

## Research Article

# Effective Feature Selection for 5G IM Applications Traffic Classification

Muhammad Shafiq,<sup>1</sup> Xiangzhan Yu,<sup>1</sup> Asif Ali Laghari,<sup>1</sup> and Dawei Wang<sup>2</sup>

<sup>1</sup>School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China

<sup>2</sup>National Computer Network Emergency Response Technical Team, Coordination Center of China, Beijing, China

Correspondence should be addressed to Xiangzhan Yu; yuxiangzhan@hit.edu.cn

Received 11 January 2017; Accepted 4 April 2017; Published 22 May 2017

Academic Editor: Ben Niu

Copyright © 2017 Muhammad Shafiq et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Recently, machine learning (ML) algorithms have widely been applied in Internet traffic classification. However, due to the inappropriate features selection, ML-based classifiers are prone to misclassify Internet flows as that traffic occupies majority of traffic flows. To address this problem, a novel feature selection metric named weighted mutual information (WMI) is proposed. We develop a hybrid feature selection algorithm named WMI\_ACC, which filters most of the features with WMI metric. It further uses a wrapper method to select features for ML classifiers with accuracy (ACC) metric. We evaluate our approach using five ML classifiers on the two different network environment traces captured. Furthermore, we also apply Wilcoxon pairwise statistical test on the results of our proposed algorithm to find out the robust features from the selected set of features. Experimental results show that our algorithm gives promising results in terms of classification accuracy, recall, and precision. Our proposed algorithm can achieve 99% flow accuracy results, which is very promising.

## 1. Introduction

Accurate network traffic classification is extremely important for the network management including IP network management, deploying QoS-aware mechanisms, monitoring security, bandwidth management, and intrusion detection. For instance, it is useful for the Internet Service Providers (ISPs), network operators, and network administrators to understand the traffic composition and prioritize some sensitive bandwidth traffic such as video conferencing and voice over IP (VoIP). Moreover, from the perspective of network security, network traffic classification technique can help us in blocking unwanted or attack traffic.

In the last few years, several traffic classification models [1, 2] have been proposed in this regard. Traditionally, port-based technique was proposed, which is based on well-known port numbers for traffic classification. This technique is easy to be deployed and implemented. However, many Internet applications use dynamic port number for their communication instead of well-known port number, which makes it difficult for the network operator to identify network traffic

composition by port numbers. Moore and Papagiannaki [3] showed that port-based traffic classification technique does not give more than 50–70 percent accuracy.

To address the above problems, payload-based technique was proposed [4, 5] which inspects the packet payload signatures. Though this traffic classification technique is easy and accurate, this classification technique is ineffective for encrypted applications as numerous applications such as Skype use encrypted methods to protect their data from being detected. Furthermore, this technique is against the privacy laws to inspect the packet payload. In recent years, machine learning (ML) algorithms have been presented to classify Internet traffic flows, whose features are extracted from the flow statistics. Moreover, this ML technique is user privacy friendly and does not inspect the packet payload. However, the accurate feature selection problems challenge this technique. Features selection refers to the selection or filtration of accurate features in more than available features. For instance, Moore et al. in 2005 [6] presented the most wide features of extraction and selection method and they selected 248 statistical features, based on a whole traffic flow.

Using these selected statistical features, classifiers can achieve high performances results in Internet traffic classification. Nevertheless, in real words, it is not applicable to use these features in traffic classification. Recently, Zhang et al. [7] presented feature selection algorithm based on imbalance traffic classification. However, it is important to study more deeply accurate features selection techniques for Internet traffic classification.

In the previous work [8], we classify instant messages (IM) WeChat application message services (including text and picture messages) using four machine learning (ML) classifiers. We select 50 statistical flow-based features and got very effective accuracy results. Similarly, in [9, 10], we classify IM traffic accurately and find out effective features for IM traffic classification using machine learning algorithms. However, more than 50 features increase computational complexity and decrease accuracy results. That is why it is important to study more deeply accurate features selection technique for IM applications traffic classification.

In this paper, we proposed a feature selection algorithm to improve the performance of ML-based traffic classification technique and select effective features set for IM applications traffic classification. Our main contributions are given as follows:

- (i) In order to deal with accurate features selection problem in Internet traffic identification, a hybrid feature selection algorithm named WMI\_ACC is proposed. It includes two metrics for accurate features selection: weighted mutual information (WMI) metric and accuracy (ACC) metric of the selected classifiers. Firstly, WMI\_ACC algorithm assigns WMI values to the features. After assigning the WMI values, WMI\_ACC algorithm selects the best features with high ACC values for the specific classifier. Our study is the first study where the WMI metric is put forward in Internet traffic classification. Furthermore, this is the first time we apply WMI combined with ACC metrics for feature selection in traffic identification.
- (ii) We present the features that are selected by our proposed algorithm and report their actual values with metric values. Experimental results show that ten common flow-based features selected from different network environment traces have discriminative power for classifying instant messaging (IM) applications TCP and UDP flows. These features are (1) max\_fpktl, (2) mean\_fpktl, (3) max\_bpktl, (4) std\_fpktl, (5) min\_bpktl, (6) mean\_bpktl, (7) max\_bpktl, (8) std\_bpktl, (9) max\_fiat, and (10) total\_fpacket.
- (iii) Our third contribution is to select the robust features from the selected features of our proposed WMI\_ACC algorithm. We use Wilcoxon statistical test to find out the robust feature from the selected features.

The rest of the paper is organized as follows: Section 2 gives an overview of the related works. Section 3 elaborates our proposed WMI\_ACC algorithm. Section 4 demonstrates

the evaluation methodology with details, experimental work, and utilized datasets. Analysis and discussions are given in Section 5. Finally, conclusions and future works are shown in Section 6.

## 2. Related Work

Recently, machine learning (ML) algorithm has widely been applied in Internet traffic classification in [11–19]. Some of them are applied for traffic flow classification and some of them are applied for bandwidth management. However, most of the methods are applied for improving the performance of classification by using ML algorithms methods. These proposed methods are able to get 80% accuracy results. For feature extraction and selection, mostly using feature selection method as presented by Moore et al. in 2005 [6], based on whole traffic flow, they selected 248 statistical features, such as RTT and minimum, maximum, and average values of packet size. Using these selected statistical features, the applied classifiers can get very effective performance results in Internet traffic classification. But, in real circumstances, it is not good for traffic classification.

Thus, we must select accurate features for Internet traffic classification so that we can manage subsequent management and security policies. In 2012, in [7], Zhang et al. proposed two different algorithms for feature selection for optimization of traffic classification. They evaluated their results based on true positive rate (TPR) and false positive rate (FPR) and proved that their algorithm can achieve greater than 90% flow accuracy. Similarly, Peng et al. in 2016 [20] evaluated the effectiveness of statistical features. But their research study was limited to only early stage Internet traffic classification. Bernaille et al. [21] studied the problem of effective features in network traffic classification. In their study, they used  $K$ -Means and GMM and HMM model for Internet traffic classification. They selected packet size as a feature and extracted more features for early stage Internet traffic classification. Lim et al. in [22] used the packet size, connection level, and statistical features for Internet traffic classification. Van Der Putten and Van Someren in [23] revealed that features selection is very important for performance optimization compared to the choice of classification classifiers.

Zheng et al. in [24] showed that how many types of selected features metrics affects the identification performances. For accurate features selection purpose, Chen and Wasikowski in [25] proposed new feature selection metric using area under the ROC curve to evaluate features for Internet traffic classification. Kamal et al. [26] proposed three different filtering techniques, Balanced Minority Repeat (BMR), Differential Minority Repeat (DMR), and Higher Weight (HW), for the identification of effective features.

From the last few years, daily use of Internet applications increases day by day due to free of cost availability of instant messaging (IM) applications. It is also important to accurately classify IM applications. For instance, WeChat is an IM and free calling application developed by Tencent Holdings in China. After launching the WeChat application, its online users reached 300 million [27] and, in November 2015, its

active users reached 650 million all over the world. Apart from China, its online users reached 100 million [28] in the rest of the world. So, day to day increasing number of active users and traffic of this application can affect performance of the network. It is also important to classify WeChat messages and audio and video call traffic accurately to manage the quality of services (QoS) as Huang et al. [29] proposed measurement ChatDissect tool to measure WeChat application traffic and distinguish 150K users and 16 GB traffic of WeChat from real-world network traces. In 2013, Church and Rodrigo de Oliveira [30] studied the performance of mobile instant messaging sending service with traditional short messages. In 2014, O'Hara et al. [31] studied instant messaging application WhatsApp in smartphone and conducted some interviews and a survey to study the user activity using WhatsApp application. In 2014, Fiadino et al. [32] also studied WhatsApp application's flow stream and collected data in European Network, which consisted of millions of data flow streams. They also studied audio and video flow data stream. In 2014, Liu and Guo [33] studied video messaging services in WeChat and WhatsApp applications. They captured the traffic using mobile devices for their study. Furthermore, in our previous work in [8], we classify IM applications; however, we only classify WeChat text messages service flow traffic. Nevertheless, it is more important to select accurate features for IM applications traffic classification.

### 3. Proposed Method

In this section, we explain our proposed WMI\_ACC algorithm in detail. First of all, we examine the problem of features selection. Then we introduce WMI based feature metric and then design the WMI\_ACC algorithm that uses the WMI combined with ACC metric to select effective features for Internet traffic classification.

#### 3.1. Feature Selection Metrics

**3.1.1. Mutual Information Based Metric.** In information theory, mutual information is extensively used for features selection [9, 34], image processing [35], speech recognition [36], and so forth. It measures the mutual dependency between two random variables  $X$  and  $Y$ , which describes the amount of information held by random variable. The mutual information between two random variables is described as

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

In (1), the marginal entropies of  $X$  and  $Y$  are  $H(X)$  and  $H(Y)$ , while conditional entropies are  $H(X | Y)$  and  $H(Y | X)$  and joint entropies of  $X$  and  $Y$  are  $H(X, Y)$ . Moreover, the relationship between  $H(X)$ ,  $H(Y)$ ,  $H(X | Y)$ ,  $H(Y | X)$ ,

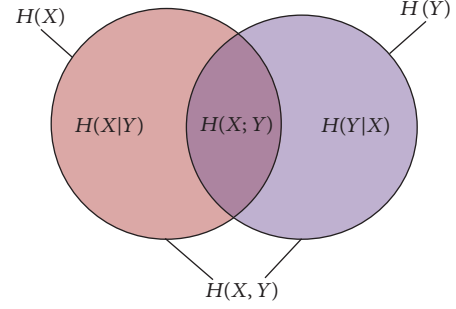


FIGURE 1: The relationship between mutual information and entropies.

$H(X, Y)$ , and  $I(X; Y)$  is shown in Figure 1. According to Shannon definition of entropy theory, we have

$$\begin{aligned} H(X) &= - \sum_{x \in X} p(x) \log(p(x)), \\ H(Y) &= - \sum_{y \in Y} p(y) \log(p(y)), \\ H(X, Y) &= - \sum_{x \in Y} \sum_{y \in Y} p(x, y) \log(p(x, y)), \end{aligned} \quad (2)$$

where  $p(\cdot)$  indicates the probability distribution of a random variable. As in [11], use the three equations in (1) to achieve the computational formula for mutual information. We also use the same method that the authors therein have used for mutual information.

$$I(X; Y) = - \sum_{x \in Y} \sum_{y \in Y} p(x, y) \log \left( \frac{p(x, y)}{p(x)p(y)} \right). \quad (3)$$

In case of continuous random variables, the summation will be replaced by a definite double integral.

$$I(X; Y) = \int_Y \int_X p(x, y) \log \left( \frac{p(x, y)}{p(x)p(y)} \right) dx dy. \quad (4)$$

**3.1.2. Weighted Mutual Information (WMI) Metric.** To address accurate features selection problem, we proposed weighted mutual information based on weighted entropy. If the total number of features is  $N$ , the weight value is calculated as follows:

$$W_i = 1 - \frac{ni}{N}, \quad (5)$$

where  $ni$  is the number of features assigned to features set; then the weighted mutual information (WMI) between two random variables can be defined as

$$\begin{aligned} I_w(X; Y) &= H_w(X) - H_w(XY) = H_w(X) - H_w(XY) \\ &= H_w(X) + H_w(XY) - H_w(X, Y) \\ &= H_w(X, Y) - H_w(XY) - H_w(XY). \end{aligned} \quad (6)$$

In (5), the weighted marginal entropies of  $X$  and  $Y$  are  $H_w(x)$  and so on. Again  $p(\cdot)$  is the probability distribution of random

```

Input:  $D(F_1, F_2, F_3, \dots, F_N, F)$  // training data set,
Output:  $feature []$  //selected feature set
(1) begin
(2) for  $i = 1$  to  $M$ 
(3) calculate weight value  $w[i]$  for each features;
(4) end for
(5) for  $i = 1$  to  $N$ ;
(6) calculate  $WMI(F_i)$ ;
(7) if  $(WMI(F) > \delta)$ ;
(8) insert  $F_i$  into descending order;
(9) end if
(10) end for
(11)  $F_p = \text{getfirstfeatures}(\text{list})$ ;
(12) end until  $(Fp == \text{NULL})$ ;
(13)  $X$  is a data set of samples
Values of features;
(14)  $\text{last\_accuracy} \leftarrow \text{classify } X$ ;
(15) Insert the feature into  $S_{\text{wrappers}}$ ;
(16)  $\text{Feature} = \text{get next features}$ ;
(17) For feature is not NULL
(18) insert the feature into  $S_{\text{wrappers}}$ ;
(19)  $X$  is a data set of samples values for  $S_{\text{wrappers}}$ ;
(20)  $\text{Accuracy} \leftarrow \text{classify } X$  with a specific classifiers;
(21) If  $(\text{ACC} \leq \text{last\_ACC})$ 
(22) Remove features from  $S_{\text{wrappers}}$ ;
(23) else
(24)  $\text{feature} = \text{getNextfeature}(\text{list}, \text{feature})$ ;
(25) end if
(26) end for
return  $S_{\text{wrappers}}$ ;

```

ALGORITHM 1: Feature selection algorithm based on weighted mutual information combined with ACC (WMI\_ACC).

variable. So, by using the three equations in (6), the weighted mutual information can be obtained as follows:

$$H_w(X) = - \sum_{x=X} w_i p(x) \log(p(x)),$$

$$H_w(Y) = - \sum_{x=Y} w_i p(Y) \log(p(Y)),$$

$$H_w(X; Y) = - \sum_{x \in Y} \sum_{y \in Y} w_i p(x) \log(w_i p(x)), \quad (7)$$

$$I_w(X; Y)$$

$$= \int_Y \int_X w_i p(x, y) \log \left( \frac{w_i p(x, y)}{w_i p(x) w_i p(y)} \right) dx dy.$$

For mutual information computational analysis, there is a bundle of software applications publically available, but we select H. Peng's mutual information Matlab toolbox [37] for our study.

**3.1.3. ACC Metric.** After using WMI metric, it is essential to select the effective features for specific ML classifier to obtain effective performance results. For this purpose, a wrapper method based on accuracy (ACC) metric is applied. On

the other hand, to achieve high performance accuracy with regard to classification of applications, the AUC metric is not suitable to rank the features. The highest ACC implies that ML classifier can obtain effective performance. Thus, we rank the features by using ACC metric and select those features with highest ACC values. We used C4.5, Random Forest, and Random Tree machine learning classifiers for ACC metric.

**3.1.4. WMI\_ACC Algorithm.** In this section, we propose effective features selection algorithm, named as WMI\_ACC. WMI\_ACC is a hybrid features selection algorithm based on WMI combined with ACC metric. Firstly, it filters most of the features with WMI metric and then selects the effective features with ACC metric for a specific algorithm. Algorithm 1 shows the detailed pseudocodes for our implemented features selection algorithm.

In Algorithm 1, there are two steps, given dataset  $D$  with  $M$  classes and  $N$  features. In the first step (lines (1)–(10)), WMI\_ACC algorithm filters most of the features with WMI value. The weight values for each of the features (line (3)) are calculated according to (4) (illustrated in Section 3.1.2). A good feature has greater mutual information values related to other features. WMI\_ACC firstly calculates the value of WMI between each of the features (line (6)). However, if the value of WMI is greater than the predetermined threshold value

(line (7)), it inserts features in the list in descending order. The greater threshold value speeds up the feature selection process but decreases the classification accuracy [20]. Thereafter, in line (11), the algorithm will get the list of WMI features set.

In the second step (lines (13)–(26)), WMI\_ACC algorithm selects effective features with ACC metric for a particular ML classifier. It gets the features from the desired list one by one and finds the feature that produces high ACC (accuracy) value. Exactly, from lines (13)–(16), firstly it achieves the values of ACC based on  $S_{\text{wrapper}}$  which consists of first feature list and then it takes the next feature from the list and then inserts it into  $S_{\text{wrapper}}$ . If the ACC value of new inserted feature is low, WMI\_ACC algorithms remove the features from the list in line (21). Lastly,  $S_{\text{wrapper}}$  includes the effective features set.

**3.1.5. Statistical Test.** In more depth, to select the robust features from the selected features list of our proposed WMI\_ACC algorithm and to find the significant difference among the results of the applied method, statistical tests are conducted. In this study, we executed Wilcoxon pairwise statistical test on the results of methods [38, 39]. The detailed introduction to the Wilcoxon pairwise statistical test is given as follows:

- (i) *Wilcoxon test:* we also used Wilcoxon signed-rank statistical test in this research. Wilcoxon test is also a nonparametric method used for pairwise comparison between two methods [40] and is also used in many research areas [9]. If  $d_i$  is the variance between two methods' performance scores on  $i$ th out of  $n$  problem and the score is in different ranges, then it can be normalized on interval 0 and 1 in [41]. Afterwards, the variations are ranked by their absolute values and in ties practitioner will be conducted on one method as in [42]. In this case, the positive values indicate that the method performed well and vice versa.

$$R^+ = \sum_{d_i > 0} \text{rank}(d_i) + \frac{1}{2} \sum_{d_i = 0} \text{rank}(d_i), \quad (8)$$

$$R^- = \sum_{d_i < 0} \text{rank}(d_i) + \frac{1}{2} \sum_{d_i = 0} \text{rank}(d_i).$$

$R^+$  is used for the sum of positive values and  $R^-$  will be used for the sum of negative variation values. It means that if the difference between these  $R^-$  and  $R^+$  is very high, then the hypothesis will be disallowed, that is, rejected. This statistical test is also used like Friedman test to determine whether the hypothesis will be rejected or not on the specific significant values  $\alpha$ .

## 4. Evaluation Methodology

This section includes traces traffic, evaluation criteria, and analysis of experimental results.

TABLE 1: Characteristics of HIT Trace 1 dataset.

Application	Duration time	# instances	Date
WTCP	1 hour	20512	28 April 2016
WUDP	1 hour	16400	28 April 2016
P2P	1 hour	1501	27 December 2015
IM	1 hour	7911	27 December 2016
IMAP	1 hour	15832	27 December 2015
FTP	1 hour	25251	27 December 2015

**4.1. Datasets.** In this paper, we select two sets of network traces for our experimental study. One dataset is our set of traces collected in our lab, while the other set is an open network trace dataset. The selected two traces are different network environment datasets. We applied our proposed feature selection algorithm on both datasets, respectively, not on only one dataset for better understanding of the composition of Internet traffic. We used two different network environment datasets, because these datasets are different from each other; for example, in our trace dataset, we capture mostly WeChat instant messaging application's traffic, while in NIMS dataset GTALK IM application's traffic is traced.

**4.1.1. HIT Trace 1 Dataset.** In this paper, we used the same dataset that we have used in our previous paper in [9]. However, for developing HIT Trace 1 dataset, we capture WeChat instant messaging (IM) application traffic. WeChat IM application includes multiple functions, but we only trace text messages, pictures messages, and audio and video calls traffic; also, in more depth, we trace IM, IMAP, and FTP applications' traffic for our research study. In this research study, we are interested in finding out the effective features for IM application traffic classification. Thus, we trace only WeChat IM application text messages, pictures messages, and audio and video calls traffic, respectively, with a Wireshark tool [43]. We captured the traffic with duration of one hour for our research study at the laboratory at School of Computer Science & Information Technology, Harbin Institute of Technology, Harbin, China, on 27 December 2015 and 28 April 2016. It should be noted that we only trace the traffic that has none zero payload packets and we are only interested in TCP and UDP traffic of WeChat IM application and P2P, IM, IMAP, and FTP traffic. In this dataset, WTCP traffic and WUDP traffic mean TCP traffic and UDP traffic of WeChat application. The detailed characteristics of HIT Trace 1 dataset are shown in Table 1.

**4.1.2. NIMS Dataset.** NIMS dataset includes packets collected at the research tested network. The dataset consists of SSH servers outside connection and application behaviors traffic such as DNS, HTTP, SFTP, and P2P traffic. However, we are also interested in instant messaging applications traffic classification. In this case, we also added NIMS GTALK trace traffic, which includes TCP GTALK traffic and UDP GTALK traffic. Moreover, in NIMS dataset, we select only DNS, HTTP, SFTP, GTALK TCP, and GTALK UPD traffic for

TABLE 2: Characteristics of NIMS dataset.

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

TABLE 3: Accuracy result of HIT Trace 1 dataset.

Applications	Bayes Net	Naïve Bayes	C4.5	R/Forest	Random Tree
WTCP	98.40	92.68	99.56	99.82	99.70
WUDP	100	97.75	99.56	99.87	99.93
P2P	99.95	97.78	99.87	99.93	99.95
IM	94.09	92.08	94.39	91.59	90.94
IMAP	93.31	91.81	94.17	91.42	91.05
FTP	99.98	97.53	99.85	99.98	99.95

our research work study. The detailed characteristics of NIMS data are shown in Table 2.

**4.2. Performance Measures.** For the evaluation performance of five machine learning (ML) classifiers/algorithms, classification accuracy, recall, and precision values are employed. All the measuring metrics are described as follows:

- (i) Classification accuracy: it is the number of correctly classified traffic flows divided by total classified flows
- (ii) Recall: it is the percentage of specific traffic flows Class Z correctly classified as belonging to that Class Z
- (iii) Precision: it is the percentage of the traffic flows which exactly have Class Z between all those that were classified as Class Z

These performance evaluation metrics are important for flow-based traffic classification in network traffic identification. However, flow accuracy is used to measure the overall performance of an ML classifier.

**4.3. Experimental Results and Analysis.** In this section, our objective is to evaluate the performance of our proposed algorithm, comparing the results of HIT Trace 1 dataset with NIMS dataset's results. Our experiments include three phases. First of all, on HIT dataset and NIMS dataset, we validate that our proposed feature selection algorithm is effective for feature selection with respect to accuracy results. Then, we validate that our proposed algorithm is effective with precision and recall results. Table 3 depicts the classification accuracy results of HIT trace dataset using Bayes Net, Navies

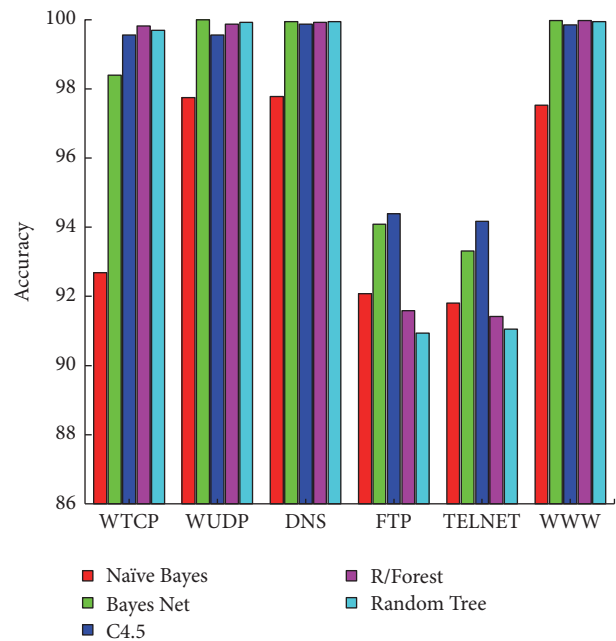


FIGURE 2: Accuracy result of HIT Trace 1 dataset.

Bayes, C4.5 decision tree, Random Forest, and Random Tree machine learning algorithms. Figure 2 shows the detailed accuracy result chart of HIT Trace 1 dataset. However, we use Weka application for our experiments using training and testing method to classify IM applications traffic accurately.

It is clear from Table 3 and Figure 2 that the applied machine learning classifiers give maximum classification

TABLE 4: NIMS dataset's accuracy results.

Applications	Bayes Net	Naïve Bayes	C4.5	R/Forest	Random Tree
GTALK TCP	100	98.60	100	100	99.99
GTALK UDP	100	86.55	100	100	100
DNS	99.78	86.43	100	100	100
FTP	100	99.34	99.97	100	100
HTTP	100	98.58	99.97	100	99.99
SFTP	100	99.97	100	100	100

TABLE 5: Recall results of HIT dataset.

Applications	Bayes Net	Naïve Bayes	C4.5	R/Forest	Random Tree
WTCP	97.81	96.09	99.61	99.74	99.61
WUDP	100	92.25	100	100	100
P2P	96.10	89.61	90.91	94.81	96.10
IM	36.82	14.77	38.86	31.14	45.91
IMAP	80.91	49.90	80.32	50.89	31.61
FTP	100	89.69	92.79	98.97	99.95

accuracy results and values, but C4.5 decision tree machine learning classifier provides overall maximum classification accuracy for HIT Trace 1 dataset, 97.90%, which is effective classification accuracy value as compared to other machine learning classifiers' accuracy results. On the other side, all applications are classified with very effective values; however, FTP and Telnet applications are slightly classified when compared to other classified applications for HIT Trace 1 dataset. Telnet and FTP applications give maximum 94% accuracy results but WTCP, WUDP, and WWW applications are classified with very effective accuracy results as compared to other applications. While classifying Telnet and FTP applications, the maximum accurate results are given by Bayes Net and C4.5 machine learning classifiers when compared to other machine learning classifiers.

From these experimental results, it is evident that Random Forest and C4.5 machine learning classifiers give better performance in terms of classification accuracy as compared to other machine learning classifiers for the NIMS dataset. However, Random Forest machine learning classifier gives very effective results in terms of classification accuracy. The details are shown in Table 4 and Figure 3. In NIMS dataset, SFTP application is classified 100% as compared to other traffic applications and the applied machine learning classifiers give very accurate identification results for SFTP. The applied machine learning classifiers give very accurate performance results for NIMS dataset, but Naïve Bayes ML classifiers give slightly low accuracy results for all traffic applications except SFTP application. However, all the traffic applications are classified using five machine learning classifiers accurately.

Similarly, Bayes Net machine learning classifier gives better recall values for HIT Trace 1 dataset as shown in Table 5 and Figure 4. All the traffic applications are classified accurately with respect to recall metrics, but IM and IMAP traffic applications give very poor results with respect to recall metrics, particularly the IM application that gives very low performance results of recall metric for HIT Trace 1 dataset

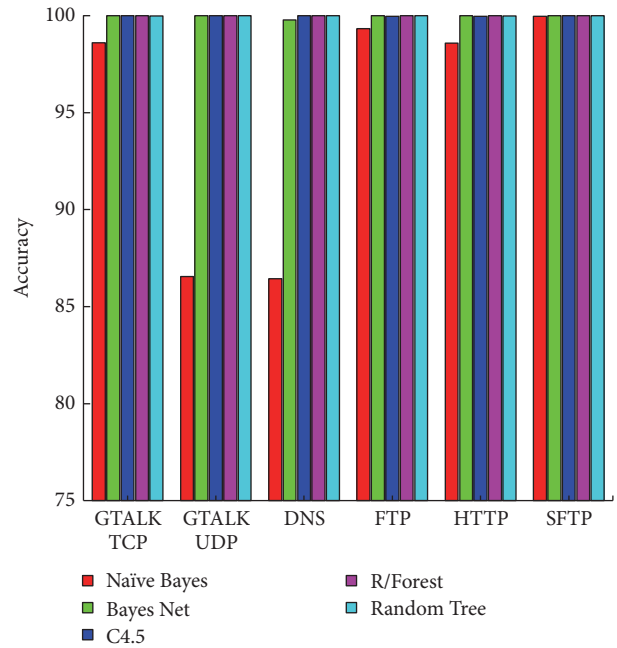


FIGURE 3: Accuracy result of NIMS dataset.

traffic classification, while ML classifiers C4.5, Bayes Net, and Random Tree give effective recall results. From the table, WTCP application is classified very effectively using the ML classifiers with respect to recall metrics but Random Forest ML classifier gives maximum recall results as compared to other machine learning classifiers and then Random Tree and C4.5 ML classifiers give maximum same performance recall results using WTCP application. Similarly, using WUDP traffic application, only Naïve Bayes ML classifier gives low results and all other ML classifiers give 100 recall results for WUDP traffic application. For P2P, Random Forest and Bayes Net give good recall results, while, for IM, Random Forest

TABLE 6: Precision result of HIT dataset.

Applications	Bayes Net	Naïve Bayes	C4.5	R/Forest	Random Tree
WTCP	99.63	92.73	99.69	99.97	99.92
WUDP	100	99.31	99.81	99.97	99.74
P2P	100	36.51	98.59	100	100
IM	60.67	37.14	63.81	35.69	36.27
IMAP	54.41	50	58.89	44.83	39.75
FTP	97.98	39.55	97.83	100	100

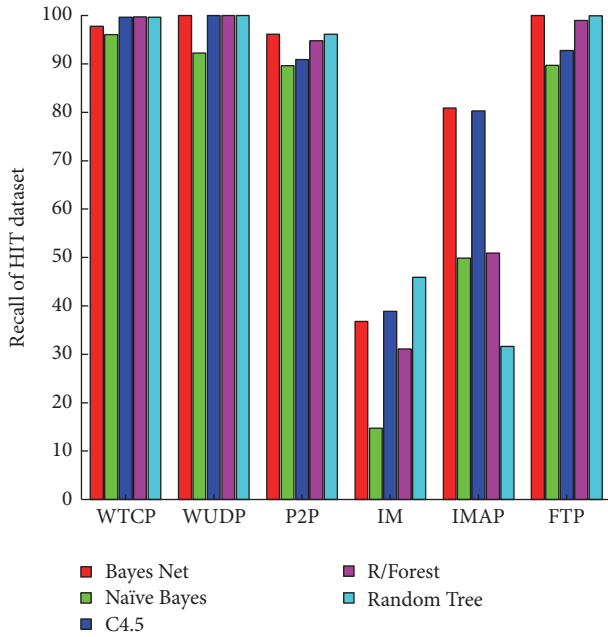


FIGURE 4: Recall results for HIT Trace 1 dataset.

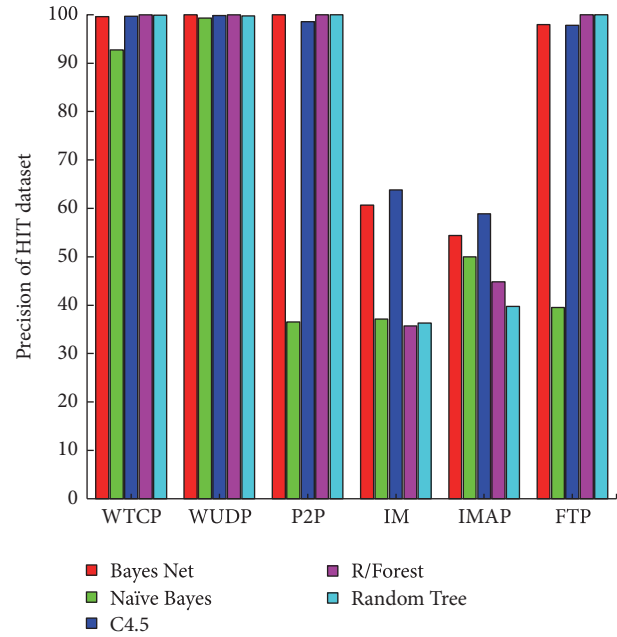


FIGURE 5: Precision results for HIT Trace 1 dataset.

gives maximum recall results. Similarly, for FTP, Bayes Net and Random Tree give maximum recall results.

However, for precision results, as shown in Table 6 and Figure 5, Random Forest machine learning classifier gives effective precision results for HIT Trace 1 dataset. It is clear from the experimental results using HIT Trace 1 dataset that the entire selected machine learning classifiers get high performance results values in terms of classification accuracy, recall, and precision.

Though all the traffic applications are classified very efficiently using machine learning classifiers, IM and IMAP traffic applications give very low precision performance results for HIT Trace 1 dataset as compared to other traffic applications. Using machine learning algorithms, WTCP application is classified efficiently and the applied ML classifiers give good results with respect to precision metric, which are more than 99% results. Similarly, WUDP traffic application is also classified very accurately and the applied classifiers got more than 99% precision results, but Bayes Net classifier gives 100% precision results, which are promising precision results. For P2P application, Random Forest, Random Tree, and Bayes Net ML classifiers give 100% precision results, while IM application traffic is classified very effectively and

the applied classifiers do not get more than 63% precision results. Similarly, IMAP application is also poorly classified with respect to precision metric. However, FTP traffics are classified with respect to precision values and mostly classifiers got 100% precision results for HIT Trace 1 dataset.

From the experimental result of NIMS dataset, it is evident again that C4.5 and Random Forest ML classifiers give very effective precision results for NIMS dataset. However, all the applied ML classifiers give very attractive results but C4.5 and Random Forest ML classifiers' results are very promising results in terms of precision values. Using ML classifiers, all the applied machine learning classifiers give very effective precision results for GTALK TCP application, but Random Forest, C4.5, and Bayes Net ML classifiers give 100% results. For GTALK UDP application, all the applied classifiers got good precision results, but only Naïve Bayes got low precision result. Similarly, for DNS, FTP, HTTP, and SFTP, all the applied classifiers got promising precision results but only Naïve Bayes and Random Tree got slightly low precision results. However, all the precision results are good using NIMS dataset.

Recalling results for NIMS dataset, C4.5, Random Forest, and Random Tree ML classifiers give accurate results in terms



TABLE 7: Precision result of NIMS dataset.

Applications	Bayes Net	Naïve Bayes	C4.5	R/Forest	Random Tree
GTALK TCP	100	58.11	100	100	99.42
GTALK UDP	99.26	71.85	100	100	100
DNS	100	91.77	100	100	100
FTP	100	100	100	100	100
HTTP	100	99.58	100	100	99.92
SFTP	100	100	100	100	100

TABLE 8: Recall results of NIMS dataset.

Applications	Bayes Net	Naïve Bayes	C4.5	R/Forest	Random Tree
GTALK TCP	100	100	100	100	100
GTALK UDP	100	91.63	100	100	100
DNS	99.47	75.05	100	100	100
FTP	100	89.88	99.48	100	100
HTTP	100	90.60	99.97	100	100
SFTP	100	99.63	100	100	100

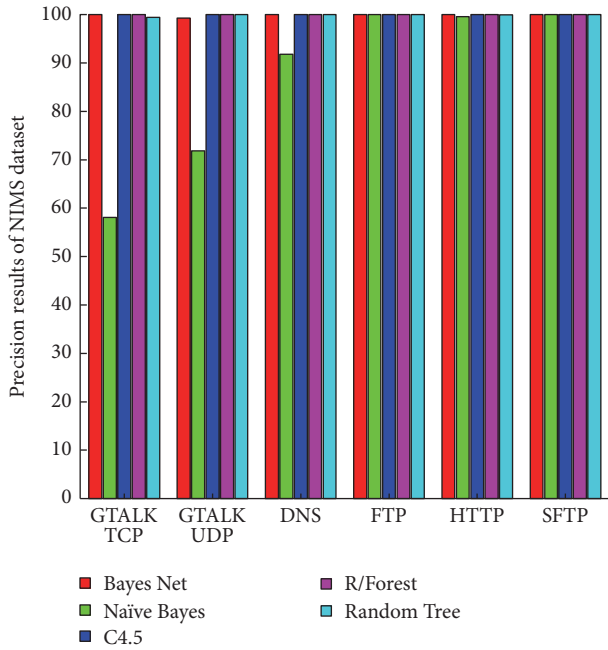


FIGURE 6: Precision results for NIMS dataset.

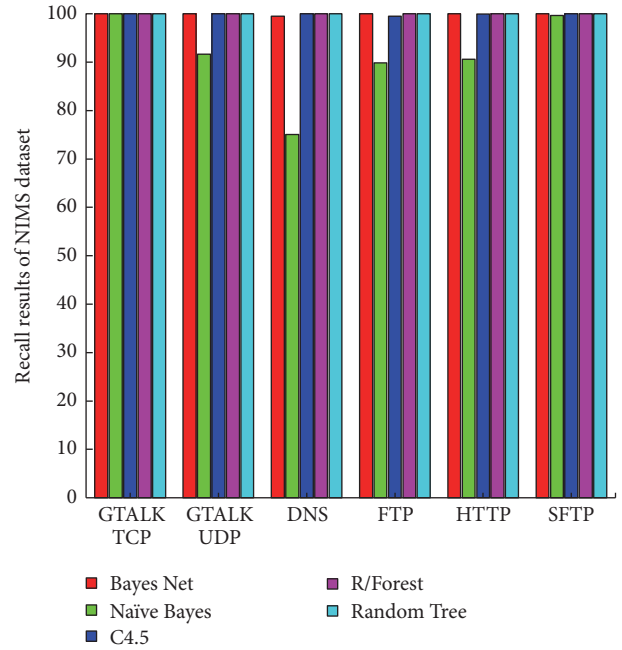


FIGURE 7: Recall results for NIMS dataset.

of recall for the NIMS dataset. The detailed results are shown in Tables 7 and 8 and Figures 6 and 7. Similarly, all the utilized traffic applications of NIMS dataset are classified accurately with respect to recall metric but SFTP and GTALK TCP got 100% recall results using the applied ML classifiers. From the classifiers' point of view, only Naïve Bayes ML classifier got low recall results as compared to other ML classifiers. Moreover, using ML classifiers for GTALK TCP application, the traffic is classified very accurately as all the applied ML classifiers got 100% recall results, while for GTALK UDP only Naïve Bayes ML classifier gives slightly low recall results. For DNS application, only Naïve Bayes got poor recall value and

the remaining applications got 100% recall values. Similarly, FTP, HTTP, and SFTP applications are classified accurately about 100% but only Naïve Bayes gives slightly low recall results.

**4.3.1. Wilcoxon Pairwise Statistical Test Result.** Table 9 shows the Wilcoxon pairwise test results for the robust features selection from the selected features of WMI\_ACC proposed algorithm. From Table 9,  $p$  value of features is greater than 0.05 for the accuracy results. Thus, we conclude that there is no significant difference between the results of 9 features and other features for the selected features. We conclude that

TABLE 9: Wilcoxon pairwise test results.

9 Ftrs versus min_fpctl	Accuracy		
	$R^+$	$R^-$	$p$ value
Mean_fpctl	0.00	3.00	0.180
Max_fpctl	0.00	1.00	0.317
Std_fpctl	3.00	3.00	1.000
Min_bpctl	0.00	1.00	0.317
Mean_bpctl	3.00	3.00	0.317
Max_bpctl	0.00	1.00	1.000
Std_bpctl	3.00	3.00	1.000
Max_fiat	5.00	1.00	0.276
Total_fpackets	0.00	1.00	0.317

TABLE 10: Selected feature of our proposed algorithm.

S. number	Feature name
1	Mean_fpctl
2	Max_fpctl
3	Std_fpctl
4	Min_bpctl
5	Mean_bpctl
6	Max_bpctl
7	Std_bpctl
8	Max_fiat
9	Total_fpackets
10	Max_fpctl

all the selected features of our proposed algorithm are very effective.

**4.3.2. Selected Features of Our Proposed Algorithm.** The features that are selected by our proposed algorithm are given in Table 10.

**4.3.3. Average Accuracy Comparison of Datasets.** In this section, we compare average accuracy results of both (HIT and NIMS) datasets to find out the effective machine learning classifier out of five applied machine learning classifiers. In Figure 8, it is clear that NIMS dataset application traffic flows are classified very effectively and all the applied ML classifiers got very promising accuracy results as compared to HIT Trace 1 dataset. However, all the applied ML classifiers got very effective average accuracy results, but Bays Net ML classifier gives maximum average performance results for HIT Trace 1 dataset as compared to other applied classifiers. Similarly, using NIMS dataset, the applied classifiers performances are very effective, but Random Forest and Random Tree ML classifiers' performance results are very effective as shown in Table 11. However, using Wilcoxon statistical test, we conclude that Random Forest ML classifier is a very effective classifier for IM applications traffic classification.

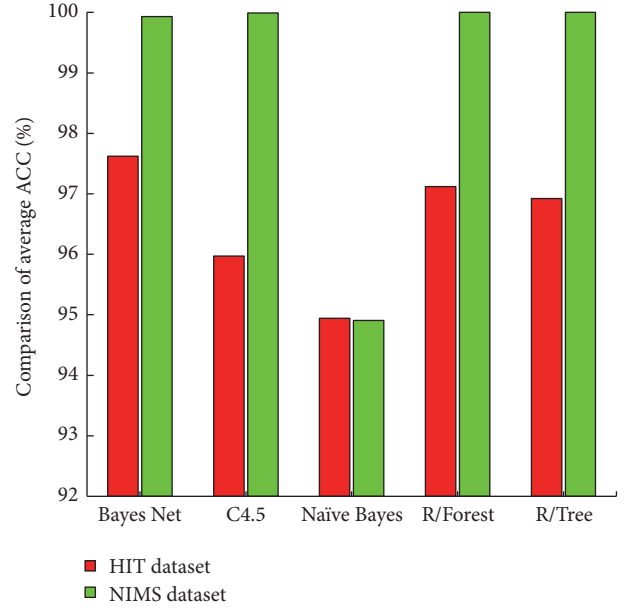


FIGURE 8: Datasets' average accuracy comparison.

## 5. Analysis and Discussion

Though the results of the five applied machine learning classifiers are different with respect to accuracy, recall, and precision using HIT Trace 1 dataset and NIMS dataset, some information can be obtained from experimental study for IM traffic classification:

- (i) From this study, it is clear that our proposed algorithm selects effective features set for IM traffic classification using two different network environment datasets in terms of classification accuracy, recall, and precision metrics.
- (ii) From the experimental results, all the applied machine learning classifiers give very effective performance results for all application classifications, but only FTP and Telnet applications are classified a little bit low in both utilized datasets as compared to other applications.
- (iii) In this research study, our proposed algorithm gives effective features sets and it is evident that all the features carry enough identification information for IM traffic classification.
- (iv) Through accuracy results, the classification performance can be easily evaluated for the instant messaging (IM) traffic classification. But, in some cases, some classifiers get high identification performance results and in some cases they do not get very effective results. It is due to imbalance traffic composition found in the datasets.
- (v) We discuss that all the applied ML classifiers give very effective performance results. However, C4.5 decision tree and Random Forest ML classifiers give very accurate performance results as compared to other machine leaning classifiers.

TABLE II: Datasets' average accuracy comparison.

Algorithms	Average accuracy of HIT dataset	Average accuracy of NIMS dataset
Bayes Net	97.62%	99.93%
C4.5	95.97%	99.99%
Naïve Bayes	94.94%	94.91%
Random Forest	97.12%	100%
Random Tree	96.92%	100%

## 6. Conclusion

This paper proposed feature selection algorithm named WMI\_ACC used to select effective features for IM traffic classification. The performance of our proposed algorithm WMI\_ACC is very promising for 5G IM traffic classifications. The experimental results showed that our approach is able to improve the classification accuracy, recall, and precision mostly in 5G high dimension traffic. Furthermore, ten flow-based features selected by our approach are very important for 5G IM traffic classification. They are (1) max\_fpctl, (2) mean\_fpctl, (3) max\_bpctl, (4) std\_fpctl, (5) min\_bpctl, (6) mean\_bpctl, (7) max\_bpctl, (8) std\_bpctl, (9) max\_fiat, and (10) total\_fpacket. Using Wilcoxon pairwise statistical test, it is evident from the experimental study that these features carry enough classification information. Moreover, all the applied ML classifiers get very effective performance results, but we found that C4.5 and Random Forest ML classifiers with WMI\_ACC selected features have very effective performance as compared to other applied machine learning classifiers. In our experiments, some ML classifiers get very efficient performance results in terms of classification accuracy, recall, and precision and some ML classifiers get little bit low classification results. It is due to imbalance of dataset. However, there is still a gap for further research in the 5G instant messaging (IM) traffic classification. A new approach should be designed to select robust feature for IM applications traffic classification and this is our future research work.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

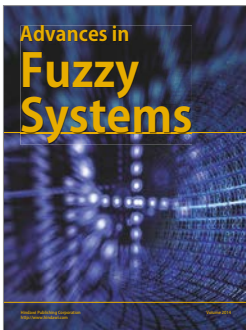
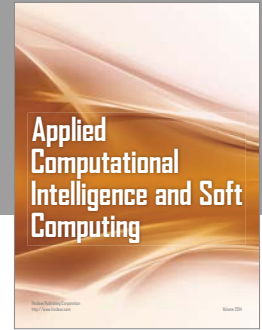
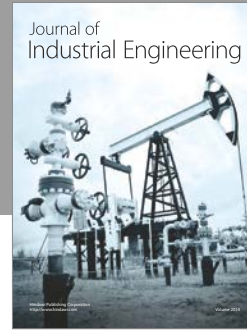
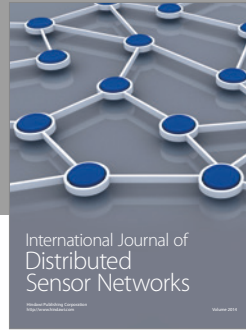
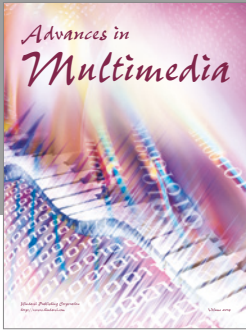
## Acknowledgments

This work was supported by National Natural Science Foundation of China under Grant no. 61571144.

## References

- [1] T. T. T. Nguyen and G. Armitage, "A survey of techniques for internet traffic classification using machine learning," *IEEE Communications Surveys and Tutorials*, vol. 10, no. 4, pp. 56–76, 2008.
- [2] P. A. W. E. L. Foremski, "On different ways to classify Internet traffic: a short review of selected publications," *Theoretical and Applied Informatics*, vol. 25, 2013.
- [3] A. W. Moore and K. Papagiannaki, "Toward the accurate identification of network applications," in *Proceedings of the International Workshop on Passive and Active Network Measurement*, Springer, Berlin, Heidelberg, Germany, 2005.
- [4] S. Sen, O. Spatscheck, and D. Wang, "Accurate, scalable in-network identification of P2P traffic using application signatures," in *Proceedings of the 13th International World Wide Web Conference (WWW '04)*, pp. 512–521, ACM, May 2004.
- [5] P. Haffner, S. Sen, O. Spatscheck, and D. Wang, "ACAS: automated construction of application signatures," in *Proceedings of the ACM SIGCOMM 1st Workshop on Mining Network Data (MineNet '05)*, pp. 197–202, ACM, August 2005.
- [6] A. Moore, D. Zuev, and C. Michael, *Discriminators for Use in Flow-Based Classification*, Queen Mary and Westfield College, Department of Computer Science, 2005.
- [7] H. Zhang, G. Lu, M. T. Qassrawi, Y. Zhang, and X. Yu, "Feature selection for optimizing traffic classification," *Computer Communications*, vol. 35, no. 12, pp. 1457–1471, 2012.
- [8] M. Shafiq, X. Yu, and A. A. Laghari, "WeChat text messages service flow traffic classification using machine learning technique," in *Proceedings of the 6th International Conference on IT Convergence and Security (ICITCS '16)*, pp. 1–5, IEEE, Prague, Czech Republic, September 2016.
- [9] M. Shafiq and X. 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.
- [10] M. Shafiq, X. Yu, A. A. Laghari, L. Yao, N. K. Karn, and F. Abdesssamia, "WeChat text and picture messages service flow traffic classification using machine learning technique," in *Proceedings of the IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS), IEEE 18th International Conference on (IEEE '16)*, 2016.
- [11] L. Peng, B. Yang, and Y. Chen, "Effective packet number for early stage internet traffic identification," *Neurocomputing*, vol. 156, pp. 252–267, 2015.
- [12] L. Peng, H. Zhang, B. Yang, Y. Chen, M. T. Qassrawi, and G. Lu, "Traffic identification using flexible neural trees," in *Proceedings of the IEEE 18th International Workshop on Quality of Service (IWQoS '10)*, June 2010.
- [13] G. Lu, H. Zhang, X. Sha, C. Chen, and L. Peng, "TCFOM: A robust traffic classification framework based on OC-SVM combined with MC-SVM," in *Proceedings of the 2010 International Conference on Communications and Intelligence Information Security (ICCIIS '10)*, pp. 180–186, chn, October 2010.
- [14] A. W. Moore and D. Zuev, "Internet traffic classification using bayesian analysis techniques," *ACM SIGMETRICS Performance Evaluation Review*, vol. 33, no. 1, p. 50, 2005.

- [15] T. Auld, A. W. Moore, and S. F. Gull, "Bayesian neural networks for internet traffic classification," *IEEE Transactions on Neural Networks*, vol. 18, no. 1, pp. 223–239, 2007.
- [16] D. A. Cieslak, N. V. Chawla, and A. Striegel, "Combating imbalance in network intrusion datasets," in *Proceedings of the IEEE International Conference on Granular Computing (GrC '06)*, pp. 732–737, May 2006.
- [17] D. Nechay, Y. Pointurier, and M. Coates, "Controlling false alarm/discovery rates in online internet traffic flow classification," in *Proceedings of the 28th Conference on Computer Communications, IEEE (INFOCOM '09)*, pp. 684–692, April 2009.
- [18] W. Li, M. Canini, A. W. Moore, and R. Bolla, "Efficient application identification and the temporal and spatial stability of classification schema," *Computer Networks*, vol. 53, no. 6, pp. 790–809, 2009.
- [19] D. G. Gomes, N. Agoulmine, Y. Bennani, and J. N. de Souza, "Predictive connectionist approach for VoD bandwidth management," *Computer Communications*, vol. 30, no. 10, pp. 2236–2247, 2007.
- [20] L. Peng, B. Yang, Y. Chen, and Z. Chen, "Effectiveness of statistical features for early stage internet traffic identification," *International Journal of Parallel Programming*, vol. 44, no. 1, pp. 181–197, 2016.
- [21] L. Bernaille, R. Teixeira, I. Akodkenou, A. Soule, and K. Salamati, "Traffic classification on the fly," *ACM SIGCOMM Computer Communication Review*, vol. 36, no. 2, pp. 23–26, 2006.
- [22] Y.-S. Lim, H.-C. Kim, J. Jeong, C.-K. Kim, T. Kwon, and Y. Choi, "Internet traffic classification demystified: on the sources of the discriminative power," in *Proceedings of the 6th International Conference on Emerging Networking Experiments and Technologies (Co-NEXT '10)*, December 2010.
- [23] P. Van Der Putten and M. Van Someren, "A bias-variance analysis of a real world learning problem: the CoLL challenge 2000," *Machine Learning*, vol. 57, no. 1-2, pp. 177–195, 2004.
- [24] Z. Zheng, X. Wu, and R. Srihari, "Feature selection for text categorization on imbalanced data," *ACM SIGKDD Explorations Newsletter*, vol. 6, no. 1, pp. 80–89, 2004.
- [25] X.-W. Chen and M. Wasikowski, "FAST: a roc-based feature selection metric for small samples and imbalanced data classification problems," in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '08)*, pp. 124–132, August 2008.
- [26] A. H. M. Kamal, X. Zhu, A. Pandya, and S. Hsu, "Feature selection with biased sample distributions," in *Proceedings of the International Conference on IEEE Information Reuse and Integration (IRI '09)*, 2009.
- [27] WeChat attracts 300m users in less than 2 years, [http://www.chinadaily.com.cn/cndy/2013-01/17/content\\_16128915.htm](http://www.chinadaily.com.cn/cndy/2013-01/17/content_16128915.htm).
- [28] By the numbers: 50+ Amazing WeChat Statistics, <http://expandedramblings.com/index.php/wechat-statistics/>.
- [29] Q. Huang, P. P. C. Lee, C. He, J. Qian, and C. He, "Fine-grained dissection of WeChat in cellular networks," in *Proceedings of the 23rd IEEE International Symposium on Quality of Service (IWQoS '15)*, pp. 309–318, Portland, Ore, USA, June 2015.
- [30] K. Church and R. de Oliveira, "What's up with whatsapp?: comparing mobile instant messaging behaviors with traditional SMS," in *Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '13)*, pp. 352–361, ACM, Munich, Germany, August 2013.
- [31] K. P. O'Hara, M. Massimi, R. Harper, S. Rubens, and J. Morris, "Everyday dwelling with whatsapp," in *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing*, 1131 pages, ACM, Baltimore, Md, USA, February 2014.
- [32] P. Fiadino, M. Schiavone, and P. Casas, "Vivisecting WhatsApp through large-scale measurements in mobile networks," in *Proceedings of the ACM conference on (SIGCOMM '14)*, pp. 133–134, usa, August 2014.
- [33] Y. Liu and L. Guo, "An empirical study of video messaging services on smartphones," in *Proceedings of the 24th ACM Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV '14)*, pp. 79–84, ACM, March 2014.
- [34] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.
- [35] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Transactions on Medical Imaging*, vol. 16, no. 2, pp. 187–198, 1997.
- [36] L. R. Bahl, P. Brown, P. de Souza, and R. Mercer, "Maximum mutual information estimation of hidden Markov model parameters for speech recognition," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '86)*, Tokyo, Japan, April 1986.
- [37] H. Peng, Mutual information Matlab toolbox, <http://www.mathworks.com/matlabcentral/fileexchange/14888-mutual-information-computation>.
- [38] S. García, A. Fernandez, J. Luengo, and F. Herrera, "Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: experimental analysis of power," *Information Sciences*, vol. 180, no. 10, pp. 2044–2064, 2010.
- [39] F. Gringoli, L. Salgarelli, M. Dusi, N. Cascarano, F. Risso, and K. C. Claffy, "GT: picking up the truth from the ground for internet traffic," *ACM SIGCOMM Computer Communication Review*, vol. 39, no. 5, pp. 12–18, 2009.
- [40] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, CRC Press, 2003.
- [41] D. Quade, "Using weighted rankings in the analysis of complete blocks with additive block effects," *Journal of the American Statistical Association*, vol. 74, no. 367, pp. 680–683, 1979.
- [42] J. D. Gibbons and S. Chakraborti, *Nonparametric Statistical Inference*, Springer, Berlin, Heidelberg, Germany, 2011.
- [43] Trace traffic WireShark, <http://www.wireshark.org/> (last Accessed Sept 2015).



**Hindawi**

Submit your manuscripts at  
<https://www.hindawi.com>

