A Novel Algorithm for Encrypted Traffic Classification based on Sliding Window of Flow's First N Packets

Yang Liu, Jinfu Chen, Peng Chang
Academy of Sciences
School of Cyber Security, University of Chinese
Beijing, China
e-mail: liuyang5599@163.com, chenjinfu@iie.ac.cn, changpeng@iie.ac.cn

Xiaochun Yun
Institute of Information Engineering
Chinese Academy of Sciences
Beijing, China
e-mail: yunxiaochun@cert.org.cn

***Abstract*—Network applications are getting more and more prevalent along with the development and the widespread use of encrypted network applications. However, traffic classification methods may need to be improved to realize more stable classification in a more sufficient way. Here, we proposed a novel Sliding Window First N Packets algorithm for the encrypted network traffic classification. With this method, one could evidently reduce the flow characteristics feature dimension, as well as the number of packets in each traffic flow. The experimental results show that under a reduced dimension of encrypted traffic flow characteristics and also a reduced number of each flow data packets, average classification accuracy using the Sliding Window First N Packets algorithm we proposed is more than 95%. By using our approach, one can achieve a general increase of the traffic classification accuracy by about 3% compared with the existing methods.**

Keywords-traffic classification; protocol recognition; Window First N Packets (WFNP) algorithm; feature selection

I. INTRODUCTION

Based on the statistic data from China Internet Network Information Center, the internet users in China reached at 668 million in 2015, and the penetration rate is 48.8%. The latest report from Canadian broadband management Agency found that the global encrypted Internet traffic is on the increase, since the Snowden exposed the US National Security Agency (NSA) surveillance program. According to this report, the encrypted Internet flow rate of European users at peak times ranged from 1.47% to 6.10%. The surge of encrypted Internet traffic is not only consuming a huge proportion of network bandwidth, but also causing harm to quality services and network accounting management. Thus, improving the classification performance and the accuracy of the encryption application traffic is of central importance.

In previous studies, some researchers have used port and data packet load to classify the traffic. In 2005, Moore and others employed port identification to improve the traffic classification. They experimentally achieved an accuracy of the protocol classification slightly less than 70% [1], [2]. Because it is difficult to parse the data packet load from the encrypted traffic and most of the encrypted protocols are using port hopping technology, one may not be able to identify the encrypted traffic with a considerable accuracy, based on the port and data packet load method. Cao [3] and others summed up that the current classification methods of encrypted traffic are mainly on P2P and SSL encryption traffic. However, the present encryption traffic classification algorithm is mainly based on the network behavior and machine learning. Many studies show that using the host network behavior to classify the encrypted traffic is not stable since machine learning algorithms are developing very quickly. Thus, researchers started to utilize different algorithms to train off-line network flows, extract network flow characteristics and build the corresponding classification model, and finally adopted the trained model to classify the traffic online. Many studies show that in this way, one can achieve higher classification accuracy. Li [4] and others showed that the main challenge of flow classification is the online classification performance and feature extraction. The training set of machine learning methods relies on many computational flow characteristics, and highly depends on the data sets. Because the calculation of the characteristics from the network flow needs the computational flow characteristics of each packet, the surge in traffic will lead to a decline in the performance of classification and the classification accuracy would be unstable.

In order to ease the above-mentioned problems, we presented an encrypted traffic classification method based on the network flow of sliding time window of first few packets. This method only considers first few packets of the encrypted traffic flows, and considers time as a sliding window to quantify the continuity of the encrypted traffic. We also experimentally show that our method can not only yield stable classification results but also with a higher accuracy for encrypted traffic.

II. RELATED WORK

Encrypted traffic classification methods can be categorized into two: one is based on host behavior characteristics; the other is based on machine learning. The following of this section will discuss those two methods.

A. Host Behavior Characteristics Classification

Host behavior classification is based on a host transport layer behavioral characteristics. In 2005, Karagiannis [5]

978-1-5386-2030-4/17/$31.00 ©2017 IEEE

proposed a blind classification (BLINC), which is based on the behavior characteristics of the interactive communication process of host applications. In BLINC, the port number and packet load are not considered, and the results show that the classification model can reach 95% accuracy. However, it is difficult for the blind classification to meet the requirements of real-time classification and more difficult to implement with. In 2011, Iliofotou [6] puts forward a method based on behavioral diagrams to classify traffic. Classification accuracy has been improved by a certain degree compared with a BLINC protocol. But the computational cost of this method is high, and it could not meet the real-time classification, either. In 2011, Zhang [7] and others proposed a method based on user behaviors to classify network traffic online, including browser access, online chat and other activities of uploading and downloading. Relevant experiments showed that the classification accuracy rate following this method can reach 90%. In 2013, Xiong [8] presented a method based on associated host behavior to classify Thunder encrypted traffic and the experimental results showed that most of the classification accuracy is more than 95%, but this method is mainly associated with the Thunder traffic. It would be better to associate Thunder, DNS servers and other resources.

Without depending on the port number and the packet load, host interactive behavior classification could meet the performance issues for the backbone traffic classification. However, the host in the presence of dynamic interaction, such as dynamic routing communication process, may leave the classification method unstable.

B. Machine Learning Classification Method

Machine learning classification for encrypted traffic is becoming a hot-spot. Either ordinary or encrypted network traffic flow between communicating hosts contains certain statistical characteristics. Thus, machine learning methods can extract network traffic statistical features and then do feature selection and dimension reduction. Based on 249 flow characteristics [9] raised by Moore, researchers could select certain subsets of features and utilize appropriate classification algorithms to establish a classification model train to effectively classify the on-line encrypted traffic. However, machine learning depends on data sets. Moreover, some of the flow characteristics are required to calculate the flow's each packet so that performance of application identification has declined. The commonly used machine learning approaches for traffic classification are naive Bayes, K neighbors, C4.5 decision tree and support vector machine (SVM) classification. In 2009, Alshammari [10] chose 22 network flow characteristics and used C4.5 classification method to classify a variety of protocols. The results showed that classification accuracy of SSH encrypted traffic could reach an average of 97%. However, since the entire flow statistical features are considered, the performance is low. In 2011, Gu [11] used four classification methods to handle Skype encrypted traffic. Experiments showed that the effect of C4.5 decision tree classification is better. The average accuracy is higher than 93%. In 2012, Liu [12] proposed a semi-supervised classification method based network integration features to classify encrypted traffic HTTPS, SKYPE and Thunder traffic. The results showed that the average accuracy rate is less than 94%. In 2014 Korczynski [13] and others used Markov chain fingerprinting method to classify SSL/TLS encrypted traffic, the highest classification accuracy rate can reach 98%. But the classification method is unstable and some applications classification accuracy is only 59.7%.

In summary, these two categories of classification have certain limitations. The current studies are mainly focused on the characterization of each network flow and using machine learning methods to classify encrypted traffic and the leading accuracy may be unstable and may not meet the real-time requirements. In order to tackle the above-mentioned problems, for encrypted traffic flow, we utilize sliding time window of flow first N packets to classify encrypted traffic in our work. The results show that our approach may not only be stable and explanatory, but also has higher classification accuracy for encrypted traffic.

III. WFNP Traffic Classification Algorithm

This section describes the principles of encrypted traffic flow statistics feature selection process and WFNP classification algorithm.

A. Encrypted Traffic Flow Features Selection

In the past research on network flow feature selection, Moore [3] extracted more than 200 basic characteristics of the flow in 2005. In 2006, Nigel [14] integrated two feature selection algorithms, feature selection based on correlation (CFS) and feature selection algorithm based on consistency (CON). Finally, he chose 8 flow basic characteristics. In 2009, Li [15] adopted fast filtering algorithm (FCBF) to feature selection. Considering the stability of the space and time, they chose 12 flow statistic characteristics. In this paper, we combine the information gain ratio [16], the Correlation--based Feature Selection (CFS) and Consistency--based Feature Selection (CON) to select features. In machine learning, the information gain can be used to define a preferred sequence of attributes to investigate to most rapidly narrow down the state of X [17]. Let T denote a set of training examples, each of the form $(x, y) = (x_1, x_2, x_3, \ldots, x_k, y)$. Where x_k denotes the k-th attribute in X and y is the corresponding class label. The information gain ratio for an attribute x is defined by Quinlan in terms of entropy as follows:

$$GR(x)=\frac{Gain(x)}{SpInfo_x(D)} \quad (1)$$

where GR(x) means the information gain ratio. Gain(x) is information gain, which represents the difference between the original information requirement and the new requirement. $SpInfo_x(D)$ is split information, which means that the potential information generated by splitting the training set D based on the particular attribute x [16]. The split information is with respect to classification that is acquired based on the partitioning by the particular attribute x. So back to the (1) equation, the attribute that is with maximum value of gain ratio is chose as the splitting

attribute, which means that the corresponding attribute have a higher weight in the classification. Particularly, if the value of split information is 0, the gain ratio becomes unstable. To avoid this problem, the training set selected must be large enough.

Based on the characteristics of each grade, we finally choose 11 statistic features. Our selected features combine two parts: one is the traditional high grade features and another aspect is the features we proposed that relevant to SSH and P2P protocols. According to different evaluation criteria for each feature selection algorithms, different algorithms select different flow properties.

TABLE I. FLOW STATISTICS FEATURE

statistics feature	description
protocol	Transport layer protocol
first_bbytes	The first packet length backward
win_bbytes	Initialization window backward
win_fbytes	Initialization window forward
mean_fpktl	Mean packet length forward
sec_fbytes	The second packet length forward
min_bpktl	The minimum packet length backward
max_bpktl	The maximum packet length backward
duration	Traffic flow duration
fpsh_cnt	Psh packet number forward
third_bbytes	The third packet length backward

To balance the weight of the three feature selection algorithms and avoid skewness, we employ different coefficients as the specific weight to different algorithms. Each value $A(i)$ in a set may be associated with a weight $P(i)$. The weights reflect the significance, importance, or occurrence frequency attached to their respective values. It is defined as weight arithmetic mean or the weighted average. The statistical characteristics of the score are calculated for each flow by the Eq. (2).

$$A_{Score}(i) = P_{Gain} \times A_{Gain}(i) + P_{cfs} \times A_{cfs}(i) + P_{con} \times A_{con}(i) \quad (2)$$

where, i represents the i-th attribute. AGain(i), Acfs(i) and $A_{con}(i)$ denote the i-th selected feature attribute scores in three algorithms. For each algorithm, Eq. (3) describes weight relationship.

$$P_{Gain} + P_{cfs} + P_{con} = 1 \quad (3)$$

where PGain sets to 0.4, and the other two weights Pcfs and Pcon have taken 0.3. From previous studies [11], we can find that C4.5 decision tree classification method is relatively stable. And C4.5 classification select features through information gain algorithm. So we design more weight to information gain (PGain as 0.4). Based on the above feature selection method, flow statistics features have been selected and shown in Table I.

Firstly, based on Moore's basic characteristics of flow, we remove some redundancy features. There are 249 features proposed by Moore [9]. The amount of the subset of the features is redundancy. For instance, the feature mean_fpktl and min_fpktl, max_fpktl have a high correlation, so we will remove the redundancy features. Afterward, we add some features related to encrypted traffic flow. In the encrypted traffic flow, the first three packets carry important interactive information and we find that the feature first_bbytes and third_bbytes have a high weight. Finally, we get the 11 features shown in Table I. Protocol comprises of TCP and UDP, which means that different types of transport layer protocol may belong to different classes. As the aforementioned explanation, the three features comprising of first_bbytes, sec_fbytes and third_bbytes contain high classification weight in encrypted traffic. The features related to packet length include min_bpktl, mean_fpktl and max_bpktl. It is consistent with our manual experiment. The length of min_bpktl and max_bpktl in the encrypted traffic is greater than the length of normal traffic. However, the length of mean_fpktl in encrypted traffic is less than the length of normal traffic.

B. WFNP Classification Algorithm

Most of the research on traffic classification consider encrypted traffic across a complete flow. However, our work is mainly dependent on the statistical characteristics of each flow on the front N data packets. So, to a large degree it could reduce the computational cost. Moreover, this algorithm employs sliding window mechanism to quantify continuities of the encrypted traffic, combined with C4.5 decision tree classification method to classify encrypted traffic [11]-[18]. We consider C4.5 as our basic classifier because the following reasons: 1) C4.5 classifier is highly interpretive and comprehensive. 2) C4.5 algorithm can achieve a relatively high and stable accuracy in the traffic classification. 3) C4.5 classifier ease to apply into practice and employ to real-time classification.

The WFNP online classified method consists of two phases. The first stage is to extract the front N packets of each flow to calculate the statistical characteristics. In the second step, WFNP utilizes the C4.5 decision tree classification method to classify online traffic and store classification results such as traffic flow IP, port and 30 bytes load every two minutes. Most of the TCP connection time is 120 seconds so we set 2 minutes as a time window. In the third step, calculate information entropy of each protocol flow in each time window as well as compute the mean information entropy for the same kind of protocol flow log, and set a threshold for current window based on historical window of entropy. Then remove the wrong protocol flow flag in phase one. Finally, WFNP will associate multi-flow by IP and port to identify the unlabeled protocol flow. WFNP algorithm's detailed process is in the following:

1) Capture online traffic and extract first N packets of each flow to calculate the statistical characteristics;

2) Set 2 minutes as a time window. Employ C4.5 decision tree to classify each flow and log the classification flow protocol flag. Store the first 30 bytes from traffic load and the pair of IP-port in each time window;

3) Calculate the entropy of per flow according to the 30 load bytes and the mean entropy if the flags of classification flow are same. To distinguish between encrypted traffic and unencrypted traffic flow we can calculate load information entropy. The following Eq. (4) shows how to calculate the entropy of each flow:

$$I(F) = -\sum_{i=1}^{m} p_i \log_2(p_i) \tag{4}$$

where F denotes a flow, P_i is the probability of i-th byte number divided by the total number of bytes, m represents a total of load bytes, i represents the i-th byte.

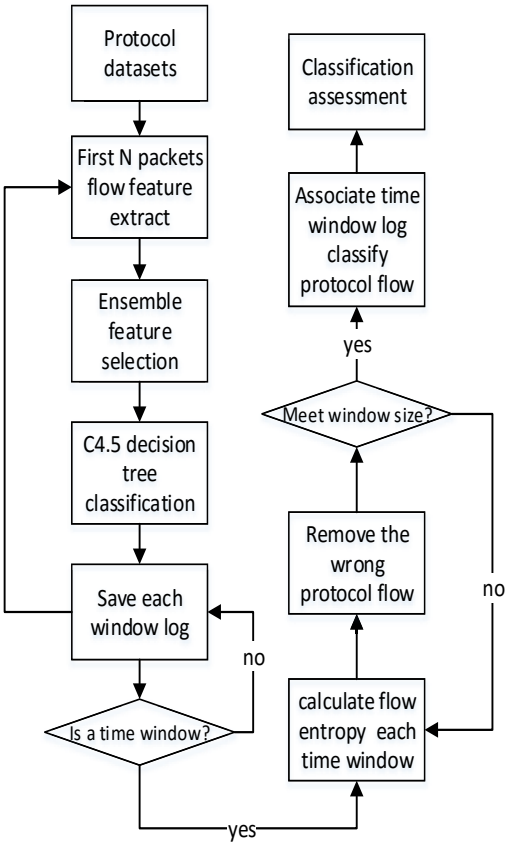

Figure 1. WFNP classification algorithm process.

4) Set up the value of information mean entropy of the same protocol flow as the entropy of each protocol in the current window. Compare each flow information entropy to the former mean entropy. If they are different and the difference is greater than the threshold, the current flow flag would be removed;

5) Slide time window. When the size of the window is equal to 5, WNFP will associate multi-flow by the IP and port of each flow. Classify the unclassified protocol flow based on the classified protocol flow.

The time window size in the paper is set to 5 and each time window is set to 2 minutes. We have already explained why these values were chosen above. The classification algorithm process is shown in Fig. 1.

The classification algorithm selects flow statistics feature, then use C4.5 classification method to classify the encrypted traffic directly and record online log automatically. WFNP also calculates the entropy of each flow of information, and then apply a sliding window mechanism to associate the flows by time and a pair of IP-port. Finally, it integrates two processes to classify encrypted traffic. Contrasted with existing classification techniques on the network traffic, through a combination of C4.5 decision tree classification and multiple flows association mining, WNFP not only considers the small size of traffic packets, but also remains a high a stable classification accuracy.

IV. EXPERIMENTAL METHOD

A. Dataset Description

The dataset adopted in this article was collected from the entrances of a LAN. We collected two data sets of every protocol with the tool of GT [19] in the different time in Jan, 2016. The full information load of the data packet had been all collected every time. The experiment condition was a network connector with 1000Mbps bandwidth. More detailed description of datasets is shown in detail in Table II.

TABLE II. PROTOCOL DATASET DESCRIPTION

dataset / protocol	data1		data2	
	size(M)	flow number	size(M)	flow number
SSH	103.6	1743	202.9	5140
FTP	30.7	327	80.94	577
P2P	116.36	3979	427.46	5760
SMTP	8.0	1128	44.4	2333
DNS	1.3	1855	4.0M	6075
HTTP	330.6	395	466.3	1385

It consists of two datasets. The size scale of Data1 is smaller than the Data2. The two datasets contain the encrypted traffic and non-encrypted traffic. Encrypted traffic mainly contains two types: One is SSH protocol traffic and the other one is P2P traffic, which is the hotspot in current research. The others are application level traffic about some well-known protocol. Generally, our data sets can represent common network data sets.

B. Experimental Procedure

The purpose of the experiment is to compare different protocols for each flow with the different N and to find the best N packets per flow classification. We use the NetMate tools to extract protocol network traffic statistics feature and

the mining tool named Weka-3.7.13 [20] to complete the flow feature selection and build the off-line classification model for encrypted traffic.

We adopt single factor packets to check four different classification algorithms by changing the number of data packets for each flow, so as to count the average accuracy of feature classification. Single factor packets methodology is that only one single variable changes and simultaneously other variables remain the same. The single variable here means that the number of first packets in each traffic flow. So we will change the number of first packets and control other variables in the same state and obverse the classification accuracy. Then we validate the results and choose the appropriate number of the data packets which could yield best statistical accuracy.

Based on the above datasets and the statistical features presented in section 2.1, we utilize the NetMate to calculate the statistical characteristics of each flow with different numbers of data packets. At the same time, we write a simple script to add protocol categories label to corresponding flow. In the experiments, we apply three popular classification algorithms and WFNP classification methods to classify different protocols, including K-nearest neighbor classification algorithm (KNN), Naive Bayesian classification algorithm (NB) and support vector machine (SVM) classification algorithm. Single factor mean analysis method verifies the number of packets to improve classification accuracy, where the single factor is the number of packets of flow. It ranges from 5 to 10 packets and compares with the entire flow packets. During the experiment, we adopt a ten-fold cross-validation method to verify the correctness of our approach. Each sample is randomly divided into ten parts, which in turn will be added to the test sample. The remaining nine will be added to the training sample. The final calculation of ten classification results precision and recall average rates are as a classification algorithm precision and recall rates.

C. Experimental Evaluation

Past research are measured with two common measure metrics, including the precision and recall, to validate the approach they proposed. As in this paper we also select the two as the metrics of the experiments and are calculated with the following variables:

(1) The true positive rate (True Positive, TP): the network data flow, which is identified by the algorithm to a protocol and it indeed belongs to the protocol.

(2) The false positive rate (False Positive, FP): the network data flow, which is identified by the algorithm to a protocol but it, does not belong to the protocol.

TABLE III. TOP N PACKET CLASSIFICATION ACCURACY

data set	packet / algorithm	5	6	7	8	9	10	*
1	KNN	93.2	94.5	94.8	96.0	95.9	95.8	96.1
	NB	93.3	93.8	94.2	95.2	95.3	95.1	95.1
	SVM	92.2	93.3	93.8	94.5	94.2	94.5	94.6
	WFNP	95.3	96.2	96.4	97.9	97.8	97.9	98.1
2	KNN	91.4	91.8	91.6	93.1	92.8	93.2	93.5
	NB	90.9	90.5	91.8	91.9	92.0	92.1	92.2
	SVM	91.2	92.1	91.8	93.0	92.9	92.8	93.0
	WFNP	94.8	95.5	95.3	96.9	96.1	96.3	96.5

(3) False negative rate (False Negative, FN): the network data flow which is identified by the algorithm to a non-protocol but it belongs to the protocol.

The measure of the precision and recall is widely used in the classification. The precision can be seen as a measure of the accuracy, and the recall rate is a measure of the completeness. Their computation is shown as Eq. (5) and (6).

$$R_{precision} = \frac{TP}{TP \quad FP} \tag{5}$$

$$R_{recall} = \frac{TP}{TP + FN} \tag{6}$$

D. Experimental Results Analysis

The first experiment shows the number of packets (5-10) in the front of each flow and the entire data packet flow is considered. It compares our WFNP algorithm with three other common classification algorithms. And the following Table III shows the result of classification accuracy.

In Table III, column * represents the number of packets, based on the entire data packet flow to calculate statistical characteristics. It can be concluded that in the classification of different protocol traffic there is basically no difference between the flow front N packets and the entire flow packets when we classify the encrypted traffic by statistical flow characteristics. From this result, we only need to calculate the statistical characteristics by each flow N front packets when we classify encrypted traffic because it can meet higher classification accuracy. Compared with other research, Su [21] presented the method that KNN-based classifier can achieve 92.21% precision. For NB classifiers, Moore [22] employed Bayesian analysis to popular protocol classification and achieved average percentage of 65.26%. By combining FCFB method, it can get higher precision but it is still around 84.06%. SVM classifier can do better because of its stability. Most of the researches on traffic classification by SVM can be recognized with precision above 90% [23], [24]. Support vector machines for TCP traffic classification got a high accuracy rate at 94.9% [23]. Looking deep into the result of Table III, we can conclude that our approach also can get a high accuracy and it is more interpretive and easier to implement. As follows, we will analyze the experimental results in more detail.

Fig. 2 shows the results of average classification precision in the front N packets in the two different data sets. This figure also presents the average classification precision of four classification accuracy.

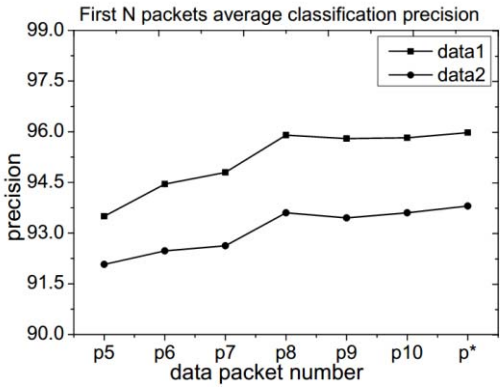

Figure 2. First N packets average classification precision.

Classification mean accuracy gradually increases according to the single factor number of packets range from 5 to 8. In data1, protocol traffic average accuracy is 93.5% when we only take the first 5 packets of each flow in feature extraction for four classification algorithms. When we take six, seven and eight packets, the average traffic classification accuracy are 94.45%, 94.8% and 95.9%, respectively. However, when setting 9 and 10 packets. Traffic classification average accuracies are 95.8% and 95.825%, respectively. Another polyline is data2. It is clear that in the case of larger data sets classification accuracy has declined.

But it also has the same trend from which we can see if we take front packets as 8 we will get more stable classification accuracy.

The above analysis shows that when we classify the online encrypted traffic we can only compute statistical features from front 8 packets of each flow. It will certainly and absolutely improve the classification performance. In this paper we consider only the first eight packets of each flow to calculate flow statistics feature. Then the next experiment applies four classification algorithms KNN, NB, SVM and WFNP to classify different protocol data sets data1 and data2. The protocols classification accuracy of each data set data1 and data2 are shown in Fig. 3 and Fig. 4.

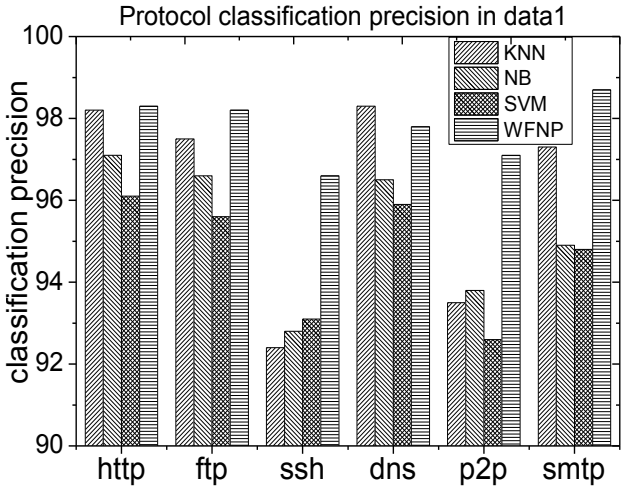

Figure 3. Fig. 3 Data1 protocol classification precision.

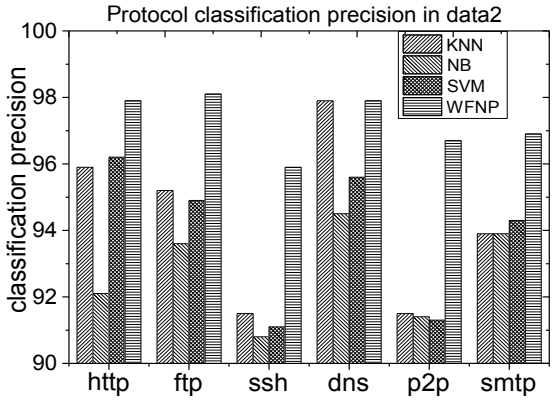

Figure 4. Data2 protocol classification precision.

Fig. 3 shows that there is no difference between the classification of non-encrypted traffic including http, ftp, dns and smtp traffic when we use the four-classification method. Firstly, we can find that KNN classifies http and dns protocol. Classification accuracy is up to 98.2% and 98.3% respectively. The accuracy of SVM also reached 97.1% and 96.1%. When we check into the encrypted traffic classification we can infer that WFNP classification accuracy was significantly higher than the other three classifications. KNN classifies SSH and P2P with only 92.4% and 93.5% respectively. SVM also has similar results, where the SSH and P2P are 93.1% and 92.6% respectively. While the use of the proposed WFNP classification, encrypted traffic classification precision are 96.6% and 97.1% accordingly.

Fig. 4 indicates the classification result of the data2 dataset. Compared with data1, the classification accuracy of data2 declines in a certain degree. Especially KNN and NB methods, average classification accuracy of various protocols falls by 3% in traffic classification. KNN and NB classification accuracies are 91.5% and 91.4% respectively. For WFNP classification, the average classification accuracy only fell less than 1%. And the classification accuracy of encrypted traffic, including SSH and P2P, are still 95.9% and 96.7% respectively. One can safely conclude that WFNP method we proposed not only could lead to a higher classification accuracy and classification stability, but also is irrelevant with the size of the dataset.

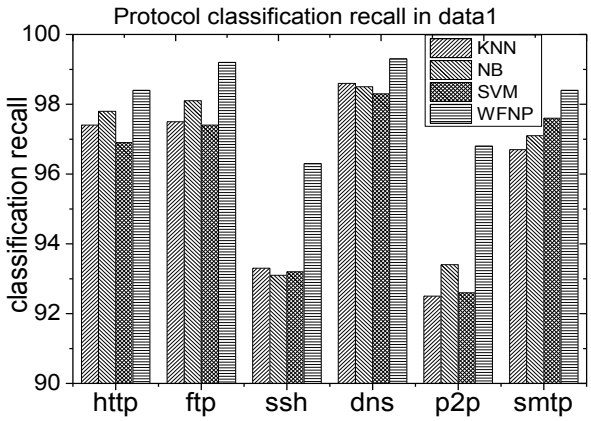

Figure 5. Data1 protocol classification recall.

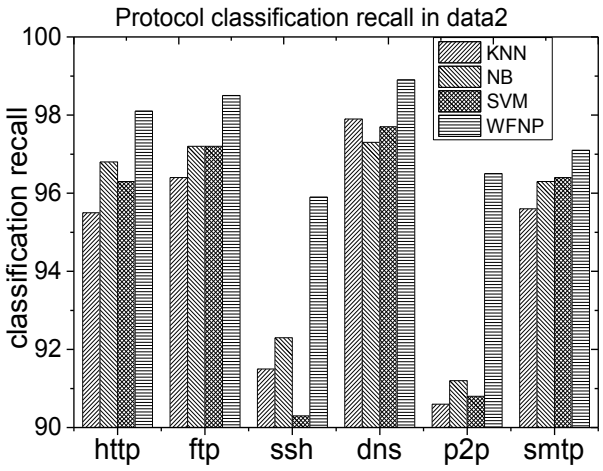

Figure 6. Data2 protocol classification recall.

Base on the above analysis we can draw the conclusion. For the classification of encrypted traffic, WFNP classification can achieve higher classification accuracy than the other three classification methods with stable outputs. In this paper, we measure both accuracy and recall rate as classification evaluation criteria. Fig. 5 and 6 show the classification recall in two data sets for the four classification methods.

The experimental results in Fig. 5 and Fig. 6 show that WFNP gets a higher recall for either encrypted or non-encrypted traffic, and the flow average recall is more than 96%, especially for the P2P traffic case, it can reach 97% on the test set of data1. Remarkably, the recall curve is still relatively stable. The recall of KNN and NB in encrypted traffic classification ranges from 90.6% to 93.4%.

V. Threat to Validity

A. External Validity

In our case study, we only focus on two datasets from the entrances of a LAN. Both of the datasets are mainly captured in six protocols. Some of the results might not be generalizable to other encrypted protocols. However, our approach can be applicable to more encrypted protocols only if we retrain the dataset and adjust the selected flow features. Future studies may consider more encrypted datasets from more encrypted protocols.

The non-linearity of data size possibly influences the accuracy and stability of classification. To reduce this threat, future study can combine other non-machine learning algorithms, such as regular expression to handle part of the relatively simple data.

B. Internal Validity

C4.5 classification can give us the intuitive split information and interpret the prediction easily. However, most of the time models based on other deep learning algorithms such as neural network can achieve a higher precision. In our future work, we will employ deep learning models to classify traffic.

Simply using precision and recall as validation metrics may be misleading when evaluating the binary class. In our dataset, test case without regression accounts for high proportion so it makes class imbalance. The precision and recall will be not stable. Further, we will utilize receiver operator characteristic (ROC) as a validation to eliminate this thread.

C. Construct Validity

Our study is based on the ability to accurately monitor the traffic of our subject dataset. This is based on the assumption that the traffic monitoring library, i.e. GT can success- fully and accurately capturing different types of protocol. This tool monitoring library is widely used in performance engineering research [17]. To further validate our results, other traffic monitoring tools can be used.

VI. Conclusions

With the increasing complex and diverse encrypted traffic, efficient stable classification for encrypted network traffic is essential to effectively manage network traffic. For encrypted traffic, mainly SSH and P2P traffic, this paper proposes an approach based on sliding windows and only considers first N packets of each flow classification algorithm. It effectively improves the precision and recall in encrypted traffic classification. The experimental result shows that the average accuracy of the proposed WFNP algorithm could reach 96% for classifying encrypted traffic including SSH and P2P traffic. Compared with other classification methods, the average accuracy could be improved by 3% using our algorithm. Moreover, our method can evidently improve the online classification performance, which is achieved by calculating the first certain number of packets of each flow but not counting the entire flow characteristics. Finally, future studies may consider more encrypted traffic data in even different environments. Future work can leverage and combine more in-depth network behavior mining to improve the classification accuracy.

Acknowledgement

This work was supported by the National Key Research and Development Program of China. (No.2016YFB0801600)

References

[1] Moore A, Papagiannaki K. Toward the accurate identification of network applications [C] Proceedings of Passive and Active Measurement Workshop (PAM2005). Boston(USA), 2005.

[2] DEWS C, WICHMANN A, FELDMANN A. An analysis of Internet chat system s[C] / /Proc of IMC. 03. New York: ACM Press, 2003: 51-64

[3] Cao Z, Xiong G, Zhao Y, et al. A Survey on Encrypted Traffic Classification[M]//Applications and Techniques in Information Security. Springer Berlin Heidelberg, 2014: 73-81.

[4] Li B, Springer J, Bebis G, et al. A survey of network flow applications[J]. Journal of Network and Computer Applications, 2013, 36(2): 567-581.

[5] Moore A, Zuev D, Crogan M. Discriminators for use in flow-based classification[M]. Queen Mary and Westfield College, Department of Computer Science, 2005.

[6] Karagiannis T, Papagiannaki K, Faloutsos M. BLINC: multilevel traffic classification in the dark[C]//ACM SIGCOMM Computer Communication Review. ACM, 2005, 35(4): 229-240.

[7] Iliofotou M, Kim H, Faloutsos M, et al. Graption: A graph-based P2P traffic classification framework for the internet backbone[J]. Computer Networks, 2011, 55(8): 1909-1920.

[8] Zhang F, He W, Liu X, et al. Inferring users' online activities through traffic analysis[C]//Proceedings of the fourth ACM conference on Wireless network security. ACM, 2011: 59-70.

[9] Xiong G, Huang W, Zhao Y, et al. Real-time detection of encrypted thunder traffic based on trustworthy behavior association[M]//Trustworthy Computing and Services. Springer Berlin Heidelberg, 2013: 132-139.

[10] Alshammari R, Zincir-Heywood A N. A Preliminary Performance Comparison of Two Feature Sets for Encrypted Traffic Classification[J]. Advances in Soft Computing, 2009, 53:203-210.

[11] Gu C, Zhang S, Sun Y. Realtime encrypted traffic identification using machine learning[J]. Journal of Software, 2011, 6(6): 1009-1016.

[12] Liu H, Wang Z, Wang Y. Semi-supervised encrypted traffic classification using composite features set[J]. Journal of Networks, 2012, 7(8): 1195-1200.

[13] Korczynski M, Duda A. Markov chain fingerprinting to classify encrypted traffic[C]//INFOCOM, 2014 Proceedings IEEE. IEEE, 2014: 781-789.

[14] Williams N, Zander S, Armitage G. Grenville Armitrage A Preliminary Performance Comparison of Five Machine Learning Algorithms for Practical IP Traffic Flow Classification[J]. Sigcomm Computer Communication Review, 2006, 30(5):5-16.

[15] Li W, Canini M, Moore A W, et al. Efficient application identification and the temporal and spatial stability of classification schema[J]. Computer Networks, 2009, 53(6): 790-809

[16] Michalski, Ryszard S., Jaime G. Carbonell, and Tom M. Mitchell, eds. Machine learning: An artificial intelligence approach. Springer Science & Business Media, 2013.

[17] Quinlan, J. Ross. "Improved use of continuous attributes in C4. 5." Journal of artificial intelligence research 4 (1996): 77-90.

[18] Shafiq, Muhammad, et al. "Network Traffic Classification techniques and comparative analysis using Machine Learning algorithms." *Computer and Communications (ICCC), 2016 2nd IEEE* International *Conference on*. IEEE, 2016.

[19] Gringoli F, Salgarelli L, Dusi M, et al. Gt: picking up the truth from the ground for internet traffic[J]. ACM SIGCOMM Computer Communication Review, 2009, 39(5): 12-18.

[20] Hall M, Frank E, Holmes G, et al. The WEKA data mining software: an update[J]. ACM SIGKDD explorations newsletter, 2009, 11(1): 10-18

[21] CHEN Li-nan, LIU Yang, et al. Efficent-cutting packet classification algorithm based on the statistical decision tree. Journal on Communications, 2014, 1000-436X(2014)Z1-0058-07.

[22] Su, Ming-Yang. "Using clustering to improve the KNN-based classifiers for online anomaly network traffic identification." Journal of Network and Computer Applications 34.2 (2011): 722-730.

[23] Moore, Andrew W., and Denis Zuev. "Internet traffic classification using bayesian analysis techniques." ACM SIGMETRICS Performance Evaluation Review. Vol. 33. No. 1. ACM, 2005.

[24] Este, Alice, Francesco Gringoli, and Luca Salgarelli. "Support vector machines for TCP traffic classification." Computer Networks 53.14 (2009): 2476-2490.