ISBN 978-981-14-4787-7
Proceedings of 2020 the 10th International Workshop on Computer Science and Engineering (WCSE 2020)
Shanghai, 19-21 June, 2020, pp. 347-352
doi: 10.18178/wcse.2020.06.052

Measure the Non-standard Port of Popular Protocols*

Chengshang Hou [1,2], Junzheng Shi [1,2 +], Gaopeng Gou [1,2], Zhen Li [1,2] and Gang Xiong [1,2]

[1] Institute of Information Engineering, Chinese Academy of Sciences, China

[2] School of Cyber Security, University of Chinese Academy of Sciences, China

Abstract. With the prosperity of the Internet, numerous network services have emerged. Many network services use non-standard, unregistered TCP/IP ports, which make the port-based firewall policy no longer effective. However, port-based firewalls are still deployed on the gateway of some local networks. As the application layer protocol increasing, network failures may increase due to the obsolete firewall policies. To make up this gap, our study aims to measure the port distribution of the popular Internet applications (e.g. WWW, P2P, multimedia). Using the flow-based traffic classification technique, we analyze the traffic collected from ISP and profile the characteristic of the protocols using non-standard ports. Our results show that a lot of application layer protocols are deployed on a wide range of non-standard ports. Particularly, both HTTP and SSL protocol, the two most popular protocols, account for a large proportion of non-standard port traffic, which is caused by the increasing web-based applications and the increasing of SSL-based encrypted traffic, respectively. Our measurement results demonstrate it is necessary to develop more fine-grained port-based firewall policies.

Keywords: traffic classification, non-standard port, port-based firewall, passive measurement

1. Introduction

Application layer protocol usually used the IANA-assigned ports [1] as service port in the past. However, as more and more network applications are deployed on the Internet, application layer protocols using non-standard ports are becoming more and more serious. Thus, network operators must have a clear understanding of the common ports for popular network services.

Newly emerged network services, especially P2P applications make network management getting more complicated. As the emergence of cloud services, a large number of enterprises deploy applications/services to cloud hosts. For example, multiple HTTP services that being deployed on same hosts will share same IP addresses but will be assigned with different port numbers [2]. The increasingly encrypted traffic also contributes to the widely employment of non-standard port. The problem results in the effectiveness of payload-based firewall. However, the accurate port-based firewall polices will be useful to improve the result of application layer protocol identification.

In this paper, we study the ISP traffic traces by identifying applications and quantify non-standard port of major Internet applications in two levels (i.e. bytes, flows). We find that there are numerous Internet applications using non-standard port from the sight of the summary of flow statistics. The major contribution of our work including:

- We quantify the proportion of protocols running on non-standard ports using real traffic traces collected from a large ISP. The results show that HTTP, SSL and P2P applications using non-standard ports account for a high proportion. In particular, the percentage of HTTP use TCP port 80 less than 50%.

[+] Corresponding author. *E-mail address*: shijunzheng@iie.ac.cn.

* This work is supported by The National Natural Science Foundation of China (No. U1636217) and The National Key Research and Development Program of China (No.2016QY05X1000 and No. 2018YFB1800200) and Key research and Development Program for Guangdong Province under grant No. 2019B010137003.

And there are some application layer protocols such as DNS and PPTP that always observed on standard port.

- We investigate the typically non-standard port for both HTTP protocol and SSL protocol. By ranking the non-standard port according to the observed traffic volumes and the number of unique IP addresses, we show the dominating port among the hundreds of non-standard ports for HTTP and SSL protocol.
- We explain why there is so much traffic on non-standard ports. We find many applications using non-standard port protocol are corresponded to specific service. For example, there are high proportion of SSL traffic using a certain non-standard port for mobile message push service (e.g. Google and Apple).

The remainder of this paper is organized as follows. we introduce the related works in Section 2. In section 3, we describe our research framework and experiments in detail. Data sets and protocol analysis is given in section 4.

2. Related Work

Many studies analyze the characteristic of the ISP traffic [3][4][5]. New trends of the Internet are revealed by measuring network traffic. For example, Labovitz et al. [6] analyze inter-domain traffic on using a large traffic show that p2p traffic is increasing. There are several of works involving the analysis of protocol ports including [7][8][9], which are more relevant to our work. Richter et al. [7] propose a packet sampling technique to collect traffic and analyze the distribution of network applications. They find that the most popular encountered non-standard ports of HTTP are 8080, 1935 and 8000, which is coincident with our results. Dainotti et al. [8] analyze the application distribution of port 80. Alock et al. [9] identify and quantify non-standard applications that make use of standard port of HTTP, HTTPS, DNS and NTP protocol. The difference between our work and other port measurement is the ports range. Previous studies focus on the limited well-known port, while our measurement expand to a wide range of non-standard ports which accounts for large fractions of traffic volume but ignored by previous port measurement studies.

The literature of traffic classification can be divided into three major methodology [5]. Port-based method is the simplest and fastest technique. Port-based classification is based on analysis the association of Internet application and port numbers. The most famous tool is CoralReef [10], developed by CAIDA, which is often used as a baseline for traffic classification.

However, P2P applications use dynamic port number to avoid detection. Payload-based methods are proposed to tackle the problem. Payload-based traffic classification take use of the invariant and unique content to identify application layer protocol. The wildly used stateful inspection firewall is developed on the basis of payload-based method.

Recently, a lot of machine learning and statistical classifiers have been developed [11][12]. Machine learning methods, which including supervised machine learning and unsupervised machine learning, learn from statistical features from network layer or transport layer. However, machine learning based traffic classification have not been applied in the large passive measurement for it lack the interpretability. In this paper, we adopt an open source payload-based classifier to identify the application layer protocols.

3. Methodology

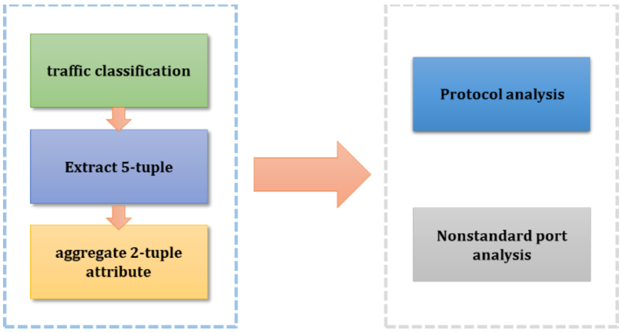

Fig. 1: The measurement framework for non-standard port analysis

First of all, we define this paper's terminology based on [13]. In order to distinguish the server and client side, we introduce the concept of bidirectional flow. Specifically, a bidirectional flow is represented by client IP address, client port, server IP address, server port, and transport layer protocol. The flow establishment time is implicit. We omit the flow establish time in order to better describe the research framework. For a TCP flow, it finishes when both communication side send TCP FIN flag. While for TCP flow without FIN flag and UDP flow, there is a maximum alive period.

Fig. 1 illustrates the main architecture of our measurement. In the first step, we first identify the application layer protocol for each flow. We use an open source classifier, nPDI[14], which has been tuned for the research network. According to the comparison of Bujlow et al. [15], nDPI is the most accurate classifier comparing with the state-of-the-art payload-based traffic classification methods. In this step, we use a five-tuple (i.e. {server IP, server Port, client IP, client Port, proto}) to represent a flow. We identify the client and server based on the direction of the first packet in a bidirectional flow. The flow attribute is record such as the number of packets transmitted, the number of bytes, and the result of the application layer protocol identification.

Secondly, we aggregate the flow according to the server IP, server port and transport layer protocol. The IP and port correspond to the protocol identified in the first step. After the aggregation, the protocol corresponding to an IP and port pair can be understood as a network service provided by a server. At the same time, we accumulate the number of packets and bytes of each service per host in this step.

The third step is to divide the data set. First, we divide the transport layer protocol data set according to the transport layer protocol TCP and UDP, and then divide the data set according to the standard port and the non-standard port. Finally, the result of protocol analysis and non-standard port analysis is presented in the next section.

4. Result

4.1. Dataset

The traffic traces were collected from an ISP of China in February 2017 for 5 days. The data volume of the traffic is roughly 30 TB and 6.2 million server IPs detected from the traffic traces. When collect the traffic trace we perform the following operation. To shield the unrelated factors, such as network scanning traffic, we ignore incomplete flows, such as flow just observed one direction. We discard the packet which does not start from a SYN flag and ignored the flow that transport layer was not TCP or UDP (e.g. ICMP, IGMP).

Table 1: The standard ports of the major protocols

TCP Protocol	**Standard ports**	**UDP Protocol**	**Standard ports**
HTTPS	443	BitTorrent	6881-6889
HTTP	80	Thunder	12345
SSH	22	Youku	4466
TELNET	23	PPLive	5041
RDP	3389	RTP	6000
POP3	110	QQ Download	1863
SMB	139,445	STUN	3478
BitTorrent	6881-6889	RADIUS	1812,1813,2865,2866
Youku	4466	Sopcast	4030
PPTP	1723	YY Voice	1146

Particularly, we consider the port 80 for HTTP. For SSL, it is difficult to select a standard port number, since many applications use the SSL protocol, such as HTTPS. We chose port 443 as the standard HTTPS port. Table 1 shows the standard ports for top 10 TCP and UDP protocols among all identified protocols.

We finally identified 72 unique protocols in this dataset. Table 2 and 3 show the number of the packets and bytes of the top 10 of TCP and UDP protocols identified from the real traces. Protocols listed in tables are sorted by the number of server IPs. Space limitation makes us unable to show all protocols. We give out the top 10 protocols on top of both TCP and UDP which account for more than 95% of total packets. The unclassified traffic is ignored because this portion of the traffic is only a small fraction and is not our research goal. In addition, since our packet was captured in a Chinese ISP, the results of application identification show high proportion of Chinese application traffic.

Table 2: Comparison of standard and non-standard ports traffic volume of major protocols on top of TCP

Protocol	**#packets**		**#bytes**	
	standard ports	**non-standard ports**	**standard ports**	**non-standard ports**
HTTPS	15 B	2.3 B	9.1 T	1.3 T
HTTP	3.4 B	9.6 B	2.6 T	4.1 T
SSH	443 M	21.5 M	122 G	15 G
TELNET	26.6 M	75.1 K	0.36 G	1.26 M
RDP	13.8 M	138 K	3.0 G	6.8 M
POP3	28.2 M	41.3 M	21.5 G	5.8 G
SMB	8.4 M	3.8 K	1.8 G	450 K
BitTorrent	131 M	121 M	58.6 G	110 G
Youku	5.8 M	830 K	5.9 G	0.97 G
PPTP	4.5 M	0	82 M	0

4.2. TCP Protocol

Table 2 shows the comparison on both traffic volumes and packets for top 10 TCP protocol by the number of servers. The SSL traffic is responsible for 45% of total traffic volume. Both HTTPS and HTTP traffic with non-standard ports account for the vast majority of the total non-standard port traffic. The detail of port distribution of the two protocols is described in the next chapter. The non-standard ports SSH protocol accounts for 15G traffic including port 2222, 222 and 443. In our data set, TELNET, RDP, SMB, and PPTP protocol show low percentage of non-standard port usage. The non-standard port of these four protocols account for a low number of packets, and bytes. The percentage of POP3 protocol on its standard port exhibits is more than the percentage of non-standard ports on the basis of the number of bytes. While the percentage of non-standard ports of POP3 protocol is more than the percentage of standard port on the basis of the number of packets. Port 25 is the non-standard port that is mostly used by POP3 protocols, which is the well-known port for SMTP. The TCP version of BitTorrent protocol accounts for 30% of total BitTorrent protocol traffic.

Table 3: Comparison of standard and non-standard ports traffic volume of major protocols on top of UDP

Protocol	**#packets**		**#bytes**	
	standard ports	**non-standard ports**	**standard ports**	**non-standard ports**
BitTorrent	260 M	1.8 B	34 G	780 G
Thunder	11.5 B	5.2 B	523 G	2.9 T
Youku	8.5 B	17.8 M	18.1 G	1.5 G
PPLive	763 M	13.5 M	29.1 G	10.9 G
RTP	1.3 B	625 M	4.3 G	355 G
QQ Download	1.6 B	2.5 M	8.5 G	1.2 G
STUN	29.3 M	84 M	1.4 G	1.6 G
RADIUS	0	886 M	0	503 G
Sopcast	0.3 M	6.5 M	1.1 G	3.4 G
YY Voice	4.9 M	4.1 M	6.2 M	0.34 G

4.3. UDP Protocol

Table 3 shows the comparison on both traffic volumes and packets for top 10 UDP protocol by the number of servers. The UDP version BitTorrent traffic show a low fraction use standard port. The alternative ports for

BT protocol include 8999, 51413, 11101, 4385. Thunder protocol is a Chinese P2P download application which is similar to BitTorrent. Thunder always use 1024-1028 and 54321 as alternative port. Youku, a Chinese video hosting and streaming application, show a high proportion use standard port 92% of total Youku traffic and 96% of total Youku packets. Another Chinese P2P streaming application, PPLive, and P2P file sharing application QQ Download show a high proportion use standard port. Because our dataset collected from a Chinese ISP, our results are high proportions of Chinese multimedia and P2P application. VoIP services, such as STUN and YY Voice, show a low fraction use standard port. In our data set, we do not observe RADIUS protocol use standard port. RADIUS frequently adopt non-standard ports include 7273, 4500, 17777 and 7274.

4.4. Measurement Result and Analysis

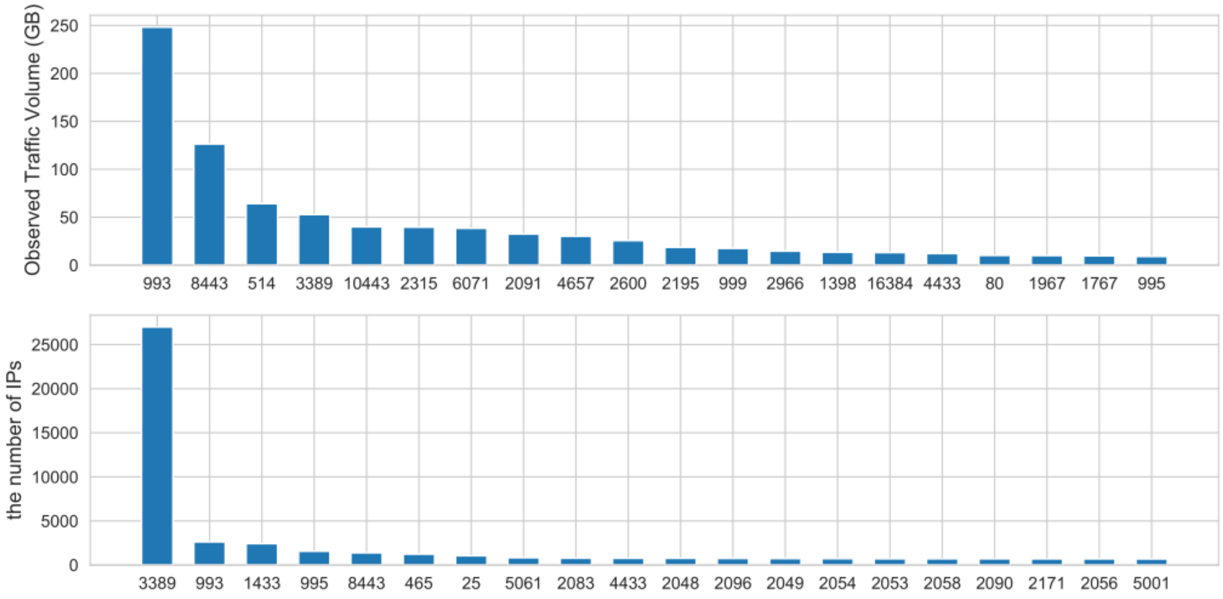

Fig. 2: SSL port usage distribution (except port 443)

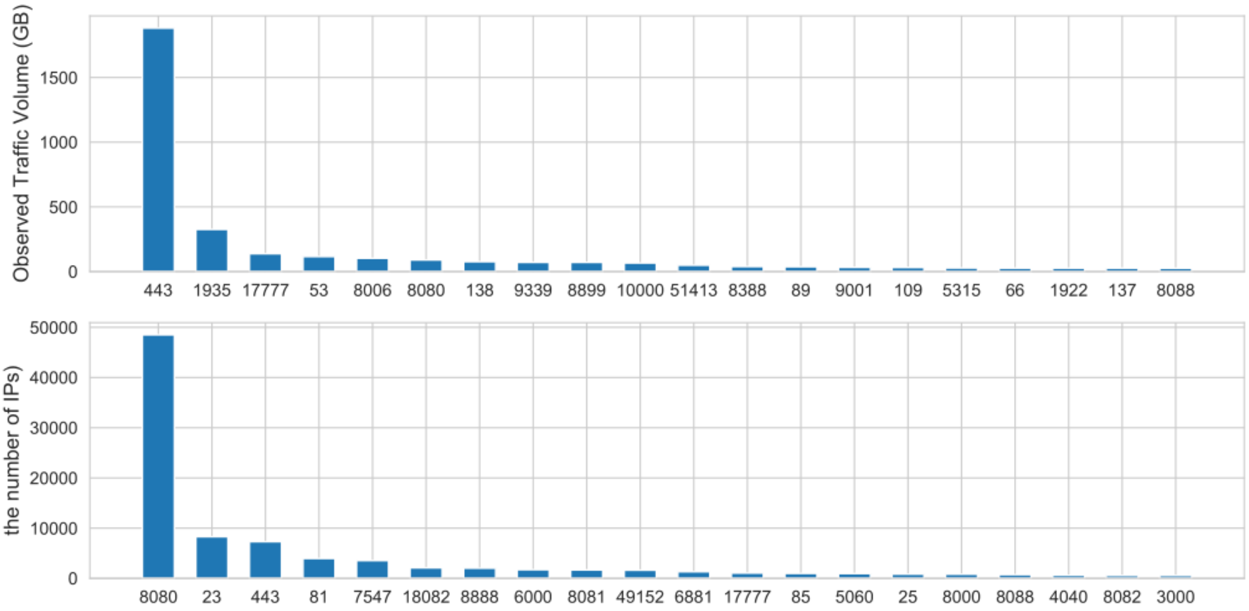

Fig. 3: HTTP port usage distribution (except port 80)

Fig. 2 lists the top twenty ports (except port 443) used by the SSL protocol, ranking based on the number of server IP and traffic volume. We observe a large number of servers using the SSL 3389 port, which is the port that the remote desktop . The 993, 995, 25, 465 ports ranked in the top 10 are related to E-mail protocols. 8443 and 1443 rank 3rd and 5th as the default SSL port of the application server such as Tomcat. Port 5061 is the default port for SIP over TLS. At the same time, we observed protocol with non-standard port is used as specific purpose, such as hosting and message notification. For example, port 2083 is used for the control port of Cpanel. We observed the Apple push service and the Google Cloud Message service on ports 5223 and 5228 respectively.

Fig. 3 lists the top 20 ports used by HTTP except for port 80, ranking based on the number of server IP and traffic volume. The port 8080 are always treated as an alternative of well-known port 80 for HTTP. However, from the view of traffic volume, port 443 is the dominating port among all nonstandard port. Surprisingly, we identified the HTTP protocol on port 23, which is the standard port of TELNET.

5. Conclusion

In conclusion, we investigate non-standard port usage on the traffic gathered from an ISP. By identifying Internet applications and quantify non-standard port of the popular Internet protocols and applications, we outline the characteristic of the non-standard port usage of popular protocols (e.g. HTTP, SSL, P2P protocol). We find many protocols heavily use non-standard port. Especially, the non-standard ports of HTTP and SSL protocol account for a huge proportion.

6. References

[1] IANA: Service name and transport protocol port number registry (2019), https://www.iana.org/assignments/service-names-port-numbers/service-namesport-numbers.xhtml

[2] Pop, A.: Delignat-Lavaud, A. and Bhargavan, K., 2015, May. Network-based origin confusion attacks against HTTPS virtual hosting. In Proceedings of the 24th International Conference on World Wide Web (pp. 227-237). (2015)

[3] Carela-Espan˜ol, V., Barlet-Ros, P., Bifet, A., Fukuda, K.: A streaming flow-based technique for traffic classification applied to 12+ 1 years of internet traffic. Telecommunication Systems 63(2), 191–204 (2016)

[4] Pujol-Gil, E., Richter, P., Feldmann, A.: Understanding the share of ipv6 traffic in a dual-stack ISP. In: PAM 2017 (2017)

[5] Benson, T., Akella, A., Maltz, D.A.: Network traffic characteristics of data centers in the wild. In: Internet Measurement Conference (2010)

[6] Labovitz, C., Iekel-Johnson, S., McPherson, D., Oberheide, J., Jahanian, F.: Internet inter-domain traffic. In: SIGCOMM (2010)

[7] Richter, P., Chatzis, N., Smaragdakis, G., Feldmann, A., Willinger, W.: Distilling the internet's application mix from packet-sampled traffic. In: PAM 2015 (2015)

[8] Dainotti, A., Gargiulo, F., Kuncheva, L.I., Pescap`e, A., Sansone, C.: Identification of traffic flows hiding behind tcp port 80. 2010 IEEE International Conference on Communications pp. 1–6 (2010)

[9] Alcock, S., Mo¨ller, J.p., Nelson, R.: Sneaking Past the Firewall: Quantifying the Unexpected Traffic on Major TCP and UDP Ports Shane. Proceedings of the 2016 ACM on Internet Measurement Conference IMC '16 pp. 231–237 (2016).

[10] CAIDA: Coralreef (2019), https://www.caida.org/tools/measurement/coralreef/

[11] Grimaudo, L., Mellia, M., Baralis, E., Keralapura, R.: Select: Self-learning classifier for internet traffic. IEEE Transactions on Network and Service Management 11(2), 144–157 (2014)

[12] Finamore, A., Mellia, M., Meo, M., Rossi, D.: Kiss: Stochastic packet inspection classifier for UDP traffic. IEEE/ACM Transactions on Networking 18, 1505–1515 (2010)

[13] Dainotti, A., Pescap`e, A., Claffy, K.C.: Issues and future directions in traffic classification. IEEE Network (2012)

[14] Deri, L., Martinelli, M., Bujlow, T., Cardigliano, A.: nDPI: Open-source high-speed deep packet inspection. 2014 International Wireless Communications and Mobile Computing Conference (IWCMC), 617-622 (2014)

[15] Bujlow, T., Carela-Espan˜ol, V., Barlet-Ros, P.: Independent comparison of popular DPI tools for traffic classification. Computer Networks 76, 75–89 (2015)