2017 IEEE 19th International Conference on High Performance Computing and Communications; IEEE 15th International Conference on Smart City; IEEE 3rd International Conference on Data Science and Systems

Robust Feature Selection for IM Applications at Early Stage Traffic Classification Using Machine Learning Algorithms

Muhammad Shafiq[1], Xiangzhan Yu[1] and Dawei Wang[2]
[1]School of Computer Science and Technology Harbin Institute of Technology
Harbin 15001 PR, China.
National Computer Network Emergency Response Technical Team Coordination Center of China
{muhammadshafiq, yuxiangzhan}@hit.edu.cn
stonetools@yeah.net

***Abstract*— Identification of network traffic accurately at its early stage is very important for network traffic management and application traffic classification. In recent years, this becomes very hot topic to identify traffic at its early stage. Unidirectional and bidirectional statistical features are effective features and widely used in Internet traffic classification. However, it is important to evaluate and select effective features for Instant Messaging (IM) application traffic classification at early stage. In this paper we are interested to find out robust and effective features at early stage. We firstly extract 22 statistical features of the first flow on two different network environment traffic datasets include on HIT and NIMS datasets. Then mutual information is conducted between the extract statistical features to select the effective features. Additionally to select robust features, we execute attribute selection cfsSubsetEval with Best search evaluator that select the robust and stable features from the result achieved by mutual information. And then, we execute 10 well-known machine learning classifiers. Our experimental results show that max_fpktl, std_bpktl, max_biat, mean_fpktl, mean_bpktl and min_biat feature are robust features at early stage traffic classification.**

***Keywords*— Robust; Effective Feature Selection; IM WeChat; Traffic Classification; Machine Learning.**

I. INTRODUCTION

In our previous works [1,2,3, 28,29], we classify Instant Messages (IM) applications includes with WeChat IM application using two data set at early stage to identify effective packet number for IM application traffic classification. In these published work our main contribution was to identify effective packets number at early stage for IM applications. However, effective features selection is also important for accurate Internet traffic classification. In 2005 A.W. Moor et al. [4] present a feature extraction method, which is mostly widely used for features extraction. They extract 248 statistical features, such as packet size, minimum, maximum and RTT features etc. Using these features ML classifiers can get very efficient performance results in Internet traffic classification. B. Qu et al [5], found in their study that it is possible to classify Internet traffic at early stage based on accuracy of early stage problem. Keeping in view the security policies and network traffic management, it is very significant to identify Internet traffic with early stage. Recently in 2012 in [6], Internet traffic classification at early stage become very hot topic in Internet traffic classification. Similarly in 2006 L. Bernaille et.al in [7] use the first few packets sizes of each TCP flow as a features and then apply K-means clustering technique, they got very high identification performance rates for 10 different types of application traffic. Using packet sizes of the first flow, this technique become very famous in Internet traffic classification. A. Este et al. in 2009 [8] proved using round trip time (RTT), packet size, packet direction and inter arrival time (IAT) of the early stage packets and found that early stage traffic carry enough information for traffic identification. In Zhang Hongli et at. [11] Proposed two different algorithms for features selection to select feature for imbalance traffic classification. Likewise, Lizhi Peng et al. in 2016 [12] evaluate the effectiveness of statistical features at early stage traffic classification.

However, it is very important to select robust features for IM applications at early stage particularly for WeChat application and no study has been carried out on robust features selection for IM application at early stage traffic classification. However this is the first study which is related to IM application traffic classification at early stage. For example, WeChat is an instant messaging and free of cost application developed by Tencent Holding in China. After launching WeChat application, it online clients users reached to 300 million in [9] and after that in November 2015 its active clients was about 650 million in all over the world. While, it's active users reached to 100 million from outside China in [10]. So it is also important to select robust features for IM applications include with WeChat application at early stage. Qun Huang et al. in [13] proposed a measurement ChatDissect tool for the measurement of IM WeChat application traffic. Using this they identify 150K users and 16GB traffic of WeChat application in network traces. Lim et al. in [14] utilized the packet size, connection level and statistical features for the classification Internet traffic. Similarly Der Puttern and Someren in [15] showed that features selection is very important for the optimization of performance. Church and Rodrigo de Oliveira in 2013 in [16] study the mobile messaging services sending performance with traditional messages. Similarly O'Hara et al. in [17] in 2014 study the mobile application WhatsApp in smart phone to study the user activity. For this technique they take some surveys and interviews. Fiadino et al. in [18] also study WhatsApp IM application, but their study was application flow stream. For this study they collect data in European Network which incudes on millions of audio and video flow data stream. Liu et al. [19] in 2014 study two application different IM application video messages services WeChat and WhatsApp. For their study they trace the traffic using mobile devices.

In the above mentioned study, it is clear that packet level features and derived packet level features are applied at early stage to classify Internet traffic. however, it is also important to identify network traffic at early stage using unidirectional and bidirectional statistical features and also essential to select robust features for IM applications at early traffic classification.

Contributions: In this paper, we set out to evaluate and select robust features for Instant Messaging (IM) applications at early stage. For the robust and effective features selection, we use mutual information analysis and attribution selection method with best search algorithm and experimental methods. Two different network environment traffic data sets and 10 well-known machine learning classifiers are conducted for our experimental work. We use unidirectional and bidirectional flow for this study and then we extract 22 statistical features at early stag flow traffic. Firstly the mutual information analyses are conducted between the statistical features to evaluate its effectiveness and also to select effective features. Then we use attribute selection method with best search method to select the robust features from the result achieved by mutual information method. After that 10 machine learning classifiers are applied using the robust features to validate the robustness of the selected features.

The rest of the paper is organized as follows: Section 2 demonstrate the characteristics of the selected data sets used in this study. Section 3 elaborates the features extraction in details while methodologies are discussed in Section 4. The details experimental results and analysis are given in Section 5. Finally, conclusion and future work draw in Section 6.

II. DATA SETS

In this study we select one set of open data set and a set of collected in our University laboratory for our study.

A. HIT Trace I Dataset

We trace HIT Trace I dataset at our University laboratory, we capture WeChat four functionality such text messages, pictures messages, audio call and video call TCP and UDP traffic unidirectional and bidirectional traffic. We also capture four other applications traffic such as P2P, IM, IMAP, and FTP application traffic. As we are interested to set out select robust feature for IM application traffic classification at early stage. For this purpose, firstly we capture IM WeChat application four functionality traffic (as we discussed in above line) using Wire Shark tool [20] duration of 1 hour at research lab School of Computer Science and Technology, Harbin Institute of Technology Harbin China in 27 December 2015 and 28 April 2016 respectively. But we select the traffic that none zero payload packets. In this process of capturing, we are interested to capture WeChat TCP, UDP traffics of text messages, pictures messages, audio call and video call. After that we also capture P2P, IM, IMAP and FTP traffics. After capturing the traffic, the trace file is saved as dot PCAP extinction. The characteristics of this datasets are given in Table 1. However, TCP mean WeChat TCP traffic and UDP also WeChat UDP traffic.

Table 1 Characteristics of HIT Trace 1 data set

Application	Duration Time	#Instances	Date
TCP	1 hour	20512	28 Apr 2016
UDP	1 hour	16400	28 Apr 2016
P2P	1 hour	1501	27 Dec 2015
IM	1 hour	7911	27 Dec 2016
IMAP	1 hour	15832	27 Dec 2015
FTP	1 hour	25251	27 Dec 2015

B. NIMS Dataset

NIMS data set includes on packets collected at the authors' research tested network. The data set consist on SSH servers outside connection and application behaviors traffic such as DNS, HTTP, SFTP and P2P traffic. However, we are also interesting in instant messaging application traffic classification, in this base we also added NIMS GTalk trace traffic, which includes on TCP Gtalk traffic and UDP Gtalk Traffic. Moreover, in NIMS data set we select only DNS, HTTP, SFTP, Gtalk TCP and Gtalk UPD traffic for our research work study. The detail characteristics of NIMS data are shown in Table 2.

Table 2 Characteristics of NIMS Data Set

Application	#Instances	Location	Date
GTalkTCP	482	Dalhousie University network	2010
GTalkUDP	9176	Dalhousie University network	2010
DNS	12734	Dalhousie University network	2010
FTP	1728	Dalhousie University network	2010

HTTP	3840	Dalhousie University network	2010
SFTP	2269	Dalhousie University network	2010

III. Flow Based Features

In this study we use the network traffic as a bidirectional flow connection between the two hosts where the two hosts have the same 5-tuple (source and destination IP address, source and destination port numbers and protocol). The generated flows represent client to server connection forward while server to client flow represent backward direction. We use NetMate tool [21] to generate the flow and statistical feature as we shown in Table 3 with details.

S.NO	Feature	Features Name
1	min_fiat	Minimum of forward inter-arrival time
2	mean_fiat	Mean of forward inter-arrival time
3	max_fiat	Maximum of forward inter-arrival time
4	std_fiat	Standard deviation of forward inter-arrival times
5	min_biat	Minimum of backward inter-arrival time
6	mean_biat	Mean backward inter-arrival time
7	max_biat	Maximum of backward inter-arrival time
8	std_biat	Standard deviation of backward inter-arrival times
9	min_fpkt	Minimum of forward packet length
10	mean_fpkt	Mean of forward packet length
11	max_fpkt	Maximum of forward packet length
12	std_fpkt	Standard deviation of forward packet length
13	min_bpkt	Minimum of backward packet length
14	mean_bpkt	Mean of backward packet length
15	max_bpkt	Maximum of backward packet length
16	std_bpkt	Standard deviation of backward packet length
17	proto	Protocol
18	Duration	Total duration
19	f_packets	Number of Packets in forward direction
20	f_bytes	Number of Bytes in forward direction
21	b_packts	Number of Packets in backward direction
22	b_bytes	Number of Bytes in backward direction

IV. Methodology

A. *Mutual Information Based Metric*

Mutual information analysis is extensively used in Information theory for features effective features selection [22], image processing [23] and speech recognition [24] etc. it is widely used for measuring the mutual dependencies between two random variables X and Y, and define the quantity of information held by random variable. The mutual information between two random variables is described as.

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &= H(Y) - H(Y|X) \\ &= H(X) + H(Y) - H(X,Y) \\ &= H(X,Y) - H(X|Y) - H(Y|X) \end{aligned} \quad (1)$$

In equation (1), the marginal entropies of *X* and *Y* are *H(X)* and *H(Y),* while conditional entropies are *H(H-Y)* and *H(Y-X)* and joint entropies of *X* and *Y* are *H(X,Y).* Moreover, the relationship between *H(X), H(Y), H(X-Y), H(Y-X), H(X,Y)* and *I(X;Y)* are shown in Figure 1. According to Shannon definition of entropy theory, we have

$$H(X) = -\sum_{x\epsilon X} p(x) log(p(x)) \quad (2)$$

$$H(Y) = -\sum_{x\epsilon Y} p(y) log(p(y)) \quad (3)$$

$$H(X,Y) = -\sum_{x\epsilon Y}\sum_{y\epsilon Y} p(x,y) log(p(x,y)) \quad (4)$$

Where *P(.)* specify the probability distribution of a random variable. As in [25] use the three equations in equation (1) to achieve the computational formula for mutual information. We use the same method that they have used in [25] for mutual information.

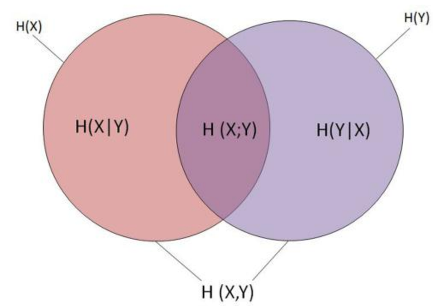

Fig.1 The relationship between mutual information and entropies

$$I(X,Y) = -\sum_{x\epsilon Y}\sum_{y\epsilon Y} p(x,y) log(\frac{p(x,y)}{p(x)p(y)}) \quad (5)$$

For mutational information analysis there is a lot of open sources application on Internet. But we use H. Peng's Matlab toolbox [26] for our mutual information analysis.

B. *Attribute Selection*

In this paper we used attribute selection method named cfsSubsetEval for robust feature selection from the results achieved by mutual information with best search method. The attribute cfsSubsetEval evaluators evaluate the worth of attributes predictive ability with redundancy and also evaluate the subset of feature which highly correlated to the class. There are many for attribute selection but we use Weka cfsSubsetEal attribute evaluator with best search method to select robust features set.

C. *Classifiers*

We execute for our identification 10 well-known machine learning classifiers. For this study work we use Weka data mining software [27] as our experiment tool. The introductory information of the utilized machine learning classifiers is given below.

– Bayes: Bayes machine learning classifiers are widely used in Internet traffic classification. It is actually based on Bayes Theorem. In this paper we used Bays Net and Naïve Bayes machine learning classifiers.
– Meta: Using Meta category of ML classifiers, we used Bagging and AdaBoost machine learning classifiers for our study. These ML classifiers first learned and then yield learning.
– Rule: In Rule category, we choose OneR and PART ML classifiers for the identification. Rule category first creates specific rules and then performs classification on testing data.
– Tree: This category are also called decision tree which are used in many research area in recent years. However, we choose only J48, Naïve Bayesian and Random Forest ML classifiers for our study.
– SMO: In this function category, we choose Support Vector Machine which is also called SMO in function category. However it is very useful for both classification and regression in network traffic classification.

D. *Performance Measures*

In this research work we consider the confusion matrix which is based on measuring a classification work. In confusion matrix rows are called instances of class and columns are called predicted class. Fig. 2 shows the binary classification confusion matrix used in this research work. We used the following measurement for our research study.

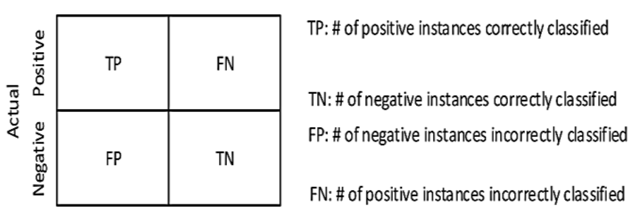

FIG. 2 CONFUSION MATRIX FOR CLASSIFICATION RESULTS EVALUATION

- Classification Accuracy: it is the number of exact correctly classified traffic flows divided by total classified flows. We use the following mathematical equation for our study.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

- Sensitivity: sensitivity and recall are the same metrics in Internet traffic classification technique. So Equation 4 can be used for sensitivity.

$$Sensitivity = \frac{TP}{TP + FN} \quad (7)$$

- Specificity: It can be defined as the ability of machine learning classifier to classify negative results. Equation 5 shows its formula and while mathematically it can be defined as TN divide by sum of FP and TN.

$$Specificity = \frac{TN}{FP + TN} \; 1 - FPR \quad (8)$$

- Area under curve: It is also called receiver operating characteristics (ROC) curve, which defines the performance of machine learning classifiers. It also shows the trade-off among FPR and TPR, while FPR is also known as Specificity and TPR is called Sensitivity. The AUC values can be computed by using confusion matrix values by TPR and FPR.

$$AUC = \frac{1 + TPR - FPR}{2} \quad (9)$$

Since $Specificity = 1 - FPR$ and $Sensitivity = TPR$

Replacing 1-FPR by Specificity and TPR by Sensitivity, we will get

$$AUC = \frac{Sensitivity + Specificity}{2} \quad (10)$$

Equation 10 shows that *AUC* is the average of Sensitivity and Specificity results performance.

V. EXPERIMENTAL RESULTS AND ANALYSIS

A. *Mutual Information Analysis between features*

In this study work, we use mutual information analysis between each features and the traffic type label for both NIMS and HIT datasets to select effective features for Instant messages traffic classification as shown in Fig. 3. In this figure all the statistical features show very effective mutual information values, but 13 features are very effective feature accord in mutual information analysis as shown in Table. 4. We perform mutual information experiment for both data sets. But the mutual information analysis of both different network environment datasets are little different as compare both datasets features. However, we select the effective feature according to the highest mutual information values for both HIT and NIMS datasets. In HIT data set min_fiat feature give the highest mutual information values as compare to other features and then second on std_biat and then std_bpktl and so on.

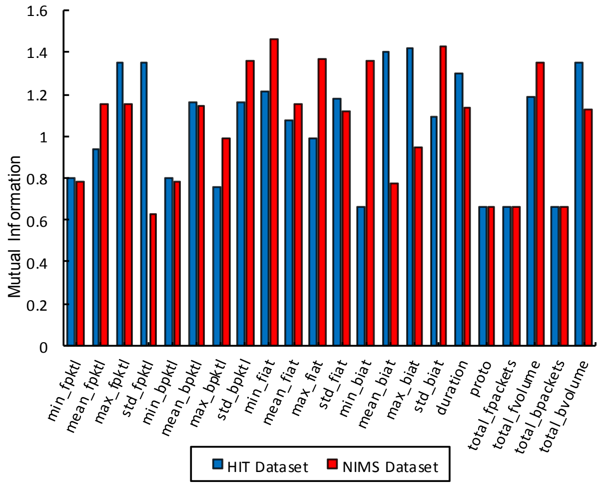

FIG.3 MUTUAL INFORMATION ANALYSIS BETWEEN FEATURES

While in HIT data set min_biat feature give highest mutual information results for IM Internet traffic classification, similarly std_bpktl on second and total_fvolume on third and son on. But again we select those features which produce highest mutual information values for both HIT and NIMS data set.

TABLE 4 SELECTED FEATURES WITH MUTUAL INFORMATION VALUES

HIT Dataset		NIMS Dataset	
Feature Name	MI Values	Feature Name	Values
max_biat	1.418295	min_fiat	1.464752
mean_biat	1.402424	std_biat	1.430143
max_fpktl	1.353784	max_fiat	1.368748
total_bvolume	1.353781	min_biat	1.361419
std_fpktl	1.353781	std_bpktl	1.357562
duration	1.303743	total_fvolume	1.348352
min_fiat	1.215905	mean_fiat	1.152738
total_fvolume	1.184157	mean_fpktl	1.151476
std_fiat	1.177029	max_fpktl	1.151476
std_bpktl	1.166105	mean_bpktl	1.145562
mean_bpktl	1.166105	duration	1.135510
std_biat	1.097662	total_bvolume	1.126514
mean_fiat	1.078614	std_fiat	1.118119

B. *Robust Feature Selection*

We select the robust feature from the selected features of mutual information analysis by using Weka attribute selection method with best search method. This method selects three features in HIT data set out of seventeen features while four features in NIMS dataset as shown in Table 5.

TABLE 5 ROBUST FEATURES OF BOTH NIMS AND HIT DATA SETS

HIT Dataset		NIMS Dataset	
Feature Name	Values	Feature Name	Values
max_biat	1.418295	min_biat	1.361419
max_fpktl	1.353784	mean_fpktl	1.151476
std_bpktl	1.166105	max_fpktl	1.151476
….	…..	mean_bpktl	1.145562

C. *Identification Results*

In this section, we carry out the classification experiment to validate the selected robust features of HIT and NIMS data sets. For the classification we performed our classification experiment by utilizing ten well-known machine learning classifiers with 10 folders cross validation method. However for the evaluation of ten machine learning classifiers we use select accuracy (ACC) and AUC metrics. Fig. 4 and 5 show the accuracy and AUC results for HIT data set respectively. The entire applied machine learning classifiers gives very effective accuracy results for HIT data set using selected robust features as shown in figure. However Navies Bayes and Hoeffding machine learning classifiers perform low as compare to other machine learning classifiers. Similarly AUC results of HIT data set is very effective but TCP, UDP and P2P traffic AUC results value are better than compare other traffic in HIT data set. Likewise all the applied classifiers AUC values are effective, but Bayes Net, Bagging, OneR, Part, RandomForest and C4.5 decision tree gives very effective AUC values performance as compare to other machine learning classifiers. Using HIT data set TCP, UDP, P2P, IM and FTP AUC results values.

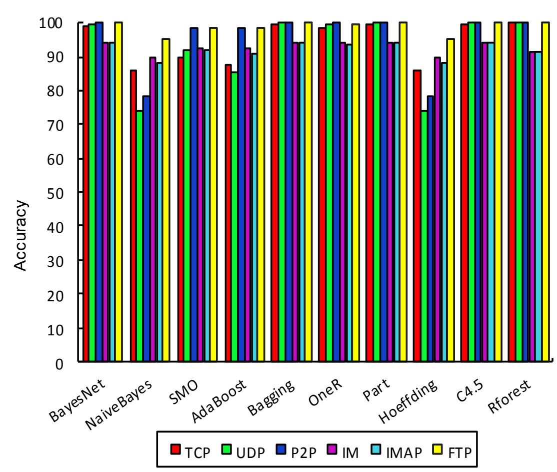

FIG.4 ACCURACY RESULT FOR HIT DATA SET

When observing the results of NIMS data set in Fig. 6 and 7. Although HIT data set gives accuracy (ACC) and AUC experimental results very efficient, but as compare to NIMS data set classification results NIMS data set performance results is very accurate. Using NIMS data set all the applied machine learning classifiers performance are very effective using the selected robust features, but Bayes Net, Bagging, Part, C4.5 and RandomForest machine learning classifiers get very accurate classification accuracy results as compare to other machine learning classifiers.

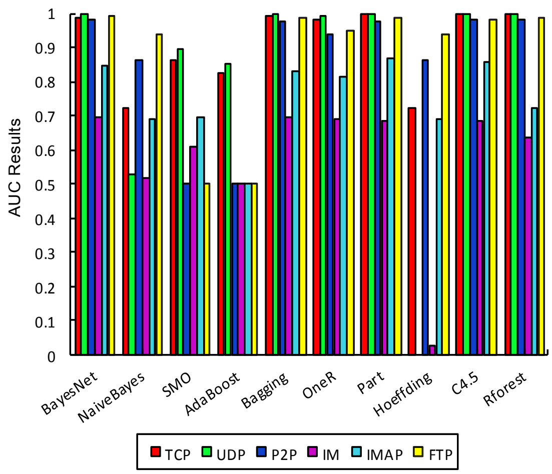

FIG.5 AUC RESULT FOR HIT DATA SET

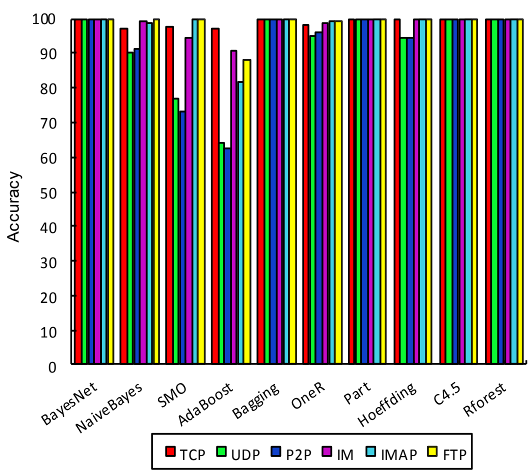

FIG.6 ACCURACY RESULT FOR NIMS DATA SET

But only Support Vector Machine (SVM) and AdaBoost ML classifiers performance are low. However, AUC result is different from accuracy results of NIMS data set, anyway all the ML classifiers AUC result for NIMS data set is very good but only SMO and Adaboost ML classifiers AUC result is low as compare to other machine learning classifiers. Nevertheless, Bays Net, Bagging, Part, C4.5 and RandomForest ML classifiers get very accurate AUC results for NIMS data set using robust feature.

VI. ANALYSIS AND DISCUSSION

Although the performance results of ten applied machine learning classifiers are different with respective to accuracy (ACC) and AUC using HIT and NIMS data sets. But some lesson can be learned.

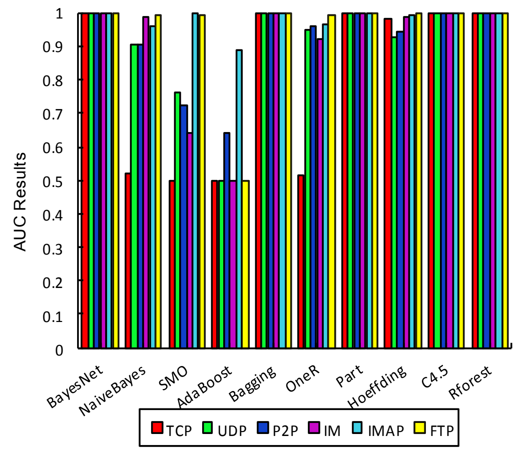

FIG.7 AUC RESULT FOR NIMS DATA SET

- In this study work, it is clear that our selected robust features set are very effective and gets very efficient performance result for both NIMS and HIT data sets. We validate the robust features with AUC and Accuracy metrics.
- From the experimental study using ten well-known ML classifiers, all the applied ML classifiers performance are very effective using robust features set. However, Bayes Net, C4.5 and RandomForest ML classifiers performance are very effective in overall experiments.
- In this research study, though classification performance are evaluated with accuracy and AUC metrics for the IM traffic classification, but some ML classifiers get high performance results and some get low performance results. It's due to imbalance traffic.
- From this experimental work, it is clear that it is possible to classify instant messaging (IM) with few statistical features set at early stage Internet traffic classification.

VII. CONCLUSION

This paper set out to select robust features for Instant Messages traffic classification at early stage. We use mutual information and Weka attribute selection method to select robust features for IM traffic classification at early stage. Using two dataset includes one opening data set and 10 well-known machine learning classifiers are applied. According to our experimental result, first we select 13 features from 22 statistical features of early stage flow using mutual information analysis. We select those features which have highest mutual information values. Then we Weka attribute selection with best search method to select the robust statistical feature from the feature achieved by mutual information. After that we execute 10 machine learning classifiers, we validate the classification results with accuracy and AUC metrics. At end we validate that our selected robust features max_biat, max_fpktl, std_bpktl, min_biat, mean_fpktl, max_fpktl and mean_bpktl are very effective feature for instant messages traffic classification at early stage. We infer that our selected robust feature are very effective, but there is still need to pick up more effective features for IM early stage traffic classification and this is our future work.

ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation of China under Grant No. 61571144 and 61771166, and National Key Research & Development Plan of China under Grant 2016QY05X1000.

REFERENCES

[1] Muhammad Shafiq and Xiangzhan Yu, “Effective Packet Number for 5G IM WeChat Application at Early Stage Traffic Classification,” Mobile Information Systems, vol. 2017, Article ID 3146868, 22 pages, 2017. doi:10.1155/2017/3146868

[2] Shafiq, Muhammad, Xiangzhan Yu, and Asif Ali Laghari. "WeChat Text Messages Service Flow Traffic Classification Using Machine Learning Technique." IT Convergence and Security (ICITCS), 2016 6th International Conference on. IEEE, 2016.

[3] Shafiq, Muhammad, et al. "WeChat Text and Picture Messages Service Flow Traffic Classification Using Machine Learning Technique." High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS), 2016 IEEE 18th International Conference on. IEEE, 2016.

[4] Moore, Andrew, Denis Zuev, and Michael Crogan. Discriminators for use in flow-based classification. Queen Mary and Westfield College, Department of Computer Science, 2005.

[5] Qu, Buyu, et al. "On accuracy of early traffic classification." Networking, Architecture and Storage (NAS), 2012 IEEE 7th International Conference on. IEEE, 2012.

[6] Dainotti, Alberto, Antonio Pescape, and Kimberly C. Claffy. "Issues and future directions in traffic classification." IEEE network 26.1 (2012).

[7] Bernaille, Laurent, et al. "Traffic classification on the fly." ACM SIGCOMM Computer Communication Review 36.2 (2006): 23-26.

[8] Este, Alice, Francesco Gringoli, and Luca Salgarelli. "On the stability of the information carried by traffic flow features at the packet level." ACM SIGCOMM Computer Communication Review 39.3 (2009): 13-18.

[9] More than 300 million users engage http://www.chinadaily.com.cn/cndy/2013-01/17/content_16128915.htm

[10] By the numbers: 50+ Amazing WeChat Statistics http://expandedramblings.com/index.php/wechat -statistics/

[11] Zhang, Hongli, et al. "Feature selection for optimizing traffic classification." Computer Communications 35.12 (2012): 1457-1471.

[12] Peng, Lizhi, et al. "Effectiveness of Statistical Features for Early Stage Internet Traffic Identification." International Journal of Parallel Programming44.1 (2016): 181-197.

[13] Huang, Qun, et al. "Fine-grained dissection of WeChat in cellular networks." 2015 IEEE 23rd International Symposium on Quality of Service (IWQoS). IEEE, 2015.

[14] Lim, Yeon-sup, et al. "Internet traffic classification demystified: on the sources of the discriminative power." Proceedings of the 6th International COnference. ACM, 2010.

[15] Van Der Putten, Peter, and Maarten Van Someren. "A bias-variance analysis of a real world learning problem: The CoIL challenge 2000." Machine Learning 57.1-2 (2004): 177-195.

[16] Church, Karen, and Rodrigo de Oliveira. "What's up with whatsapp?: comparing mobile instant messaging behaviors with traditional SMS." Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services. ACM, 2013.

[17] O'Hara, Kenton P., et al. "Everyday dwelling with WhatsApp." Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing. ACM, 2014.

[18] Fiadino, Pierdomenico, Mirko Schiavone, and Pedro Casas. "Vivisecting whatsapp through large-scale measurements in mobile networks." ACM SIGCOMM Computer Communication Review. Vol. 44. No. 4. ACM, 2014.

[19] Liu, Yao, and Lei Guo. "An empirical study of video messaging services on smartphones." Proceedings of Network and Operating System Support on Digital Audio and Video Workshop. ACM, 2014.

[20] Wireshark, http://www.wireshark.org/ (last Accessed Sept 2015)

[21] NetMate. http://www.ip-measurement.org/tools/netmate/.

[22] Peng, Hanchuan, Fuhui Long, and Chris Ding. "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy." IEEE Transactions on pattern analysis and machine intelligence 27.8 (2005): 1226-1238.

[23] Maes, Frederik, et al. "Multimodality image registration by maximization of mutual information." IEEE transactions on Medical Imaging 16.2 (1997): 187-198.

[24] Bahl, L. R., et al. "Maximum mutual information estimation of hidden Markov model parameters for speech recognition." proc. icassp. Vol. 86. 1986.

[25] Peng, Lizhi, Bo Yang, and Yuehui Chen. "Effective packet number for early stage internet traffic identification." Neurocomputing 156 (2015): 252-267.

[26] Peng, H.: Mutual infomation Matlab toolbox, http:// www.mathworks.com/ matlabcentral/ fileexchange/14888- mutual- information-computation.

[27] Weka 3: Data Mining Software in Java, http:// www.cs.waikato.ac.nz/ m l/ weka/

[28] Shafiq, Muhammad, et al. "Effective Feature Selection for 5G IM Applications Traffic Classification." Mobile Information Systems 2017 (2017).

[29] 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.