Exploring Flow Characteristics in IPv6: A Comparative Measurement Study with IPv4 for Traffic Monitoring

Qiang Li[1], Tao Qin[2], Xiaohong Guan[1,2] and Qinghua Zheng[2]
[1] Tsinghua National Laboratory for Information Science and Technology, Tsinghua University
Beijing 100084, China
[e-mail: qli06@mails.tsinghua.edu.cn, xhguan@tsinghua.edu.cn]
[2] MOE Key Lab for Intelligent Networks and Network Security, Xi'an Jiaotong University
Xi'an 710049, China
[e-mail: tqin@sei.xjtu.edu.cn, qhzheng@mail.xjtu.edu.cn]
*Corresponding author: Qiang Li

Received November 21, 2013; revised January 28, 2014; revised March 1, 2014; accepted March 17, 2014; published April 29, 2014

Abstract

With the exhaustion of global IPv4 addresses, IPv6 technologies have attracted increasing attentions, and have been deployed widely. Meanwhile, new applications running over IPv6 networks will change the traditional traffic characteristics obtained from IPv4 networks. Traditional models obtained from IPv4 cannot be used for IPv6 network monitoring directly and there is a need to investigate those changes. In this paper, we explore the flow features of IPv6 traffic and compare its difference with that of IPv4 traffic from flow level. Firstly, we analyze the differences of the general flow statistical characteristics and users' behavior between IPv4 and IPv6 networks. We find that there are more elephant flows in IPv6, which is critical for traffic engineering. Secondly, we find that there exist many one-way flows both in the IPv4 and IPv6 traffic, which are important information sources for abnormal behavior detection. Finally, in light of the challenges of analyzing massive data of large-scale network monitoring, we propose a group flow model which can greatly reduce the number of flows while capturing the primary traffic features, and perform a comparative measurement analysis of group users' behavior dynamic characteristics. We find there are less sharp changes caused by abnormity compared with IPv4, which shows there are less large-scale malicious activities in IPv6 currently. All the evaluation experiments are carried out based on the traffic traces collected from the Northwest Regional Center of CERNET (China Education and Research Network), and the results reveal the detailed flow characteristics of IPv6, which are useful for traffic management and anomaly detection in IPv6.

A preliminary version of this paper appeared in IEEE ICON 2012, December 12-14, Singapore. This version includes extended measurement and analysis works on one-way flows and group user's behavior characteristics. This work was partially supported by the National Natural Science Foundation of China (61221063, 61103240, 61103241, 91118005, 91218301), the National Science & Technology Support Program of China (2011BAK08B02), the Fundamental Research Funds for the Central University.

http://dx.doi.org/10.3837/tiis.2014.04.009

Keywords: Traffic management, Comparative Measurement, Network Monitoring, IPv4, IPv6.

1. Introduction

Internet has become one of the most important infrastructures of the society, and has penetrated and benefited the quality of our daily life. At the same time, we are in the midst of a transition from IPv4 to IPv6 as the IPv4 addresses are running out. IPv6 is the next generation IP network, which is designed to accommodate current and future growth of the Internet, providing plenty of enhancements over IPv4. Due to the adoption of IPv6 around the globe, increasing amount of attention has been paid to IPv6 in both industry and academia.

A number of papers have been published discussing different aspects of the IPv6 protocol suite, including protocol design, routing mechanisms, transition issues and performance evaluation [1-5]. Some of them are interested in comparing the performance over IPv6 and IPv4 by quantifying the differences in terms of various metrics. There also exist measurement works focus on mining the traffic characteristics of IPv6 network from specific applications or packet level. Nevertheless, all of the above works fail to provide an overall evaluation of the traffic features of IPv6 networks when compared with those of IPv4 networks. As the applications running over IPv6 are quite different from those in IPv4 network, the traffic features and model obtained from IPv4 are not readily available for IPv6 network management. Therefore, it is of great significance to investigate the traffic features of IPv6 to devise efficacious policies for IPv6 network monitoring.

To address these issues, in this paper we perform a careful comparative measurement study between the IPv4 and IPv6 traffic to explore the traffic features of IPv6 from flow level. Firstly, we analyze the difference in flow statistical characteristics, and find that approximately 90% of both IPv6 and IPv4 flows are of less than 10 packets. In IPv6 networks, we find that the percentage of elephant flows is bigger than that of IPv4. Comparative analysis results of flow duration also indicate that the percentage of flows with long duration in IPv6 networks is larger than that of IPv4. This is due to the fact that the applications running in IPv6 networks are mostly file downloading and online video, which results in a large number of long duration flows. Additionally, we find the users in IPv6 network usually generate a large number of traffic flows with huge traffic volumes. Especially, nearly 40% of flows for IPv6 traffic are generated on some specific ports which are related with some special applications. Those findings are pivotal for traffic engineering in IPv6, such as flow control.

Secondly, we analyze the characteristics of one-way flow, consisting of packets in only one direction without forward or backward packets. We find that approximately 95% of one-way flows carry less than 10 packets. In particular, about 40% and 18% of one-way flows in IPv4 and IPv6 traffic contain only one packet, respectively. Additionally, TCP traffic dominates the one-way flows in the IPv6 networks. While in IPv4 networks, the percentage of TCP one-way traffic is equal to that of UDP traffic. Deep investigations into one-way flows show that some of the one-way flows are related to scanning-like attacks due to their flow statistical characteristics and they are important information sources for abnormal behavior detection.

Finally, to address the challenge of processing massive data while monitoring the large-scale networks, we develop group flow model based on higher aggregated flow level, which can greatly reduce the number of flows while keeping the primary traffic characteristics compared with Netflow. Based on this model, we extract the connection degree as traffic

feature and use Renyi cross entropy method to measure the group users' behavior dynamics. It is found that the group users' behavior in IPv6 networks is more stable, which can be utilized for abnormal behavior detection and traffic monitoring.

To demonstrate the effectiveness of the proposed method, we analyze several real traffic data sets collected from the Northwest Regional Centre of CERNET and obtain several insightful experimental results, which reveal the detailed flow characteristics of IPv6. Other researchers who are interested in those traces can require downloading them by contracting with the authors.

The rest of the paper is organized as follows: Section 2 introduces the related work. Section 3 presents the framework of the measurement and analysis methods. Section 4 explores the statistical characteristics of traffic flows in IPv6 network, as well as in IPv4 network. Section 5 analyzes one-way traffic flows and the dynamic changes of group users' behavior. Finally, conclusions are drawn in Section 6.

2. Related Work

It is important to understand and control the Internet by measuring the network traffic, which reveals the characteristics of traffic and users' behavior. Many papers have presented measurement work on the IPv4 network including measurement methodologies [6], performance measurement [7], anomaly detection [8] and user behavior analysis [9]. Compared with IPv4 networks, there are only a few works on IPv6 network measurement. The related works are addressed as follows.

Researchers have found that IPv6 network deployment is stronger in Europe and Asia-Pacific region than in North America [10]. China has built an IPv6-only backbone in the CERNET, i.e. CNGI-CERNET2 since 2003 [11], which is provided as a good experiment platform for researchers in China. Most of the researchers perform their measurement on the packet-level traces, and focus on the usage of BT applications and the characteristics of BT traffic and BT users' behavior in IPv6 network. Zhang et al. [12] performed a measurement study on the BT traffic behaviors in IPv6 network, and presented a comparative analysis of BT flows over TCP and uTP (uTorrent Transport Protocol). Gao et al. [13] investigated the usage of Chinese IPv6 network from the perspective of user's behavior, and they found that a very small fraction of users produce most traffic, most of which are P2P sharing or video streaming. In [14], Ao and Chen collected packet traces from external links and log files from a private BT tracker, and they analyzed users' performance on BT under IPv6 network environment. In [15, 16], BitTorrent packet traffic features are mainly examined for IPv4 and IPv6 from several perspectives, such as autocorrelation, spectral density and self similarity of packet size and packet interarrival time. Li et al. [17] collected packet traces from two academic IPv6 networks, and made a comprehensive study from the perspective of an operational ISP for IPv6 network development and management. To obtain detailed characteristics of IPv6 traffic and explore its differences with that of IPv4, in this paper, we perform a comparative measurement study between IPv6 and IPv4 from flow level for traffic monitoring.

In regard to traffic flow analysis, measurement on non-productive traffic is one of the hot topics in the area of traffic engineering and abnormal detection. In 2004, Pang et al. [18] presented a study of the broad characteristics of Internet Background Radiation (IBR), which is the non-productive traffic sent to darknet (blocks of unused IP addresses). Since then, the darknet has proved to be a useful tool to monitor malicious activities (such as worms, network scanning and DDoS) or mis-configurations [19]. IBR traffic is somehow related with unsolicited one-way traffic [20]. Glatz and Dimitropoulos introduced a classification scheme

for one-way traffic, especially useful for monitoring IBR traffic [21]. In [22], Ford, Stevens and Ronan created an IPv6 darknet to observe the IBR traffic, and found that the detected darknet traffic seems to be attributable to mis-configuration rather than malicious activity. The authors believed that it may be a consequence of the huge IPv6 addresses and immaturity of IPv6 Internet at that time. Enlighten by these works, we will focus on the one-way traffic measurement, especially for the IPv6 abnormal behavior detection in this paper.

With the development of Internet, there are more and more flow records, which is hard to process for real time application. As IPv6 have a huge address space and it will generate much more packet/flow data than IPv4, it becomes more challenging to monitor and analyze massive Internet traffic. Many traffic data reduction techniques (i.e., traffic filtering, packet sampling, and flow technique) have been proposed to solve these problems. In [23], Kohler et al. analyzed the structure of IPv4 addresses (a subnet of the address space) to understand the properties of large aggregated traffic. They analyzed the packet count distribution with different destination-prefix aggregation, and obtained some interesting results for traffic monitoring. Researchers also proposed several highly aggregated flow models to measure traffic features and changes of the traffic pattern for network management purpose [24, 25]. In this paper, we perform a measurement research on the IPv6 aggregated traffic and measure the user's behavior dynamic on this level for abnormal detection.

Our previous work [26] has dealt with analyzing the differences of statistical characteristics between IPv4 and IPv6 traffic features based on the 5-tuple flow model, and measuring the users' behavior dynamics using Renyi entropy method. In this paper, we complements these works by presenting an in-depth investigation on the flow characteristics over IPv4 and IPv6 networks based on three flow level models. What's more, we analyze the characteristics of one-way flow and group users' dynamic behavior for abnormal detection. The main contributions of this paper are summarized as follows:

1) We present a comprehensive measurement study between IPv6 and IPv4 networks to explore the flow characteristics from three flow levels: 5-tuple flow level, bidirectional flow level and group flow level.

2) By analyzing the one-way flow characteristics, we find that there are many one-way flows both in the IPv4 and IPv6 traffic. Some of them are related to network scanning-like attacks.

3) We also analyze group users' behavior dynamic characteristics and find that the user's behavior of the IPv6 networks is more stable. Our proposed method could also be used for anomaly detection in future.

3. Framework of Measurement Methods

The proposed framework for comparative analysis on traffic and users' behavior in different network environments is divided into three steps, as shown in **Fig. 1**.

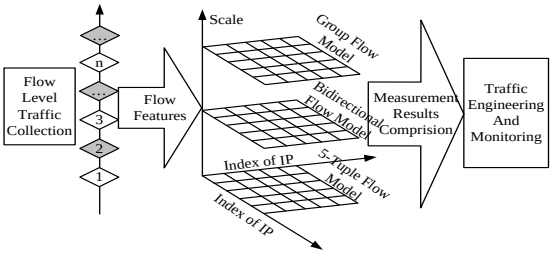


Fig. 1. Framework of Measurement Methods

Step 1: Traffic data collection. The raw online traffic data are collected using the Netflow protocol v5 and v9 for IPv4 and IPv6, respectively. The traffic collection platform is set in the Northwest Region Center of CERNET, and the network being monitored is the campus network of Xi'an Jiaotong University.

Step 2: Flow features extraction. We analzye the characteristics of traffic on differnent aggregated level. In this paper we employ the standard 5-tuple flow model, bidirectional flow model and group flow model to caputre the traffic features.

Step 3: Comparison analysis and measurement. Based on the features obtained, we measure the characteristics of IPv4 and IPv6 flows. We present several interesting findings and investigate the measurement results for security monitoring.

Based on the results obtained from above three steps, we perform an insightful comparison of the difference of traffic features between IPv4 and IPv6 networks from the view of three levels: regular 5-tuple flow level, bidirectional flow level and group flow level.

3.1 Flow Model and Features for Traffic Measurement

The technology of network flow [27] has been developed for traffic summaries and statistics, and has been used in the areas of network monitoring, traffic measurement, traffic analysis, etc. It is usually defined as a set of packets between two endpoints with the same 5-tuple: source IP/port, destination IP/port and protocol number. Netflow is the most widely used 5-tuple flow model, and has become a de facto industry standard. However, this kind of model also cannot reflect the exchange characteristics. To fully explore the difference between the IPv6 and IPv4 traffic, we develop two kinds of flow models on different aggregated level in this paper. Based on all these flow models, we extracted several flow features to capture the traffic characteristics.

Feature 1: Flow Size, which is defined as the total number of packets the specific flow holds. It is employed to describe the size of the flow.

Feature 2: Flow Duration, which is defined as the lasting length of the flow. It is employed to capture the characteristics of the communications.

Bidirectional flow model: Different with the traditional Netflow model, the bidirectional flow model is defined as a set of packets in a specific time window T with the same values of five fields: source IP/port, destination IP/port and protocol number, including the forward and backward packets. Generally speaking, the bidirectional flow is a two-way flow consisting of packets in the forward and backward directions. However, there are many one-way flows in actual networks, only with packets in forward or backward direction. These one-way flows may results from network errors (e.g., mis-configurations, operation errors), application behaviors or attacks. We use this model to extract the real one-way flows from the regular flow records for analysis.

For flow monitoring in large-scale network, it will generate massive 5-tuple flow records, and it is still hard to process for real time application. Some researchers propose high level flow model (e.g., Origin-Destination flow model) or measurement matrix (e.g., AS-level) for abnormal detection, characterizing traffic behaviors in large-scale networks from a network-wide perspective [28-30]. In this paper, we develop a group flow model on an aggregated traffic level to reduce the flow amount.

Group flow model: A group is defined as a set of IP addresses with the same network prefix. Then all the packet/byte/flow between two groups is aggregated in the group flow. It is proposed to aggregate traffic packets on higher flow level and will greatly reduce the number of flows to be processed while remaining the primary traffic characteristics. We use this model to aggregate the IPv6 traffic and measure the group users' behavior for abnormal behavior

detection.

The group flow model can be described using the matrix *M(t)* in Equation (1), where *m* is the number of groups *{S_1, S_2, …, S_m}* in the monitored network, and they communicate with *n* groups *{D_1, D_2, …, D_n}* outside at time *t*, where the groups are denoted by IP addresses with network prefix. Let in_{ij} be the traffic transferred from D_j to S_i, and out_{ij} be the traffic transferred from S_i to D_j. Note that we use in_{ij} and out_{ij} to refer either of the three types of traffic (in number of bytes, packets and flows). In this way, the traffic flow patterns can be viewed as a time series of traffic matrix *M(1), M(2), …, M(t)*.

$$M(t)=\begin{bmatrix} (in_{11},out_{11}) & (in_{12},out_{12}) & \cdots & (in_{1n},out_{1n}) \\ (in_{21},out_{21}) & (in_{22},out_{22}) & \cdots & (in_{2n},out_{2n}) \\ \cdots & \cdots & \cdots & \cdots \\ (in_{m1},out_{m1}) & (in_{m2},out_{m2}) & \cdots & (in_{mn},out_{mn}) \end{bmatrix} \quad (1)$$

Based on the matrix, we extract another flow feature for in-depth traffic measurement.

Feature 3: Connection Degree (*CD*), which is defined as the number of different groups that one group connected with in a specific time window *T*, is normally divided into In-Degree (*CID*) and Out-Degree (*COD*), as shown in Equations (2) and (3). As to two-way traffic, in-degree is equal to out-degree as the packets exist in both directions. This feature can capture the communication range effectively and measure the dynamics of the users' behavior, which can also capture the characteristics of abnormal behaviors, such as scanning behavior will generate huge number of connections.

$$CID_i=\left\|\{in_{ij}\,|\,in_{ij}\neq 0, 1\le j\le n\}\right\|, 1\le i\le m \quad (2)$$

$$COD_i=\left\|\{out_{ij}\,|\,out_{ij}\neq 0, 1\le j\le n\}\right\|, 1\le i\le m \quad (3)$$

3.2 Methods for Dynamic Characteristic Measurement

Users will join and leave the network at different times, in other words, they appear in the networks randomly. Those types of behavior dynamics can be captured using connection degree and Renyi entropy [31]. We employ the Renyi entropy to measure the dynamic changes of connection degree, in turn to monitor the users' behavior dynamics. The Renyi entropy of order α is defined as the following.

$$H_\alpha(p)=\frac{1}{1-\alpha}\log_2\sum_r p_r^\alpha \quad (4)$$

where $0<\alpha<1$, p is a discrete stochastic variable, and p_r is the distribution function. The Shannon entropy is a special case of Renyi entropy with $\alpha\to 1$. The Renyi cross entropy of order α is

$$I_\alpha(p,q)=\frac{1}{1-\alpha}\log_2\sum_r \frac{p_r^\alpha}{q_r^{\alpha-1}} \quad (5)$$

where p and q are two discrete variables, p_r and q_r are the corresponding distribution functions. One important property of the Renyi cross entropy is that if p and q have the same distributions, then $I_\alpha\to 0$. If we choose $\alpha=0.5$ in Equation (4), then the entropy measure in Equation (5) is symmetric, which means that $I_\alpha(p,q)=I_\alpha(q,p)$. If $\alpha=0.5$ is chosen, the Renyi cross entropy can be rewritten into

$$I_{0.5}(p,q)=2\log_2\sum_r\sqrt{p_r q_r} \quad (6)$$

This symmetric method is suitable for network dynamic change monitoring as the changes caused by abnormal behaviors are symmetric. So we use Renyi cross entropy to measure the changes of connection degrees at different times as show in Equation (7), where *F(·)* is the probability distribution function (PDF) of the features. In this paper, we use the frequency to approximate the PDF such as Equation (8). It is also very easy to select the threshold for abnormal detection using Renyi cross entropy method as the entropy is equal to zero if there is no abnormal behaviors.

$$\begin{aligned} f_1 &= I_{0.5}(CD(t), CD(t+1)) \\ &= 2\log_2 \sum_k \sqrt{F(CD_k(t))F(CD_k(t+1))}, 1 \le k \le m \end{aligned} \tag{7}$$

$$F(CD_k) = {CD_k} \Big/ {\sum_{i=1}^{m} CD_i} \quad 1 \le k \le m \tag{8}$$

3.3 Methods for Performance Evaluation

To evaluate the performance of the dynamic methods proposed in this paper, we employ two methods used in prior works, and they are described as follows:

1) EWMA Method

The Exponentially Weighted Moving Average (EWMA) is a statistic for monitoring the process that utilizes different weights for historical observations. It is proved to be an excellent estimation function under ill conditions to detect the abnormal behaviors by discovering abrupt changes in traffic volumes [32]. In this paper, we select L=1.96 for the 3% significance level, and select the forgetting factor λ=0.3 for better forecasting results. The connection degree is selected as the features for measurement.

2) Shannon Entropy Based Method

Entropy has been proved to be a good method to measure the uncertainty of a variable. It has been used for abnormal traffic detection in network area and received quite good results [33]. Usually the entropy method is applied on the features of traffic, such as the source/destination IP address/port number, and connection degree. To make the comparison results more reasonable, we also select connection degree as the measurement feature.

4. Exploring Characteristics of Flow Level

4.1 Traffic Collection Platform and Test Beds

The traffic traces used in this paper are collected from the Northwest Regional Center of CERNET. The network being monitored is the campus network of Xi'an Jiaotong University, which contains more than 30,000 end users with global IP addresses, including students, faculty members and contract personnel from service providing companies. Their behavior characteristics are complex enough to perform our measurement and analysis work. IPv6 is deployed in campus network using IPv4-IPv6 dual-stack. All of the traces used in the datasets are Netflow records collected on an egress router with a bandwidth of 10 Gbps for the time horizon of more than two months ranging from 2011 to 2012.

Due to the huge amount of traffic data, we use the NfDump [34] tools to collect Netflow data and store the data into binary files. According to the settings on router, the Netflow data is sent to the Netflow collector every 5 minutes, which means that the export time window is of 5-minute length. All the periodic exported Netflow records will be saved into the same file. In this paper, we take two one-day long traces for IPv4 and IPv6 respectively. The traffic traces

are described in **Table 1**.

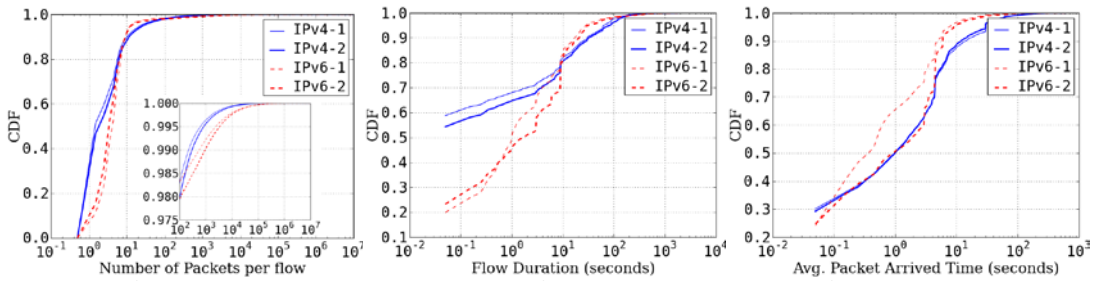
Table 1. Description of traffic traces

Trace	Collection Time	# of Packets (1e9)	# of Flows (1e6)	# of Unique Addresses (1e3)	# of Campus Users (1e3)
IPv4-1	2012/11/14 (1 day)	10.7	245.9	6325.1	15.2
IPv4-2	2012/11/18 (1 day)	11.1	204.5	6063.1	15.1
IPv6-1	2011/10/06 (1 day)	15.2	76.8	239.3	19.7
IPv6-2	2011/10/22 (1 day)	16.7	86.1	409.6	25.6

4.2 Distribution of Flow Characteristics

The difference of the flow size (i.e., the number of packets) between IPv6 and IPv4 flows is shown in **Fig. 2(a)**. As can be observed, about 90% of both IPv6 and IPv4 flows have less than 10 packets. In prior research works, it has been found that there are a large number of mice flows and relatively fewer elephant flows in the Internet traffic [35]. Few of the flows may contribute the majority of total traffic volume (in the number of packets or bytes) and those flows are usually named as elephant flows. However, the mice flows only carry little share of the total bandwidth, which are used for instant message exchange and other HTTP applications. And the small number of elephant flows usually occupies a large share of the total bandwidth. The elephant can be controlled to achieve the goal of traffic management and routing. In IPv6 networks, the traffic flows are also divided into elephants and mice. The thresholds for judging elephant flows are different in different networks. In this paper, we regard the flows whose packet number is large than 1000 as elephant flows. From **Fig. 2(a)**, we could observe that the percentage of mice flows in IPv4 network is larger than in IPv6 network. While, the percentage of elephant flows is larger in IPv6 than in IPv4.

The comparative analysis results of flow duration are shown in **Fig. 2(b)**. The percentage of flows with long duration in IPv6 networks is larger than that in IPv4 network. As we know, the applications running in the IPv6 networks are mostly file downloading and online video, which result in a large number of long-duration flows. Those findings of traffic features are important for traffic management, and in the next step we will study new methods for traffic modeling and elephant flow identification. **Fig. 2(c)** shows the CDF (cumulative distribution function) of average packet arrived time for flows. The average packet arrived time means the average time of the interval time for the packets in flow, it could be calculated as flow duration divided by number of packets minus one. For the flows with only one packet, the average packet arrived time will be set as zero. From the figure, we observe that the percentage of flows with average packet arrive time less than 10 seconds is larger in IPv6 network than that in IPv4 network, which verifies that IPv6 network provide better service performance.

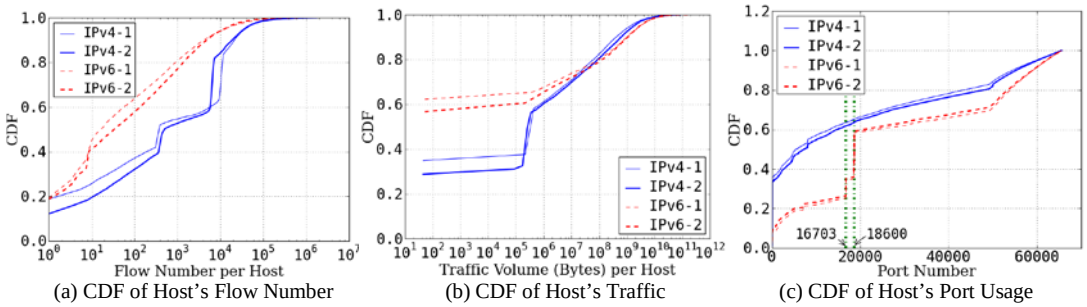


(a)CDF of Flow Size in Packets (b) CDF of Flow Duration (c) CDF of Average Packet Arrived Time

Fig. 2. Statistical Characteristics of Flow Size, Flow Duration and Average Packet Arrived Time

4.3 Analysis of Traffic Characteristics per User

We measure the traffic characteristics from the user view, and an IP address is treated as a specific user. The CDF of users' traffic is shown in **Figs. 3(a)** and **3(b)**. The users in IPv6 network usually have a large number of traffic flows and a large number of traffic volumes. We take a look at the port usage of users by analyzing the distribution of flow on ports using the same trace, and the results are shown in **Fig. 3(c)**. It illustrates that there are about 40% of flows for IPv6 on some ports which are very different from IPv4. With manual check, we select the top two ports, 18600 and 16703, which occupy about 15.5% and 8.6% of flows respectively for IPv6-1 trace. It is similar for IPv6-2 trace. Some of these flows are produced by µTorrent, which is a tiny BitTorrent client setting to use UDP port 18600 as the default port in some version.



(a) CDF of Host's Flow Number (b) CDF of Host's Traffic (c) CDF of Host's Port Usage

Fig. 3. User Behavior Characteristics in IPv4 and IPv6 Networks

With the observations of flow characteristics, we find that the usage of IPv6 network is quite different from that of IPv4, which makes the obvious differences of IPv4 and IPv6 flows. All these findings are important for traffic engineering, such as flow control.

5. Traffic Analysis for Security Monitoring

Most people believe that it is more difficult to perform address and port scanning over the entire IPv6 address range or in specific IPv6 networks since IPv6 expands the address structure length from 32 bits to 128 bits. However, it does not mean that IPv6 is safe enough to prevent from malicious attack or worms [36]. Therefore, it is important to investigate the traffic features of IPv6 in order to detect the abnormal behavior in IPv6 networks and design new policies for IPv6 traffic management. In this section, we analyze the flow characteristics for security monitoring purpose.

5.1 One-way Flow Measurement for Abnormal Detection

Usually, the communications between any two hosts are bidirectional. The applications running on the hosts generate two-way flows which could be defined as a set of flows that have reverse values of source IP, destination IP and port numbers in the same time window. However, people have found that there are one-way traffic flows without corresponding reverse flows. Compared with two-way flows, one-way flows are consisting of packets in only one direction, which are normally associated with network attacks (e.g., vulnerability scanning), mis-configurations, operation errors or application behaviors [21].

5.1.1 Extracting One-way Flows

To analyze one-way traffic, one-way flows have to be filtered out from the trace files. We write a simple script to process Netflows in every 5-minute length file into bidirectional flows. The Netflow records contain the source IP/port, destination IP/port, protocol number, byte/packet counts, timestamps and other information. Based on the 5-tuple, we will check all the flows one by one in the same time window to consider whether there are reverse flows. After processing, the flows in trace files are divided into one-way flows and two-way flows. **Table 2** shows the number of Netflows and bidirectional flows (including one-way flows and two-way flows) in the traces. We can observe that the bidirectional flow model can reduce the number of flow records to be processed while reflecting the exchanging characteristics between end hosts. We also find the percentage of flow records in IPv6 networks are reduced more than that of IPv4 networks in which there are more percentage of one-way flows. **Table 3** shows the statistic information of one-way flows and two-way flows. In IPv4 traces, the number of bytes is similar in one-way and two-way flows. But the number of packets/flows in one-way flows is larger than that in two-way flows. In IPv6 traces, the number of packets/flows/bytes in one-way flows is usually smaller than that in two-way flows. We can also find that there are obvious differences between IPv4 and IPv6 traffic for the packet/flow/byte counts in one-way and two-way flows. For the one-way flows, the number of packets is similar for both IPv4 and IPv6 traces, and the number of flows is much less in IPv6.

Table 2. Information of netflow and biflow in traces

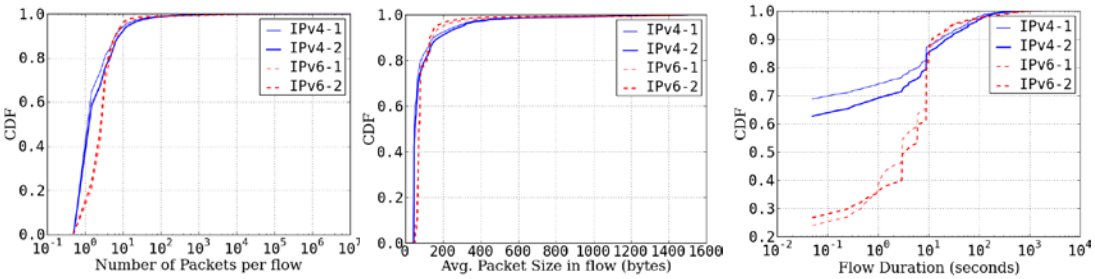
Trace	# of Netflows (1e6)	# of Biflows (1e6)	(# of Biflows)/(# of Netflows)
IPv4-1	245.9	200.6	81.57%
IPv4-2	204.5	169.0	82.64%
IPv6-1	76.8	48.1	62.63%
IPv6-2	86.1	61.7	71.66%

Table 3. Information of one-way flows and two-way flows in traces

Trace	One-way Packets (1e9)	One-way Flows (1e6)	One-way Bytes (1e12)	Two-way Packets (1e9)	Two-way Flows (1e6)	Two-way Bytes (1e12)
IPv4-1	6.015	165.1	3.149	4.676	80.82	3.521
IPv4-2	6.606	139.9	3.887	4.518	64.52	3.376
IPv6-1	5.805	26.36	5.637	9.412	50.39	8.662
IPv6-2	6.252	43.82	6.012	10.44	42.28	9.615

5.1.2 One-way Flow Characteristics

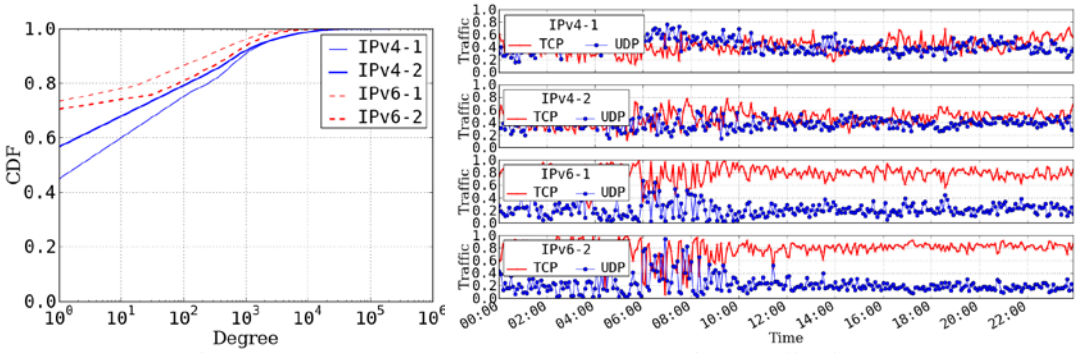
Firstly, we draw some CDF curves of packet number, average packet size and flow duration for one-way flows as shown in **Fig. 4**. As observed in **Fig. 4(a)**, about 95% of one-way flows carry less than 10 packets, and there is about 40% and 18% of one-way flows with only one packet in IPv4 and IPv6 traffic respectively. More than 90% of IPv4 and IPv6 one-way flows' average packet size is less than 200 bytes as shown in **Fig. 4(b)**. We can see that more than 90% of both IPv4 and IPv6 flows are less than 10 seconds in **Fig. 4(c)**. However, about 40% and 70% of IPv4 and IPv6 one-way flows last less than one second respectively.



(a) CDF of Packet Number (b) CDF of Avg. Packet Size (c) CDF of Flow Duration

Fig. 4. CDF of Packet Number, Average Packet Size and Flow Duration for One-way Flow

Secondly, we analyze the characteristics of user behavior for one-way flows as shown in **Fig. 5**. Based on the one-way flows, we compute the connection degree for users in the monitored network, i.e. the distinct number of hosts with which the user communicate. The CDF of hosts' connection degree is shown in **Fig. 5(a)**. We find that at least 40% of users' connection degree is one, and the percentage for IPv6 users is much higher than IPv4 users. **Fig. 5(b)** shows the percentage changes of TCP and UDP one-way traffic in 24 hours. It's interesting that the rate of TCP and UDP one-way traffic in IPv4 network are nearly equal, around 50%. However, the TCP one-way traffic in IPv6 dominates the total traffic.



(a) CDF of Host's Connection Degree (b) Rate of Total Traffic of One-way Flow

Fig. 5. User Behavior Characteristics for One-way Flow in IPv4 and IPv6 Networks

5.1.3 Abnormal Behavior Detection Based on One-way Flow Analysis

We analyze the statistical characteristics of one-flows whose average packet size is less than 200 bytes and flow size is bigger than 1000 packets. The results are shown in **Table 4**. We find that all of the flows have large number of packets, and the packet size is small without obvious changes. Then we calculate the hosts' connection degree in these flows, and find that there are some hosts with connection degree more than 5000 in IPv4 traces, the maximum connection degree reaches above 7000. Those statistical characteristics as shown in **Table 5** reveal those flows are generated by scanning-like attacks [37, 38].

Although the connection degree of hosts in IPv6 is not so high, the maximum one is about 300, but according to the results obtained in our prior works [36], those behaviors could also be scanning-like activities. Based on the above measurement results, we believe the one-way flows are generated by the abnormal behaviors and we can perform deep analysis for abnormal detection.

Table 4. Information of selected one-way flows in traces

Trace	Percentage	Mean of Avg. Packet Size (bytes)	Variance of Avg. Packet Size
IPv4-1	0.16%	59	22
IPv4-2	0.15%	64	24
IPv6-1	0.31%	69	13
IPv6-2	0.23%	68	10

Table 5. Flow statistical characteristics of scanning-like attacks

Anomaly Type	Anomaly traffic characteristics
Port Scan	Probes many destination ports. It will generate many one-way flows to some specific addresses.
Network Scan	Probes many destination addresses. It will generate many one-way flows to some specific ports.
Worms	Scanning by worms for vulnerable hosts. It is similar with the Network Scan.

5.2 Dynamic Change Detection for Large-scale Network

5.2.1 Flow Records Reduction Based on Group Flow Model

In actual networks, the specific users are often identified by unique IP address. However, the IP space is so large in IPv6 network that the number of flows generated by the IPv6 users will be extremely large. We develop the group flow model mentioned in Section 3.1 to reduce the number of flow records being processed in large-scale networks and capture the characteristics of users' behaviors. To be simple, the group is defined as the aggregated IP addresses or subnets. We will explore and compare the characteristics of IPv4 and IPv6 traffic to examine the differences on different aggregate level (using different length of network prefix to obtain the address space or groups). It helps us to learn the distribution of aggregated traffic among the groups. Since the lengths of IP address in IPv4 and IPv6 are different, we choose the prefix length according to the number of users in buildings. The preferred lengths for prefix are 24 and 64 for IPv4 and IPv6 addresses respectively. Then the flows in the same 5-minute flow file with identical protocol, identical source and destination aggregated IP address as well as identical source and destination ports are aggregated. **Table 6** gives the statistical information of groups in the traces. The in-campus group means the group in campus network, and each user in this kind of group is allocated a local IP address. **Table 7** shows the results of flow records reduced by the group flow model. We can see that the number of group flows is reduced by 75% to 80%, which is much higher for IPv4 traces than IPv6 traces.

Table 6. Information of user group in traces

Trace	Prefix Length	# of In-campus Groups	# of All Groups
IPv4-1	24	207	1499145
IPv4-2	24	207	1456714
IPv6-1	64	75	24990
IPv6-2	64	74	26906

Table 7. Information of netflow and group flow in traces

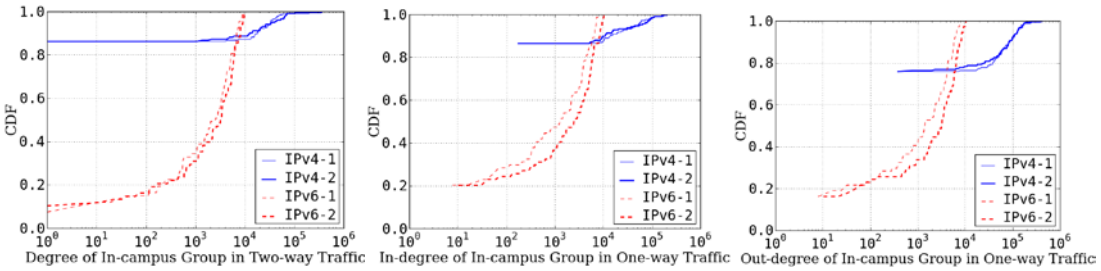
Trace	# of Netflows (1e6)	# of Group flows (1e6)	(# of Group flows)/(# of Netflows)
IPv4-1	245.9	39.3	15.98%
IPv4-1	204.5	38.2	18.67%
IPv6-1	76.8	15.3	19.7%
IPv6-2	86.1	21.2	24.62%

5.2.2 User's Dynamic Behavior Measurement for Abnormal Behavior Detection

Macious activities will cause abnormal changes in the traffic patterns. In this subsection, we choose connection degree as the flow feature. We firstly examine its distribution for one-way and two-way flows, and then measure its dynamic changes for abnormal behavior detection.

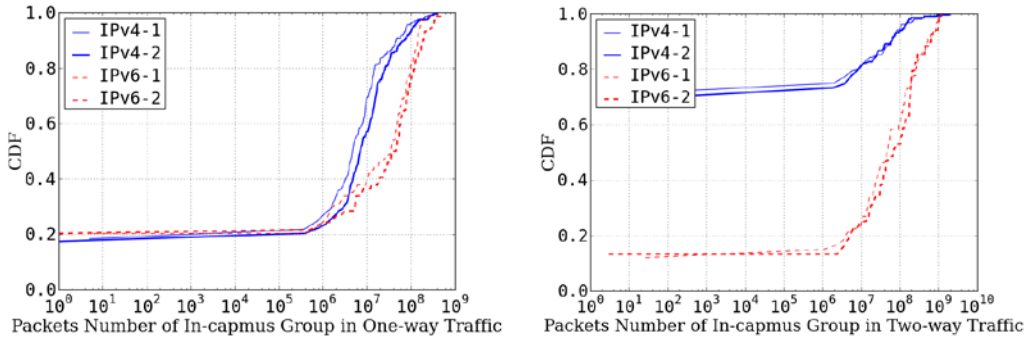
Firstly, we analyze the characteristics of group flows. In **Fig. 6(a)**, it shows the CDF of connection degree of in-campus groups in two-way traffic. It shows that there are at least 85% of in-campus groups with connection degree as one for IPv4, while it is about 10% for IPv6 traffic. The results of *CID* and *COD* are similar for in-campus groups in one-way traffic as shown in **Figs. 6(b)** and **6(c)**. However, the percentage of in-campus with only one group connected raises to 20% for IPv6 traffic, and about 80% of in-campus groups in one-way traffic connect with more than 100 groups in IPv4 traffic.

The curve shape is very similar for the CDF of packet number and the CDF of traffic volume in bytes. So here we only give the result for packet number. **Fig. 7** shows the CDF of the packet counts of one-way and two-way flows for in-campus groups. For one-way traffic, there are about 20% of in-campus groups with less than 10 packets in both IPv4 and IPv6 networks. However, we find that about 70% of in-campus groups contain 10 packets in IPv4 two-way traffic and it is about 17% in IPv6 two-way traffic, as shown in **Fig. 7(b)**.



(a) Degree of Group for Two-way Flow (b) In-degree of Group for One-way Flow (c) Out-degree of Group for One-way Flow

Fig. 6. CDF of Connection Degree of In-campus Group for Aggregated One-way and Two-way Traffic



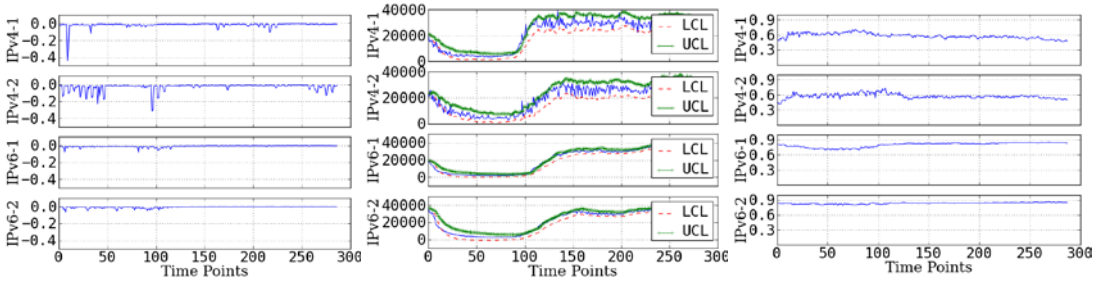
(a) CDF of Packet Number for One-way Traffic (b) CDF of Packet Number for Two-way Traffic

Fig. 7. CDF of Packet Number of In-campus Group for Aggregated One-way and Two-way Traffic

The dynamic measurement results are shown in **Fig. 8**. The performance of the Renyi cross entropy is shown in **Fig. 8(a)**, the two upper subfigures is the results of IPv4 traces and the lower two are those of IPv6 traces. As the figures show, the user's behavior is stable at most of the time points, and the dynamic changes are slight. But there are also some time points at which the dynamic changes is large, and we believe this is caused by abnormal behavior. The

measurement results also shows that the user's behaviors in IPv6 networks are more stable than IPv4 networks, which means that we need more sensitive methods for IPv6 abnormal behavior detection.

To evaluate the performance of the proposed methods, the measurement results of EWMA and Shannon Entropy methods are shown in **Fig. 8(b)** and **8(c)**. As the figures show, the measurement results of those two methods are not so efficacious, and there are too many slight dynamic change points and it is difficult to determine the real dynamic change points. In **Table 8**, we analyze statistical characteristics of the measurement results, which show that our methods are more stable and suitable for dynamic changes detection.



(a) Renyi Cross Entropy Methods (b) EWMA Methods (c) Shannon Entropy Methods

Fig. 8. Dynamic Changes of Degree of In-campus Group for Aggregated Two-way Traffic in IPv4 and IPv6 Networks

Table 8. Stability analysis of group behavior dynamics

Trace	Renyi Cross Entropy Method		EWMA Method		Shannon Entropy Method	
	Mean	Variance	Mean	Variance	Mean	Variance
IPv4-1	-0.017	0.034	21372.1	11170.8	0.576	0.050
IPv4-2	-0.027	0.043	19650.9	9597.8	0.574	0.048
IPv6-1	-0.009	0.011	20429.8	13912.2	0.802	0.047
IPv6-2	-0.008	0.010	21926.4	13206.9	0.836	0.015

6. Conclusions and Future Works

With the widespread utilization and deployment of IPv6, it has become increasingly important to understand the difference between the IPv4 and IPv6 network traffic features, which serves as the foundation of new traffic management policy design for IPv6 network. In this paper, we present a measurement study to explore the differences between traffic characteristics of IPv4 and IPv6 networks from flow level. The differences between IPv4 and IPv6 traffic features were investigated in terms of distribution of flow size, flow duration and average packet arrived time. We presented several interesting findings which are useful for designing new management policyes for traffic monitoring in IPv6. We also compared the differences of user's behavior characteristics and find that the user's behavior of the IPv6 networks is more stable, which shows there is no large-scale malicious traffic in our IPv6 datasets, but our method can be used for anomaly detection in future. All the experiments are carried out based on the traffic traces collected from the Northwest Regional Center of CERNET, and the results reveal the detailed flow characteristics of IPv6, which are quite different from those of IPv4, and a deep analysis is needed for efficacious IPv6 traffic management.

References

[1] Y. Wang, S. Ye, X. Li, "Understanding current IPv6 performance: a measurement study," in *Proc. of 10th IEEE Symposium on Computers and Communications (ISCC 2005)*, pp. 71-76, 2005. Article (CrossRef Link).

[2] W.-L. Shiau, Y.-F. Li, H.-C. Chao, P.-Y. Hsu, "Evaluating IPv6 on a large-scale network," *Computer Communications*, 29 (2006) 3113-3121. Article (CrossRef Link).

[3] T.M. Raste, D.B. Kulkarni, "Design and implementation scheme for deploying IPv4 over IPv6 tunnel," *Journal of Network and Computer Applications*, 31 (2008) 66-72. Article (CrossRef Link).

[4] E. Gamess, R. Surós, "An upper bound model for TCP and UDP throughput in IPv4 and IPv6," *Journal of Network and Computer Applications*, 31 (2008) 585-602. Article (CrossRef Link).

[5] L. Yuk-Nam, L. Man-Chiu, T. Wee Lum, L. Wing Cheong, "Empirical Performance of IPv6 vs. IPv4 under a Dual-Stack Environment," in *Proc. of IEEE International Conference on Communications (ICC'08)*, pp. 5924-5929, 2008. Article (CrossRef Link).

[6] F. Huici, A.d. Pietro, B. Trammell, J.M.G. Hidalgo, D.M. Ruiz, N. d'Heureuse, "Blockmon: a high-performance composable network traffic measurement system," in *Proc. of Proceedings of the ACM SIGCOMM 2012 conference on Applications, technologies, architectures, and protocols for computer communication, ACM*, Helsinki, Finland, pp. 79-80, 2012. Article (CrossRef Link).

[7] E. Damergi, E. Mohamed, B. Ammar, "Network performance evaluation using traffic measurements," in *Proc. of First International Symposium on Control*, Communications and Signal Processing, pp. 523-526, 2004. Article (CrossRef Link).

[8] M. Thottan, C. Ji, "Anomaly detection in IP networks," *IEEE Transactions on Signal Processing*, 51 (2003) 2191-2204. Article (CrossRef Link).

[9] K. Cho, K. Fukuda, H. Esaki, A. Kato, "The impact and implications of the growth in residential user-to-user traffic," *SIGCOMM Comput. Commun. Rev.*, 36 (2006) 207-218. Article (CrossRef Link).

[10] A. Dhamdhere, M. Luckie, B. Huffaker, k. claffy, A. Elmokashfi, E. Aben, "Measuring the deployment of IPv6: topology, routing and performance," in *Proc. of Proceedings of the 2012 ACM conference on Internet measurement conference, ACM*, Boston, Massachusetts, USA, pp. 537-550, 2012. Article (CrossRef Link).

[11] J. Wu, J.H. Wang, J. Yang, "CNGI-CERNET2: an IPv6 deployment in China," *SIGCOMM Comput. Commun. Rev.*, 41 (2011) 48-52. Article (CrossRef Link).

[12] L. Zhang, H. Wang, S. Zhong, "A measurement study on BitTorrent traffic behaviors over IPv6," in *Proc. of 2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE)*, pp. 354-357, 2012. Article (CrossRef Link).

[13] W. Shen, Y. Chen, Q. Zhang, Y. Chen, B. Deng, X. Li, G. Lv, "Observations of IPv6 traffic, in: ISECS International Colloquium on Computing, Communication," *Control, and Management (CCCM 2009)*, pp. 278-282, 2009. Article (CrossRef Link).

[14] N. Ao, C. Chen, "Understanding IPv6 user performance on private BT system," in *Proc. of 4th IET International Conference on Wireless, Mobile & Multimedia Networks (ICWMMN 2011)*, IET, pp. 288-293, 2011. Article (CrossRef Link).

[15] C. Çiflikli, A. Gezer, A. Tuncay Özşahin, Ö. Özkasap, "BitTorrent packet traffic features over IPv6 and IPv4," *Simulation Modelling Practice and Theory*, 18 (2010) 1214-1224. Article (CrossRef Link).

[16] C. Çiflikli, A. Gezer, A.T. Özşahın, "Packet traffic features of IPv6 and IPv4 protocol traffic," *Turkish Journal of Electrical Engineering & Computer Sciences*, 20 (2012) 727-749. Article (CrossRef Link).

[17] F. Li, C. An, J. Yang, J. Wu, H. Zhang, "A study of traffic from the perspective of a large pure IPv6 ISP," *Computer Communications*, 2013. Article (CrossRef Link).

[18] R. Pang, V. Yegneswaran, P. Barford, V. Paxson, L. Peterson, "Characteristics of internet background radiation," in *Proc. of Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*, ACM, Taormina, Sicily, Italy, pp. 27-40, 2004. Article (CrossRef Link).

[19] Q. Wang, Z. Chen, C. Chen, "Darknet-Based Inference of Internet Worm Temporal Characteristics," *IEEE Transactions on Information Forensics and Security*, 6 (2011) 1382-1393. Article (CrossRef Link).
[20] N. Brownlee, "One-Way Traffic Monitoring with iatmon," in *Proc. of Passive and Active Measurement*, Springer Berlin Heidelberg, pp. 179-188, 2012. Article (CrossRef Link).
[21] E. Glatz, X. Dimitropoulos, "Classifying Internet One-way Traffic," *Perform. Eval. Rev.*, 40 (2012) 417-418. Article (CrossRef Link).
[22] M. Ford, J. Stevens, J. Ronan, "Initial Results from an IPv6 Darknet," in *Proc. of International Conference on Internet Surveillance and Protection (ICISP'06)*, pp. 13-13, 2006. Article (CrossRef Link).
[23] E. Kohler, J. Li, V. Paxson, S. Shenker, "Observed Structure of Addresses in IP Traffic," *IEEE/ACM Transactions on Networking*, 14 (2006) 1207-1218. Article (CrossRef Link).
[24] A. Lakhina, K. Papagiannaki, M. Crovella, C. Diot, E.D. Kolaczyk, N. Taft, "Structural analysis of network traffic flows," in *Proc. of Proceedings of the joint international conference on Measurement and modeling of computer systems*, ACM, New York, NY, USA, pp. 61-72, 2004. Article (CrossRef Link).
[25] X. Guan, T. Qin, W. Li, P. Wang, "Dynamic feature analysis and measurement for large-scale network traffic monitoring," *Trans. Info. For. Sec.*, 5 (2010) 905-919. Article (CrossRef Link).
[26] Q. Li, T. Qin, X. Guan, Q. Zheng, "Empirical Analysis and Comparison of IPv4-IPv6 Traffic: A Case Study on the Campus Network," in *Proc. of Proceedings of 18th IEEE International Conference on Networks (ICON'12)*, Singapore, pp. 395-399, 2012. Article (CrossRef Link).
[27] B. Li, J. Springer, G. Bebis, M. Hadi Gunes, "A survey of network flow applications, Journal of Network and Computer Applications," 36 (2013) 567-581. Article (CrossRef Link).
[28] A. Lakhina, M. Crovella, C. Diot, "Mining anomalies using traffic feature distributions," in *Proc. of Proceedings of the 2005 conference on Applications, technologies, architectures, and protocols for computer communications*, ACM, Philadelphia, Pennsylvania, USA, pp. 217-228, 2005. Article (CrossRef Link).
[29] P. Casas, S. Vaton, L. Fillatre, I. Nikiforov, "Optimal volume anomaly detection and isolation in large-scale IP networks using coarse-grained measurements," *Computer Networks*, 54 (2010) 1750-1766. Article (CrossRef Link).
[30] H. Jiang, Z. Ge, S. Jin, J. Wang, "Network prefix-level traffic profiling: Characterizing, modeling, and evaluation," *Computer Networks*, 54 (2010) 3327-3340. Article (CrossRef Link).
[31] E.F. Harrington, "Measuring Network Change: Renyi cross entropy and the second order degree distribution," in *Proc. of Proceedings of passive and active measurement conference*, 2006.
[32] N. Ye, S. Vilbert, Q. Chen, "Computer intrusion detection through EWMA for autocorrelated and uncorrelated data," *IEEE Transactions on Reliability*, 52 (2003) 75-82. Article (CrossRef Link).
[33] G. Nychis, V. Sekar, D.G. Andersen, H. Kim, H. Zhang, "An empirical evaluation of entropy-based traffic anomaly detection," in *Proc. of Proceedings of the 8th ACM SIGCOMM conference on Internet measurement*, ACM, pp. 151-156, 2008. Article (CrossRef Link).
[34] P. Haag, "Watch your Flows with NfSen and NFDUMP," in *Proc. of 50th RIPE Meeting*, 2005.
[35] C. Estan, G. Varghese, "New directions in traffic measurement and accounting: Focusing on the elephants, ignoring the mice," *ACM Trans. Comput. Syst.*, 21 (2003) 270-313. Article (CrossRef Link).
[36] T. Liu, X. Guan, Q. Zheng, Y. Qu, "A new worm exploiting IPv6 and IPv4-IPv6 dual-stack networks: experiment, modeling, simulation, and defense," *IEEE Network*, 23 (2009) 22-29. Article (CrossRef Link).
[37] N. Muraleedharan, "Analysis of TCP flow data for traffic anomaly and scan detection," in *Proc. of 16th IEEE International Conference on Networks (ICON'08)*, pp. 1-4, 2008. Article (CrossRef Link).
[38] T. Qin, X. Guan, W. Li, P. Wang, M. Zhu, "A new connection degree calculation and measurement method for large scale network monitoring," *Journal of Network and Computer Applications*, 2013. Article (CrossRef Link).

Qiang Li received his B.S degree in computer science from Xi'an Jiaotong University, Xi'an, China, in 2006. He is currently Ph.D. candidate with Center for Intelligent and Networked Systems, Dept. of automation in Tsinghua University, Beijing, China. His research interests include network traffic measurement and analysis.

Tao Qin received his B.S., M.S. and Ph.D. degrees in information engineering from Xi'an Jiaotong University, Xi'an, China, in 2004, 2006, 2010 respectively. He is currently an assistant professor with the Systems Engineering Institute and SKLMS Laboratory of Xi'an Jiaotong University. His research interests include internet traffic analysis, abnormal detection and traffic modeling.

Xiaohong Guan received his B.S. and M.S. degrees in automatic control from Tsinghua University, Beijing, China, in 1982 and 1985, respectively, and his Ph.D. degree in electrical engineering from the University of Connecticut, Storrs, in 1993. From 1985 to 1988, he was with the Systems Engineering Institute, Xi'an Jiaotong University, Xi'an, China. From 1993 to 1995 he was a senior consulting engineer at PG&E. From January 1999 to February 2000, he was with the Division of Engineering and Applied Science, Harvard University, Cambridge, MA. Since 1995, he has been with the Systems Engineering Institute, Xi'an Jiaotong University, where he is also currently a Cheung Kong Professor of systems engineering and the Dean of School of Electronic and Information Engineering. He is also with the Department of Automation, Tsinghua National Lab for Information Science and Technology and the Center for Intelligent and Networked Systems, TNLIST, Tsinghua University, Beijing, China. His research interests include allocation and scheduling of complex networked resources, network security, and sensor networks.

Qianghua Zheng received the B.S. degree in computer software in 1990, the M.S. degree in computer organization and architecture in 1993, and the Ph.D. degree in system engineering in 1997 from Xi'an Jiaotong University. He is a professor in the Department of Computer Science and Technology in Xi'an Jiaotong University. His research areas include multimedia distance education, computer network security, intelligent e-learning theory and algorithm. He has managed 8 research projects and authored/coauthored approximately 5 publications in these areas. He did Postdoctoral Research in Harvard University from February 2002 to October 2002 and Visiting Professor Research in Hong Kong University from November 2004 to January 2005. He got the First Prize for National Teaching Achievement, State Education Ministry in 2005 and the First Prize for Scientific and Technological Development of Shanghai City and Shaanxi Province in 2004 and 2003 respectively.