A Survey of Techniques for Internet Traffic Identification and Classification

Mingwei Wei, *Member, IEEE*

***Abstract*—The techniques for Internet traffic identification and classification are developed rapidly in recent years, as it widely used in network management, monitor, design, security and research. In the past decade, the traffic identification and classification techniques have been evolved along with development of Internet protocols and applications, and many approaches have been proposed to optimize these techniques. Nowadays, traffic measurement remains one of the hot areas in network research. This is mostly based on the ever increasing network bandwidth, the growth number of network users, the constantly sophisticated applications and the development of technique about confusing traffic identification and classification. In this paper, we present popular traffic identification and classification techniques, include port-based, payload-based, flow-based and host-based, then analyze each technique from challenge aspect and make some remarks and recommendations that contribute to optimize traffic measurement.**

***Index Terms*—traffic identification, traffic classification, challenges, application detection, recommendations.**

I. Introduction

With the development of Internet technology and the advent of the era of mobile Internet, our life has been inseparable from the Internet nowadays. According to the 35th statistical report on development of Internet in China published by CNNIC [1], the Internet users of China have reached 6.49 hundred million by the end of 2014.

As shown in Fig. 1, the Internet population of China has been increasing rapidly in recent years, almost half of Chinese people are using Internet for work or daily life.

Fig. 1. Internet population of China

Mingwei Wei is with the Department of Computer Science and Technology, Beijing Institute of Technology, Beijing, 100081 China (phone: 152-0161-3264; e-mail: weimw0417@163.com.)

Manuscript received July 25, 2015;

As shown in Fig. 2, the Internet Penetration of China is higher and higher over the years, more and more people felt the charm of the Internet, and the Internet has penetrated into all walks of life of people.

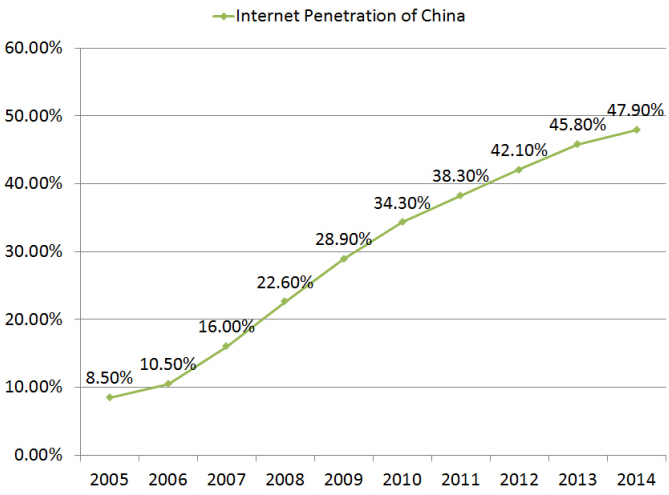

Fig. 2. Internet population of China

Increasingly serious network security problems is in contrast with the rapid development of Internet technology. In the past few years, constantly exposures of network security event make the network security problem get more and more attention. If the security problem is not solved, especially in Internet Finance and Internet Payment, the further development and employment of Internet technology will be severely impeded.

Traffic is the carrier of the Internet. In large number of Internet traffic, a wide variety of malicious traffic is hidden. These malicious traffic carrying viruses, trojans and worms threats the security of Internet. It not only affects the network service provide's service quality, but also threatens the Internet user's privacy and data security, even the national security. So how to find malicious traffic and intercept them is the challenge of Internet security.

Traffic identification and classification is a technique that can detect applications the very traffic corresponded from mass of traffic. Internet traffic identification and classification systems are deployed in gateway normally, it monitors traffic flows though gateway and intercept malicious traffic to ensure smooth operation of network. Traffic identification and classification is basic of traffic control, this technique is widely used in network audit, content audit and intrusion detection, it plays a important role in increasing network management efficiency and guaranteeing network security.

The traditional traffic identification techniques concentrate on content of traffic packet, so it is only to the recognition of unencrypted traffic effectively. On this occasion, malicious users transfer illegal data with a safe data transmission protocol becomes possible. Therefore, to identify legal and illegal data from encrypted traffic and classify traffic according to its type and source becomes a new challenge in network and Internet security.

The techniques for Internet traffic identification and classification have been evolved along with development of Internet protocols and applications, and many techniques have been proposed, including port-based techniques [2], payload-based techniques [3], flow-based techniques [4], host-based techniques [19] and graph-based techniques. Some of them have been maturely and widely deployed in the current network, and some of them are still under researching. But all the existing techniques still have their critical limitations and issues. In addition, the existing techniques are facing more and more challenges as Internet protocols are becoming safer and applications are becoming more sophisticated. In this paper, we first present popular traffic identification and classification techniques, then do in-depth analysis for them, especially their issues and challenges, and address some recommendations that can improve performance of techniques for Internet traffic identification and classification at last.

The rest of the paper is organized as follows. Section II explains port-based technique and analyzes its issues. Section III explains payload-based technique and presents its drawbacks. Section IV and Section V focus on flow-based and host-based techniques and dwell on their challenges. Then it is followed by section VI with the general challenges of traffic identification and classification. Section VII makes some final remarks on traffic analysis and provides some recommendations for solving the current issues. Finally, we conclude the paper in section VIII.

II. Analysis of Port-based Techniques

The port-based technique is used to identify application according to TCP/UDP port number on transport layer protocol. In the early stage of Internet, applications use specific port to set up communication. Most of them used well-known port numbers assigned by IANA (The Internet Assigned Numbers Authority). Table I shows the most popular well-known port number and its corresponding services and protocols.

TABLE I
POPULAR WELL-KNOWN PORT NUMBERS

APPLICATION/SERVICE	PROTOCOL	PORT-NO
THUNDER	TCP/UDP	80,8000,8888
QQ	UDP	4000
SSH	TCP	22
FTP	TCP	20,21
WEB	TCP	80,443
TELNET	TCP	23
TOMCAT	TCP	8080

To classify these applications or services, the port-based techniques only need to check source port number of IP data packet. It finished tasks perfectly, as it assumes that most applications or services use well-known fixed TCP/UDP port number. However, with the popularity of port jump technique, the port-based techniques lost their effectiveness. The so called port jump technique refers to the applications use host port for communication without fixing port number or change port in the process of data transmission. For example, utorrent can choose the random non-hold port for data transmission automatically when it has just started, this approach makes port-based technique failure. Although port-based traffic identification and classification technique is easy to implement, it's also easy to circumvent. So it is rarely to use as a main traffic identification technique now, often as an auxiliary means of other traffic identification and classification technology.

III. Analysis of Payload-based Techniques

In order to improve the low identification efficiency caused by the port jump technique, payload-based technique was proposed. Deep Packet Inspection (DPI) is one of typical approach in payload-based techniques, it compares data content and characteristics of rule set which is built in advance with feature matching algorithm, and set results of matching as the basis of identification. The item in characteristics of rule set includes type of data and feature string. Table II shows the popular payload characteristics of applications.

TABLE II
PAYLOAD CHARACTERISTICS

PROTOCOL	PAYLOAD
HTTP	'GET' 'PUT' 'POST'
SSH	'SSH'
IRC	'USERHOST'
Kazza	'Z-Kazza'
BitTorrent	'\x13BitTorrent protocol'
QQ Voice	'SIP/user-agent: Tencent-VQ'
Thunder	'\x00\x00\x00\x16\x00\x00\x00\x6a\x01'
eMule	'\xe3,\xc5,\xd4,\xe4,\xe5,\xf1'

The precision and efficiency of DPI is decided by integrity of characteristics of rule set and feature matching algorithm, as shown in Fig. 3. The feature matching algorithm includes one-mode and multi-mode. One-mode refers to one scan can only match one feature string, such as Knuth-Morris-Pratt algorithm [6] and Boyer-Moore algorithm [7]. Multi-mode refers to one scan can match a group of feature strings, such as Aho-Corasick algorithm [8], Commentz-Walter algorithm [9] and Aho-Corasick Boyer-Moore algorithm [10].

Although the payload-based technique can solve the issue of random port, they cannot deal with the encryption protocols, because the signatures of these types of protocols can hardly be found. Generally, the main challenges resulting in issues are the encryption protocols, we list issues as follows.

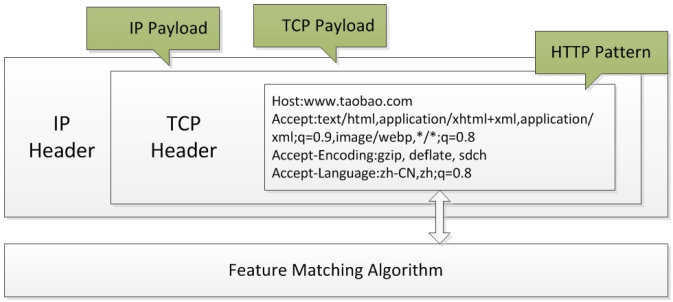

Fig. 3. Principle of Feature Matching Algorithm

A. Fail in Encryption Protocol

The encryption protocols get more and more attention in current Internet, as people have kept a watchful eye on the individual privacy. More and more applications start to use the encryption protocols (e.g. SSL/TLS) for communication, such as Skype, Alipay, Evernote and Zhihu. The signatures of packet payload can hardly be found as the specifications of these proprietary protocols remain private. Therefore, the payload-based techniques lose their effectiveness when identifying and classifying such kinds of applications employing SSL/TLS protocols.

B. Low Efficiency

With the development of Internet bandwidth, the era of Terabit bandwidth has arrived. Nowadays, the traffic identification and classification systems usually have to handle Gigabits or Terabits of data per second, which is a critical challenge task for payload-based techniques. Due to payload-based techniques have to compare every packet content with characteristics of rule set until the protocol of the flow is determined, it needs higher processing capability to handle packets in high speed network.

Furthermore, as more and more new protocols appear, the size of characteristics of rule set becomes larger and larger, so the payload-based techniques have to compare more signatures or regular expressions, the processing efficiency of payload-based techniques will drop rapidly.

IV. ANALYSIS OF FLOW-BASED TECHNIQUES

As the protocols and applications are continuing to evolve, port-based and payload-based techniques lose their effectiveness to the new coming protocols, the flow-based techniques are proposed under this circumstances.

Just as its name implies, the flow-based technique is built on flows formed by a series of data packets. It works by seeking and recording general law from data flow in a period of time. For example, through the observation, we found most of P2P applications are using TCP protocol and UDP protocol at the same time. Besides, it transmits control information by UDP and file content by TCP. We use this law as the basis to identify the P2P application. Different from the two techniques mentioned in section II and section III, the flow-based technique only analyzes the data flow and it's indifferent to the content of packet. Therefore, the flow-based technique is suitable for common Internet traffic identification and classification as well as encrypted traffic.

Most flow-based techniques use Machine Learning techniques as their identification or classification algorithms, such as Bayesian methods [11] and SVM Support Vector Machine [12]. Most flow-based techniques contain two main stages: training stage and classifying stage, as shown in Fig. 4 and Fig. 5 respectively. At the training stage, the flow-based techniques use features extracted from the training data set to train identification and classification model.

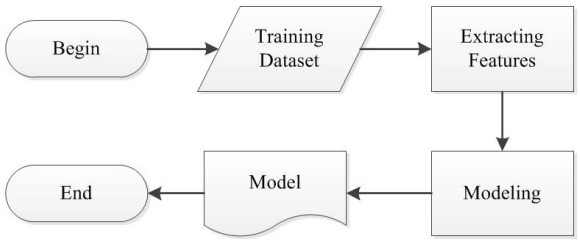

Fig. 4. Process of Training Stage in Flow-based Techniques

At the identification and classification stage, the flow-based techniques capture packets at first, then divide and refactor these packets to form flow. After that, extracting features to matching models obtained from training stage to determine the type of protocols and applications. In this way, it completes the traffic identification and classification task.

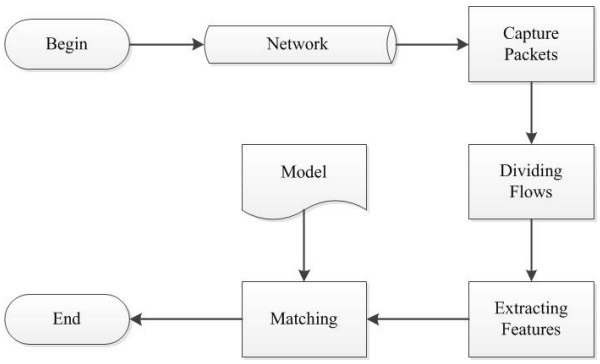

Fig. 5. Process of Identification or Classification Stage in Flow-based Techniques

In 2010 Yanfeng Sun et al. proposed a new identification method based on size of packets and clustering algorithm [15]. They studied the methods of network application identification based on the transition pattern of payload length during the start up phase of the communication. These methods didn't need to analyze all the packets, so they could identify the traffic at the early age of the communication. But when analyzing more packets, they met the contents communication problem [16]. To solve this problem and adapt the change of the packets, they consulted the inverse method, and used the inverse of the size of the packets. And associating with the improved clustering algorithm K-medoids, they finally improved the validity of our identification method.

The technique includes offline training stage and online identify stage, as shown in Fig. 6. In training stage, it cuts out and counts the size of the first N packets of every flow, inverses

the size and vectorizes them, then clusters the vector set with K-medoids algorithm to get the reference vector set that can represent certain application. In identify stage, it does the same work on testing data set to vectorize them and calculate the distance between every testing data vector and reference center vector, then identify the testing data with K-medoids and the output is application type.

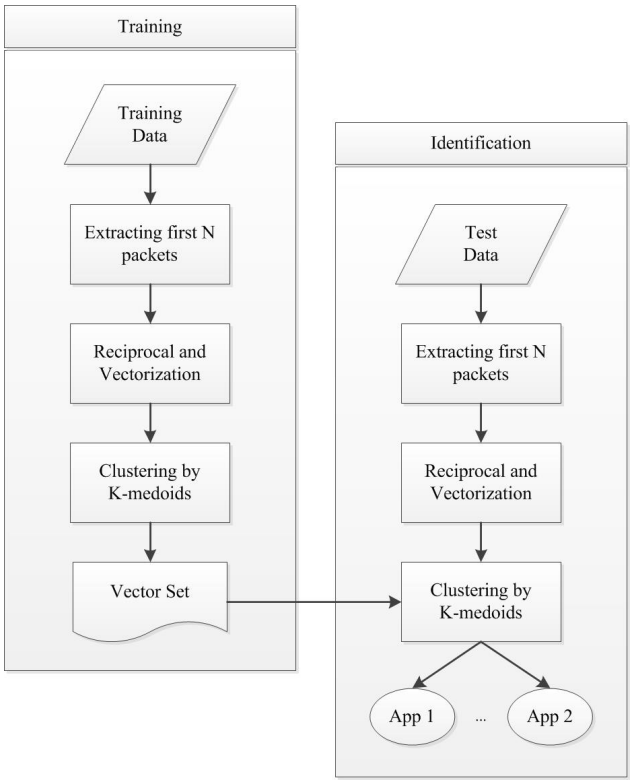

Fig. 6. Process of A New Method Based on Size of Packets and Clustering

The flow-based techniques are the hot spot in traffic identification and classification research area. Although it can solve the issues of proprietary encryption protocols, the flow-based techniques are far away from being usable due to the following reasons.

A. Selection of Training Set

In order to provide more training information, the training set usually contains traffic generated by various protocols. However, at the training stage, the target protocol can be hardly highlighted, if its traffic only occupy a small proportion in the training set. Therefore, we can increase proportion of certain protocol traffic artificially. But the artificial training set can't describe the real network environment. So building the appropriate training set is a tradeoff process. The training set should be built based on the application context and network environment.

Generally, it can be said that the training set plays an important part in flow-based techniques, we should do our best to raise the accuracy of model that got by training set. Far better to build training set in real network and consider identification and classification context at the same time.

B. Selection of Flow Statistical Features

Many statistical features can be used to describe the flow, Moore et al. listed more than 200 flow statistical features for flow-based techniques [17]. Obviously, we can't use all 200 features to describe flow. First, many ML models are sensitive to the dimensionality of feature vector. If all the proposed 200 flow features are used to train ML classification models, the dimensionality will be too high to classify traffic effectively. Second, due to the different protocol specifications, different protocols may have different measures in different flow statistical features. Some features maybe become meaningless, even make classification results worse.

Therefore, one of the most important challenge of flow-based technique is how to select the proper flow statistical features.

C. Uniqueness Proof of Flow Statistical Features

In [18], Xue et al. realized some flow-based approaches, including SVM and Bayesian, and deployed them in the backbone network to classify SSL/TLS traffic. According to testing in lab, these two models can obtain more than 95% precision. However, when they deployed them in the backbone network, in which 1 million flows are needed to classify within a second in one server, the precision of the two models reduce to less than 5%. Through data analysis, they found that there are too many non-SSL/TLS flows that have the similar statistical features as SSL/TLS flow in the backbone network. In other words, the selected flow statistical features for classifying SSL/TLS traffic are not unique.

D. Selection of Identification and Classification Models

Currently, there are thousands of models available and various models are proposed constantly. Admittedly, some models have been improved for traffic identification and classification, such as Bayesian, SVM and Decision Tree. As all the models share the same theoretical foundation, there is no statement that which model is better, but we can find the more suitable one. We summarize that there are two factors affect the performance of models, one is the selection of statistical features, the other is the purpose of traffic classification. For example, if the target traffic only occupies small part of background traffic, SVM would be the suitable model. But if the task is to identify traffic generated by multiple protocols, C4.5 would be the suitable one. So, how to select a suitable model according to the traffic characteristics should be taken into account.

E. Constantly Complicated Applications

As applications become more and more complicated, the flow-based techniques are facing more and more challenges. Currently, many applications such as QQ, use both TCP and UDP as their transport layer protocols at the same time. It brought the flow-based techniques a huge problem, because most existing flow-based techniques focus on one protocol flows, they can't deal with multiple transport protocols simultaneously. Besides, many applications employ multiple

flows to obtain services from several servers or peers. But existing flow-based techniques don't explicitly explore such information. Therefore, how to identify and classify Internet traffic completely and accurately is another mountain need to be conquered.

V. ANALYSIS OF HOST-BASED TECHNIQUES

The host-based technique refers to set up models from behavioral feature of applications. The most representative one is Karagianni et al. proposed in 2005 called BLINC [19]. BLINC set up the fingerprint model by observing features of IP address and port number when one application communicated with others. It builds up fingerprint library includes many fingerprint models of applications. When observing one application's IP address and port number, we can match these information to every fingerprint model, and mark the application which corresponds model has highest matching-degree as the identification result. BLINC only need to observe and record source IP address, destination IP address, source port and destination port to run method, without concern about content of port and payload. But this method needs a period of time to do some data statistication, so real-time identification and classification will be a problem.

Compare to flow-based techniques, the host-based techniques also contain two main stages: training stage and classifying stage, as shown in Fig. 7 and Fig. 8. In fact, these techniques also employ statistical features as their input, while the major difference from flow-based techniques is that the hostbased techniques employ the features abstracted from the multi-flows. To obtain the features from multi-flows, the first step is to integrate flows. It is obvious that integrating flows based on a host is most feasible way. This is why we call these techniques as host-based.

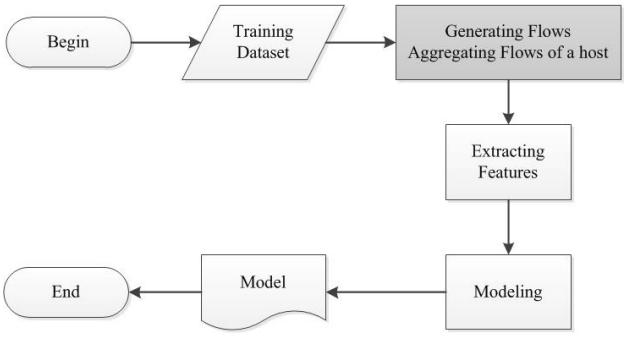

Fig. 7. Process of Training Stage in Host-based Techniques

More specifically, in 2014 Maciej proposed stochastic fingerprints for application traffic flows conveyed in Secure Socket Layer/Transport Layer Security (SSL/TLS) sessions [13]. The fingerprints are based on first-order homogeneous Markov chains for which it identifies the parameters from observed training application traces. As the fingerprint parameters of chosen applications considerably differ, the method results in a very good accuracy of application discrimination and provides a possibility of detecting abnormal SSL/TLS sessions.

We consider discrete-time random variable X_t for any $t = t_0, t_1, ..., t_n \in T$. It takes values $i_t \in \{1, ..., s\}$, where it is

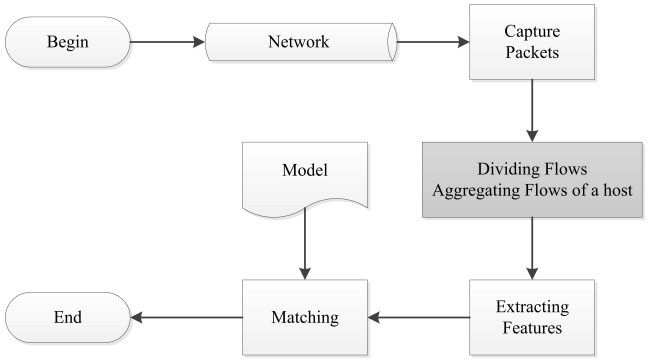

Fig. 8. Process of Identification or Classification Stage in Host-based Techniques

either an SSL/TLS message type (e.g. 22:2) or a sequence of the SSL/TLS message types transmitted in a single TCP segment (e.g. 22:11,22:14).

We assume that X_t is a first-order Markov chain [14]:

$$\begin{aligned} P(X_t = i_t | X_{t-1} = i_{t-1}, X_{t-2} = i_{t-2}, ..., X_1 = i_1) \\ = P(X_t = i_t | X_{t-1} = i_{t-1}). \end{aligned} \tag{1}$$

We further assume that the Markov chain is homogeneous, i.e. a state transition from time $t-1$ to time t is time-invariant:

$$\begin{aligned} P(X_t = i_t | X_{t-1} = i_{t-1}) = P(X_t = j | X_{t-1} = i) \\ = p_{i-j}, \end{aligned} \tag{2}$$

with the transition matrix [14]:

$$P = \begin{bmatrix} p_{1-1} & p_{1-2} & \cdots & p_{1-s} \\ p_{2-1} & p_{2-2} & \cdots & p_{2-s} \\ \vdots & \vdots & \ddots & \vdots \\ p_{s-1} & p_{s-2} & \cdots & p_{s-s} \end{bmatrix}, \tag{3}$$

where: $\sum_{j=1}^{n} p_{i-j} = 1$. We denote by:

$$Q = \begin{bmatrix} q_1, & q_2, & \cdots, & q_s \end{bmatrix}, \tag{4}$$

the ENter Probability Distribution (ENPD), where $q_i = P(X_t = i)$ at time t_0, and we define:

$$W = \begin{bmatrix} w_1, & w_2, & \cdots, & w_s \end{bmatrix}, \tag{5}$$

as the EXit Probability Distribution (EXPD), where w_i represents the probability that the session finishes when it is in state i at time t_n. Note that both probability distributions are independent of the Markov chain - they provide the probabilities to enter and quit the Markov chain. In traditional Markov chain models, there is an initial state and one or several absorbing states. In our case, ENPD defines the probability to enter one of the state of the Markov chain and EXPD gives the probability of quitting the Markov chain from any of its states. Based on these definitions, the probability that a sequence of state $X_1, ..., X_T$ representing a single SSL/TLS session occurs is as follow:

$$P(\{X_1, ..., X_T\}) = q_{i1} \times \prod_{t=2}^{T} p_{i_{t-1}} \times w_{iT}. \tag{6}$$

The resulting probability indicates how a given SSL/TLS sequence of message types during a session is close to a model

of an application flow: a larger value means that the SSL/TLS session is closer to the model. To illustrate the process of the fingerprint creation, consider the following examples of the message sequences observed during SSL/TLS sessions in a training dataset composed of only three server-side SSL/TLS flows of the PayPal application traffic:

$22:2-22:11,22:14-20:,22:-23:$
$22:2,20:,22:-23:$
$22:2-22:11,22:14-20:,22:-23:-23:-21:$

There are 6 different Markov states in the example. The transition probability between states is derived from frequencies observed in the sequences, e.g. $P_{22:2-22:11,22:14} = 1$, while $P_{23:-23:} = 0.5$. The ENPD vector is composed of two nonzero elements, namely $P_{22:2} = 0.67$ and $P_{22:2,20:,22:} = 0.33$, whereas the EXPD vector also contains two non-zero elements $P_{23:} = 0.67$ and $P_{21:} = 0.33$. The probabilities are the parameters of the Markov chain fingerprint for the PayPal traffic. Based on the model, we can find the probability that an observed SSL/TLS session conveys the PayPal application traffic (cf. Eq. 6). The probability that the following sequence of SSL/TLS message types:

$22:2,20:,22:-23:-23:-23:$

is a PayPal flow is equal to: $P(\{X_1,...,X_4\}) = 0.055$. In comparison, the probability computed from the Twitter fingerprint model (cf. Fig. 9) is equal to 0.003%, whereas the probability computed from the Skype fingerprint model (cf. Fig. 10) is equal to 0.

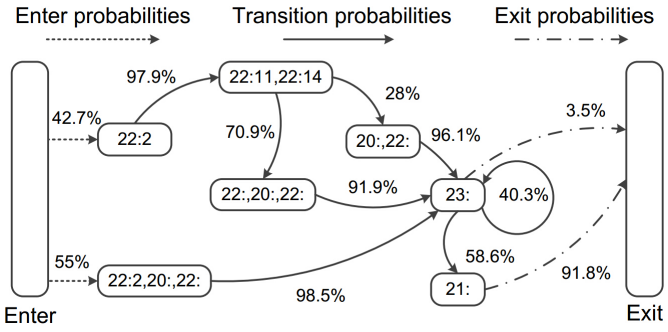

Fig. 9. Parameters of the Fingerprint for Twitter

Even though the host-based techniques are most advanced nowadays, they can classify all the traffic generated by an application, they are still suffering from the following issues.

A. Threaten from Big Data

The host-based techniques maintain the host information of either inside or outside of gateway, even both, in order to obtain enough features from multi-flows. If the host-based techniques are deployed in backbone network, there will be hundreds of millions of hosts that are needed to be monitored. And even each host in backbone network will generate plenty

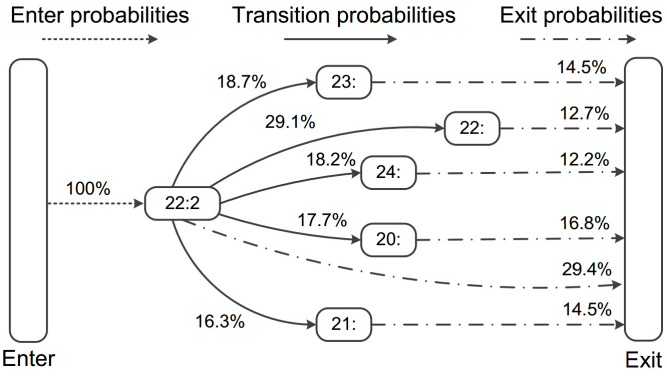

Fig. 10. Parameters of the Fingerprint for Skype

of traffic. So the host-based identification and classification system will run out of computing and storage resource, and break down in the end.

B. Unknown New Applications

As the network and protocols are developing rapidly, there are thousands of new applications published constantly, and they are more and more complicated. The host-based techniques are based on the existing applications, that is to say, with regard to new applications, there is no corresponding model in our library, we can't identify or classify that new application traffic.

C. Noise

VI. Conclusion

The conclusion goes here.

Appendix A
Proof of the First Zonklar Equation

Appendix one text goes here.

Appendix B

Appendix two text goes here.

Acknowledgment

The authors would like to thank...

References

[1] CNNIC: Statistics Report of Development of China Internet Network. (Jan 2015). [Online]. Available: http://www.cnnic.cn/hlwfzyj/hlwxzbg /201502/P020150203551802054676.pdf
[2] J. McPherson, K.-L. Ma, P. Krystosk, T. Bartoletti, and M. Christensen, "Portvis: a tool for port-based detection of security events," in *Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security*, pp. 73–81, ACM, 2004.
[3] K. Wang and S. J. Stolfo, "Anomalous payload-based network intrusion detection," in *Recent Advances in Intrusion Detection*, pp. 203–222, Springer, 2004.
[4] A.-S. Kim, H.-J. Kong, S.-C. Hong, S.-H. Chung, and J. W. Hong, "A flow-based method for abnormal network traffic detection," in *Network operations and management symposium, 2004. NOMS 2004. IEEE/IFIP*, vol. 1, pp. 599–612, IEEE, 2004.
[5] T. Karagiannis, K. Papagiannaki, and M. Faloutsos, "Blinc: multilevel traffic classification in the dark," in *ACM SIGCOMM Computer Communication Review*, vol. 35, pp. 229–240, ACM, 2005.

[6] D. E. Knuth, J. H. Morris, Jr, and V. R. Pratt, "Fast pattern matching in strings," *SIAM journal on computing*, vol. 6, no. 2, pp. 323–350, 1977.
[7] R. S. Boyer and J. S. Moore, "A fast string searching algorithm," *Communications of the ACM*, vol. 20, no. 10, pp. 762–772, 1977.
[8] A. V. Aho and M. J. Corasick, "Efficient string matching: an aid to bibliographic search," *Communications of the ACM*, vol. 18, no. 6, pp. 333–340, 1975.
[9] B. Commentz-Walter, *A string matching algorithm fast on the average*. Springer, 1979.
[10] C. J. Coit, S. Staniford, and J. McAlerney, "Towards faster string matching for intrusion detection or exceeding the speed of snort," in *DARPA Information Survivability Conference & Exposition II, 2001. DISCEX'01. Proceedings*, vol. 1, pp. 367–373, IEEE, 2001.
[11] T. Auld, A. W. Moore, and S. F. Gull, "Bayesian neural networks for internet traffic classification," *Neural Networks, IEEE Transactions on*, vol. 18, no. 1, pp. 223–239, 2007.
[12] B. Yang, G. Hou, L. Ruan, Y. Xue, and J. Li, "Smiler: towards practical online traffic classification," in *Proceedings of the 2011 ACM/IEEE Seventh Symposium on Architectures for Networking and Communications Systems*, pp. 178–188, IEEE Computer Society, 2011.
[13] M. Korczynski and A. Duda, "Markov chain fingerprinting to classify encrypted traffic," in *INFOCOM, 2014 Proceedings IEEE*, pp. 781–789, IEEE, 2014.
[14] M. Campbell, "Hidden markov and other models for discrete-valued time series," *Biometrics*, vol. 54, no. 1, p. 394, 1998.
[15] Y. Sun and S. Zhang, "a new identification method based on size of packets and clustering algorithm," *Telecommunications Information*, no. 2, pp. 26–28, 2010.
[16] Y. Waizumi, A. Jamalipour, and Y. Nemoto, "Network application identification based on transition pattern of packets," in *IEEE Wireless Rural and Emergency Communications Conference (WRECOM) 2007*, 2007.
[17] A. Moore, D. Zuev, and M. Crogan, *Discriminators for use in flow-based classification*. Queen Mary and Westfield College, Department of Computer Science, 2005.
[18] Y. Xue, D. Wang, and L. Zhang, "Traffic classification: Issues and challenges," in *Computing, Networking and Communications (ICNC), 2013 International Conference on*, pp. 545–549, IEEE, 2013.
[19] T. Karagiannis, K. Papagiannaki, and M. Faloutsos, "Blinc: multilevel traffic classification in the dark," in *ACM SIGCOMM Computer Communication Review*, vol. 35, pp. 229–240, ACM, 2005.

Mingwei Wei, is the corresponding author of this paper. He is working on master's degree in Network and Information Security Lab in school of computing in Beijing Institute of Technology. His tutor are Mingzhong Wang and Liehuang Zhu. His research interests include network security, business process management and traffic analysis.