(1-(1-F)^{\beta}\right) - \frac{\gamma}{\delta}\left(1-(1-F)^{-\delta}\right)$$



(a) (b)

Fig. 16. CDF of the number of peers per user of (a) ADSL data follows the Wakeby distribution function, and that of (b) CDMA data follows the Johnson SB distribution function.



(a) (b)

Fig. 17. Curve fitting: (a) P-P plot of fitting curve of the Wakeby distribution, and (b) Q-Q plot of fitting curve of the Johnson SB distribution

where $F = F(x) = P(X \leq x)$, and X, which is a random variable, stands for the number of peers per user in this application. After fitting, the parameters are: $\alpha = -68.046$, $\beta = 0.75523$, $\gamma = 71.063$, $\delta = 0.4764$, $\xi = 0$.

The probability density function of Johnson SB distribution is defined as:

$$f(x) = \frac{\delta}{\lambda\sqrt{2\pi}z(1-z)} exp\left(-\frac{1}{2}\left(\gamma + \delta\, ln\left(\frac{z}{1-z}\right)\right)^2\right)$$

where $z \equiv \frac{x-\xi}{\lambda}$. The fitting parameters are: $\gamma = 1.8846$, $\delta = 0.45388$, $\lambda = 587.11, \xi = -0.18362$.

Fig. 17 shows the P-P plot of the fitting curve of the Wakeby distribution and the Q-Q plot of the fitting curve of the Johnson SB distribution. In statistics, a P-P plot (probability-probability plot or percent-percent plot) is a probability plot for assessing how closely two data sets agree by plotting the two cumulative distribution functions against each other, while a Q-Q plot ("Q" stands for quantile) is a probability plot, which is a graphical method for comparing two probability distributions by plotting their quantiles against each other (from www.wikipedia.org). From Fig. 17, we can obtain that the CDF of the number of peers per user of ADSL data fits well by the Wakeby distribution function, and that of CDMA data fits well by the Johnson SB distribution function.

Hence, the Wakeby distribution and Johnson SB distribution can be useful in estimating a peer-to-peer streaming user's number of peers of elephant flows. However, we can hardly find a distribution function to approach the CDF of the number of peers of mice flows per user.

V. CONCLUSIONS

This paper presents insights of peer-to-peer streaming traffic collected from ADSL and CDMA networks in China. From our analysis, we have found that over 99% P2P streaming traffic volume is contributed by the UDP protocol. As P2P streaming applications almost always send packets with fixed length, the majority of TCP packets are either less than 100 bytes or larger than 1400 bytes, while the majority of UDP packets are either less than 100 bytes or around 1100 bytes. The distributions of ADSL packet inter-arrival time have peaks at around 18 milliseconds and 24 milliseconds. The number of flows is highly related to the number of peers, and the number of elephant flows is proportional to the total traffic volume. For ADSL traces, the average concurrent connections of elephant flows are relatively higher than those of the mice flows, while no clear distinction between the elephant and mice flows is observed in terms of the average number of concurrent connections from CDMA traces. Over 50% pairs of peers have only one flow and over 95% pairs of peers have less than 15 flows. At last, we prove that the CDF of the number of peers per user of ADSL elephant flows data follows a Wakeby distribution, while the CDF of number of peers per user of CDMA elephant flows data follows a Johnson SB distribution. This is useful to estimate a P2P streaming user's numbers of peers with elephant flows for content delivery. However, we have not found a suitable distribution to fit the number of peers of mice flows.

ACKNOWLEDGMENT

This work was supported in part by National Natural Science Foundation of China (61072061), 111 Project of China (B08004), and the Fundamental Research Funds for the Central Universities (2009RC0122).

REFERENCES

[1] C. Wu, B. Li and S. Zhao, "Characterizing Peer-to-Peer Streaming Flows," *IEEE Journal on Selected Area in Communications*, vol. 25, no. 9, pp.1612-1626, Dec.2007.

[2] B. Mandelbrot, "Self-Similar Error Clusters in Communication Systems and the Concept of Conditional Stationarity," *IEEE Trans. Communication Technology*, vol.13, no. 1, pp. 71-90, Mar.1965.

[3] R. Addie, M. Zukerman, and T. Neame, "Fractal traffic: measurements, modeling and performance evaluation," *in Proc. Fourteenth Annual Joint Conference of the IEEE Computer and Communications Societies*, vol.3, pp. 997-984, Ap. 1995.

[4] M. Garrett and W. Willinger, "Analysis, modeling and generation of selfsimilar VBR video traffic," *in Proc. ACM Sigcomm*, London, UK, pp. 269-280, 1994.

[5] L. Kaplan and C. Kuo, "Extending self-similarity for fractional browninan motion," *IEEE Trans. Signal Proc*., vol. 42, no.12, pp. 3526-350, Dec. 1994.

[6] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of ethernet traffic (extended version)," *IEEE/ACM Trans. Networking,* vol. 2, no. 1, pp. 1-15, Feb. 1994.

[7] N. Ansari, H. Liu, Y.Q. Shi and H. Zhao, "On Modeling MPEG Video Traffics, " *IEEE Trans. on Broadcasting*, Vol. 48, No. 4, pp. 337-347, December 2002.

[8] J. Erman, A. Mahanti, M. Arlitt, and C. Williamson. "Identifying and Discriminating Between Web and Peer-to-Peer Traffic in the Network Core," *in Proc. the 16th international conference on World Wide Web*, pp.883-892, May 2007.

[9] J. Zhang, J. Yang, C. An, and J. Wang. "Traffic measurement and analysis of TUNET," *in Proc. International Conference on Cyberworlds*, pp. 330-334, Nov.2005.

[10] D. Rossi, M. Mellia, and M. Meo, "A Detailed Measurement of Skype Network Traffic," *in Proc. 7th International Worshop on P2P Systems (IPTPS)*, Tampa Bay, FL, USA, pp.12-12, Feb. 2008

[11] T. Karagiannis, K. Papagiannaki, and M. Faloutsos, "BLINC: multilevel traffic classification in the dark," *in Proc. ACM SIGCOMM 2005*, Philadelphia, PA, USA, pp. 229-240, Aug.2005.

[12] P. Pang, P. Allman, P. Bennett, P. Lee, P. Paxson, and B. Tierney, "A first look at modern enterprise traffic," *in Proc. the 5th ACM SIGCOMM conference on Internet Measurement*, pp.15-28, 2005

[13] K. Papagiannaki, "Provisioning IP Backbone Networks Based on Measurements," Phd dissertation, http://citeseer.ist.psu.edu /665560.html, 2003.

[14] B. Li, S. Xie, Y. Qu, G.Y. Keung, C. Lin, J. Liu, and X. Zhang. "Inside the New Coolstreaming: Principles, Measurements and Performance Implications," *in Proc. INFOCOM 2008*, Phoenix, AZ, USA, pp.1031-1039, Ap.2008

[15] P. Chambers, P. Feng, S. Sahu, and D. Saha. "Measurement-based characterization of a collection of on-line games," *in Proc. the 5th ACM SIGCOMM conference on Internet Measurement*, pp.1-1, 2005.

[16] S. Saroiu, K. Gummadi, R. Dumn, S. Gribble, and H. Levy, "An analysis of internet content delivery systems," *in Proc.the 5th symposium on OSDI*, vol.36, no.6, pp.315-327, Dec.2002.

[17] N. Leibowitz, M. Ripeanu, and A. Wierzbicki, "Deconstructing the kazaa network," *in Proc. The Third IEEE Workshop on Internet Applications (WIAPP'03)*, pp. 112-120, June 2003.

[18] S. Sen and J. Wang, "Analyzing peer-to-peer traffic across large networks," *in Proc. ACM SIGCOMM*, Marseille, France, Nov. 2002.

[19] M-S. Kim, Y.J. Won, and J.W. Hong, "Characteristic analysis of internet traffic from the perspective of flows," *Computer Communications table of contents archive*,Vol 29,no.10, pp. 1639-1652, June.2006.

[20] M. Roughan, A. Greenberg, C. Kalmanek, M. Rumsewicz, J. Yates, and Y. Zhang, "Experience in measuring backbone traffic variability: models, metrics, measurements and meaning," *in Proc. the 2nd ACM SIGCOMM Workshop on Internet Measurment*, pp.91 – 92, Nov.2002

[21] P. Owezarski and N. Larrieu, "Internet Traffic Characterization – An Analysis of Traffic Oscillations," *High Speed Networks and Multimedia Communications*, pp.96-107, 2004.

[22] K. Thompson, G. J Miller, Rick Wilder, "Wide-area Internet traffic patterns and characteristics," *IEEE Network* vol.11, no.6, pp.10-23, Nov/Dec.1997.

[23] G. Maier, A. Feldmann, V. Paxson, M. Allman, "On dominant characteristics of residential broadband internet traffic," *IMC '09 Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, 2009, pp: 90-102

[24] G. Maier, F. Schneider, A. Feldmann, "A first look at mobile hand-held device traffic," *PAM'10 Proceedings of the 11th international conference on Passive and active measurement*,2010,pp: 161-170

[25] S. Schiff, "Resolving time-series structure with a controlled wavelet transform," *Optimal Engineering*, vol. 31, no. 1, pp. 2492-2495, Nov. 1992.

[26] J.L. Garcia-Dorado, J.A. Hernandez, J. Aracil, et al., "On the Duration and Spatial Characteristics of Internet Traffic Measurement Experiments," *IEEE Communications Magazine*, vol. 46, no.11, pp. 148-155, Nov.2008

[27] G. Haßlinger, J. Mende, R. Geib, T. Beckhaus and F. Hartleb, "Measurement and Characteristics of Aggregated Traffic in Broadband Access Networks," *Computer Science*, vol.4516, pp.998-1010, 2007

[28] S. Zhang and M.K. Vernon, "Offline and Online Network Traffic Characterization," *Computer Science*, vol.4479, pp.912-923, 2007.

[29] P. Olivier and N. Benameur, "Flow level IP traffic characterization," *Proceedings of the 17th International Teletraffic Congress (ITC17)*, Salvador da Bahia, Brazil, December 2-7, 2001, pp.25-36

[30] K. Sripanidkulchai, B. Maggs, and H. Zhang, "An Analysis of Live Streaming Workloads on the Internet," *Proc. ACM IMC*, Taormina, Italy, pp. 41-54, Oct. 2004.

[31] M. Chesire, A. Wolman, G. Voelker, and H. Levy, "Measurement and Analysis of a Streaming Media Workload," *in Proc. USITS*, San Francisco, USA, pp. 1-1, Mar. 2001.

[32] M.Zink, K.Suh, Y.Gu and J.Kurose, "Characteristics of YouTube network traffic at a campus network-measurement, models, and implications," *Computer Networks*, vol. 53, no. 4, pp. 501-514, Mar.2009.

[33] S. McCreary and kc. claffy, "Trends in wide area IP traffic patterns - A view from Ames Internet Exchange," *in ITC Specialist Seminar*, Monterey, CA, Sep.2000.

[34] N. Brownlee, kc. claffy, "Understanding Internet traffic streams: dragonflies and tortoises," *IEEE Communications Magazine*, vol.40, no.10, pp.110-117, Oct.2002.

[35] C. Wang and X. Zhou, "Design of P2P Traffic Identification Based on DPI and DFI", *International Symposium on CNMT 2009,* pp.1-4, Jan.2009

[36] M. Ye. "AutoSig-Automatically Generating Signatures for Applications," *IEEE International Conference on Computer and Information Technology*, vol.2, pp.104-109,Oct. 2009

[37] H. Liu, and L. Yu, "Toward Integrating Feature Selection Algorithms for Classification and Clustering," *IEEE Trans. Knowledge and Data Engineering*, vol.17, no.4, pp.491-502, April.2005.

[38] T.T.T. Nguyen, G. Armitage, "A Survey of Techniques for Internet Traffic Classification using Machine Learning," *Communications Surveys & Tutorials*, IEEE, vol: 10 , no.4 , pp. 56 – 76, 2008.

[39] T.M. Xia, R.Q. Song, "A Method of P2P Traffic Identification on Internet Based on the Deep Flow Inspection," *International Conference on Communication Software and Networks*, pp.777-780, Feb.2009.

[40] L. Chen, R. Cong, J. Yang et al., "P2P Streaming Traffic Classification in High-Speed Networks". *China Communications*, 2011, 8(5): 70-78

[41] K. Papagiannaki, N. Taft, S. Bhattachayya, P. Thiran, K. Salamatian, and C. Diot, "On the feasibility of identifying elephants in Internet backbone traffic," *Technical Report TR01-ATL-110918*, Sprint Labs, Sprint ATL, November 2001.

[42] Azzouna, N.B., Clerot, F., Fricker, C., Guillemin, F. "A flow-based approach to modeling ADSL traffic on an IP backbone link," *Annals of Telecommunications, Traffic engineering and routing*, 2004, pp:1260-1314.

[43] X. Hei, C. Liang, Y. Liu, and K. W. Ross, "A measurement study of a large-scale P2P iptv system," *IEEE Trans. Multimedia*, vol.9, no.8, pp.1672-1687, Dec.2007.

Jie Yang received her B.E, M.E, and Ph.D degrees from Beijing University of Posts and Telecommunications (BUPT), China in 1993, 1999, and 2007, respectively. Currently, she is an associate professor and Deputy Dean of School of Information and Communication Engineering, BUPT. Her research interests include broadband network technology, information theory in communication systems, network traffic monitoring, and P2P network technology.

Lun Yuan is currently a graduate student in the School of Information and Communication Engineering, BUPT, and received his B.E. degree in information engineering from BUPT in 2010. He is engaged in the research of broadband IP network, and traffic identification and classification.

Chao Dong is a Ph.D. senior student in the School of Information and Communication Engineering, BUPT, and received his B.E. degree from BUPT in 2007. His research focuses on broadband IP network, traffic measurement and classification, mobile network optimization, and mobile network application analyses.

Gang Cheng is a senior software development engineer at Microsoft; he received both B.E. and M.E. degrees in Information Engineering from BUPT in 1997 and 2000,

respectively, and Ph.D. in Electrical Engineering from the New Jersey Institute of Technology in 2005. His research focuses on QoS guarantee issues in high speed networks, Internet routing protocols and service architectures, especially on QoS routing, traffic scheduling, queue management, and congestion control in high speed networks.

Nirwan Ansari is Professor of Electrical and Computer Engineering at the New Jersey Institute of Technology (NJIT); he received his Ph.D degree from Purdue University, West Lafayette, IN, in 1988. His current research focuses on various aspects of broadband networks and multimedia communications. Some of his recognitions include an IEEE Fellow, the NJIT Excellence in Teaching Award in Outstanding Professional Development, IEEE MGA Leadership Award, Thomas Alva Edison Patent Award, the NCE Excellence in Teaching Award, and designation as an IEEE Communications Society Distinguished Lecturer.

Nei Kato received his M.S. and Ph.D. Degrees in Information Science from Tohoku University, Japan, in 1988 and 1991, respectively. He joined the Computer Center of Tohoku University in 1991, and has been a full professor at the Graduate School of Information Sciences since 2003. He has engaged in research on computer networking, wireless mobile communications, image processing, and neural networks, and has published more than 200 papers in journals and peer-reviewed conference proceedings.