On Monitoring and Predicting Mobile Network Traffic Abnormality

Yingxu Lai[1,*], Yinong Chen[2], Zenghui Liu[3], Zhen Yang[1], Xiulong Li[1]

1. *College of Computer Science, Beijing University of Technology, Beijing 100124, China*

[*] *Corresponding author. Tel.: +86 13439195095; fax: +86 01067391742. E-mail address: laiyingxu@bjut.edu.cn*

2. *School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe 85287, USA*

3. *Automation Engineering Institute, Beijing Polytechnic, Beijing 100176, China*

Abstract

Traffic analysis and traffic abnormality detection are emerged as an efficient way of detecting network attacks in recent years. The existing approaches can be improved by introducing a new model and a new analysis method of network user's traffic behaviors. The description dimensions to network user's traffic behaviors in the current approaches are high, resulting in high processing complexity, high delay in differentiating an individual user's abnormal traffic behavior from massive network data, and low detection rate. To improve the detection rate and efficiency, we develop a new method of establishing user's traffic behavior analysis system based on a new model of network traffic monitoring. First, we establish a more complete feature set based on the characteristics of network traffic to describe massive network user's behaviors. Then, we define a feature selection rule based on the relative deviation distance to select the optimized feature set. We use the selected feature set to locate the abnormality moment and the users who produce the abnormal traffic behavior. Finally, a traffic behavior analysis method based on prediction is developed to improve efficiency of the system. This new method is applied to evaluate the mobile users on mobile cloud. The experimental results show that the proposed method has a higher detection rate and lower delay in the analysis of abnormal user's traffic behavior than that of the existing approaches.

Key words: network attack, traffic behavior; traffic monitoring; feature set; feature selection

1. INTRODUCTION

With the explosive development of mobile applications and the support of Cloud Computing (CC) for a variety of services for mobile users, Mobile Cloud Computing (MCC) [1] is introduced as an integration of CC into the mobile environment. MCC is a model developed as a solution to overcome the challenges in mobile devices using CC services like storage and computing resources [2-3]. Experience with Internet-based services has shown that attacks from worms and viruses such as Code Red and SQL Slammer are common threats to network-oriented applications. Threats also exist in mobile devices, cloud side, and the communication channels, through attacks such as packet injection and Man-in-the-Middle. Mobile devices can access MCC in many different ways, including voice service, Short Messaging Service (SMS) and other Internet service through phone networks. In addition, most smart phones can also access to the network through Wi-Fi and Bluetooth. Wider ranges of communication ways will bring more security threats such as the sensitive information leakage or malicious attacks.

As MCC is a combination of mobile and CC, the security attacks come from the mobile device, the cloud, and the communication channels. To secure user data and applications, the mobile platform (e.g. Android, iOS, and Windows) providers implemented a variety of different strategies to secure data [4]. However, applying a high level of protection implies a burden on performance and a high level of energy consumption of the mobile device. A commonly used solution for channel security is using SSL protocols. However, these protocols are on one hand high energy consuming and on the other hand provide security properties as a block without taking into account the type of data transmitted or the user expectations for phone applications. On the server side, mobile cloud providers are responsible for securing the data in the cloud and possibly data exchanged between the devices and the cloud. Different solutions have been and are being developed to secure the data access [5-6].

In addition to preventive techniques, security analyses and threat monitoring are essential to the overall security. This paper focuses on mobile cloud security analysis and threat monitoring.

MCC security services should monitor network traffic from mobile users in real time and block any unsecure operation behaviors. Roschke [7] and Vieira [8] summarized users' requirements and proposed an extensible intrusion detection system (IDS) architecture for being easily used in a distributed cloud infrastructure. Dastjerdi [9] proposed a Mobile Agent based IDS which has been customized for CC environment in order to satisfy the cloud user's security requirements. However, due to the frequent changes of user requirements, operation behaviors and other service parameters, it is difficult and expensive to perform a real-time supervision and control in these proposed systems.

Users' traffic behavior and pattern analysis are key techniques for monitoring MCC. There are several challenges on the analysis of the massive users' traffic behaviors: (1) it is difficult to describe user's traffic behaviors because of the complexity of cloud traffic; (2) it is time consuming to analyze massive user's traffic behaviors because of the complex behavior description; and (3) it is time consuming to differentiate and identify individual network user's behaviors from massive network users' data. In order to address these challenges, Farraposo [10] developed a user's traffic behavior analysis method based on the IPFIX protocol. The method was able to detect most of the abnormal behaviors appeared in the network, but couldn't detect abnormal behaviors which had less impact on the network traffic. Cheng [11] and He [12] used cluster method to define and differentiate the normal cluster from the abnormal ones. When user's traffic behaviors that deviated from the normal cluster exceeded a preset threshold, the user would be considered to be an abnormal user. However, in the real network environment, it is difficult to construct a comprehensive normal traffic behavior's model, which can be used to differentiate different abnormal behaviors. Lakhinaetet al. [13-17] used Wavelet Analysis to detect traffic abnormality in a progressive way. Barford et al. [18] proposed a signal analysis-based network traffic abnormalities. Paxsonet al. [19] used a novel statistical method for cloud traffic classification. All these methods were time consuming. To overcome above shortages, a network traffic prediction method based on nonlinear preprocessing was proposed by Yang [20] to detect abnormal traffic behaviors. This method was able to predict the abnormal traffic behavior emerged in network. The shortage was that the threshold needed to be preset, which would either reduce the accuracy or increase false alarm rate based on the value selected. Daihee [21] proposed a network traffic analysis model using online analytical processing (OLAP) on a multidimensional data cube, which provided an easy and fast way to construct a multidimensional traffic analysis system for

comprehensive and detailed analysis of traffic data. The threshold needed to be preset too in this approach. Energy conversation is another key domain of study in mobile cloud. Mavromoustakis and Karatza proposed efficient methods in mobile devices for real-time traffic analysis and energy conservation [22-23], and Du et al. studied the energy conservation from cloud side [24].

Although many efforts have been made in traffic pattern collection and analyses, it remains a challenge to analyze massive users' traffic behaviors in terms of accuracy and efficiency. This paper attempts to address the issue by developing a more complete feature set based on the characteristics of network traffic to describe user's traffic behaviors. The method is based on network traffic monitoring and prediction, which can timely detect an abnormal traffic behavior when it emerges and identifies the user who produces the abnormal traffic behavior at the abnormality moment. In addition, the future behavior will be predicted rapidly and accurately.

The proposed method consists of these steps: first, we establish a complete feature set based on the characteristics of network traffic to describe user's traffic behavior. Then, a feature selection rule based on the relative deviation distance is proposed to select the optimized feature sets. We use these feature sets to detect the abnormal traffic behavior. Finally, a traffic behavior analysis method based on prediction is proposed to improve the efficiency of the system. The experimental results show that the proposed method has a higher detection rate and higher efficiency in the analysis of abnormal user's traffic behavior than the existing methods.

The rest of the paper is organized as follows. Representation of network user's traffic behavior and feature selection rule are presented in section 2. The user's traffic behavior system based on traffic monitoring is given in section 3. The method based on prediction that is used to analyze the massive users' traffic behavior is studied in section 4. Finally, the conclusions and future work are given in section 5.

2. REPRESENTATION OF TRAFFIC BEHAVIORS AND FEATURE SELECTION RULES

In MCC, Mobile IP (MIP) is the mobility enabling protocol developed by the Internet Engineering Task Force (IETF) to support global mobility in IP networks. If a mobile host connects to a home network, it is assigned an IP address on its home network, called the mobile host's home address. Packets from a correspondent host to the mobile host are always addressed to the home address. If the mobile host connects to a foreign network, it can acquire its care-of address either from a foreign agent or through auto-configuration methods, such as DHCP, designed for assigning temporary IP addresses. In this paper, we analyze mobile users' behaviors based on CC traffic, which are composed mainly of IP packets. Mobile users' traffic will be mingled with other non-mobile terminals' (like PC, etc.) in CC, and thus our method will be compatible with other terminal users' traffic analysis.

This section first defines the representation of network user's traffic behaviors. We use network traffic features to describe user's traffic behaviors, thus, a relatively complete set of network traffic features is proposed to describe the user's traffic behaviors. The feature set is given in section 2.1. Then the feature selection rule is given in section 2.2.

2.1. Feature set of network traffic

The feature set is obtained by analysis of the related field of the IP packet:<time, sip, sport,

dip, dport, protocol, direction, pack_len, head_len, tcp_head_len, type, syn, ack, fin>, where time field is the time when the packet is captured; sip and dip fields are the IP packet's source and destination IP address; sport and dport fields are the IP packet's source and destination port; pack_len, head_len and tcp_head_len fields are the length of IP packet; IP packet header and TCP packet header; type field is the type of ICMP packet; syn, ack and fin field are the connection synchronous field; connection acknowledgement field and connection finish field of TCP packet header.

Through the analysis of the IP packet fields, we use 110 network traffic behavior features as the feature set, which includes static and dynamic features. Static features mainly include the features like packet length, IP address, port and special packet. Dynamic features mainly include the features like speed, distribution, connection and integration. All the features are based on the statistic of the proposed vector above. For example, the type of packet length is based on the statistic of time, protocol, direction, pack_len, head_len and tcp_head_len fields of the vector. Traffic behavior of network users can be described completely and accurately with the proposed feature set.

The complete network traffic feature set is defined in Table 1. The features are described by the sequence of <protocol\statistical parameter (IP address / port / TCP connection / header.Actions.Description)>. For example, feature 77 is TAN (ip.c.bd), which use the mean number of TCP connection established with independent IP address as a feature.

Table 1. Network traffic feature set

	1	2	3	4	5	6	7	8	9	10
0	IN.u	CN.u	TN.u	UN.u	IB.u	CB.u	TB.u	UB.u	IN.d	CN.d
1	TN.d	UN.d	IB.d	CB.d	TB.d	UB.d	IAB.u	CAB.u	TAB.u	UAB.u
2	IAB(H.u)	TAB(H.u)	IBv.u	CBv.u	TBv.u	UBv.u	IBv(H.u)	TBv(H.u)	IAB.d	CAB.d
3	TAB.d	UAB.d	IAB(H.d)	TAB(H.d)	IBv.d	CBv.d	TBv.d	UBv.d	IBv(H.d)	TBv(H.d)
4	INr(u.d)	CNr(u.d)	TNr(u.d)	UNr(u.d)	IBr(u.d)	CBr(u.d)	TBr(u.d)	UBr(u.d)	CN.R	TN.R
5	UN.R	CB.R	TB.R	UB.R	TNR(dpt80)	TBR(dpt80)	CN.ip	TN.ip	UN.ip	UN.spt
6	UN.dpt	TN.spt	TN.dpt	TN(c.rq)	TN(c.bd)	TN(c.kp)	CN.erq	CN.erp	IN(H.ov)	TN(H.ov)
7	CN(d.sm)	TN(d.sm)	UN(d.sm)	CN(u.bg)	TN(u.bg)	UN(u.bg)	TAN(ip.c.bd)	TMN(ip.c.bd)	IAN(ip.u)	IMN(ip.u)
8	IAN(ip.d)	IMN(ip.d)	IAB(ip.u)	IMB(ip.u)	IAB(ip.d)	IMB(ip.d)	IAN(ip.u.bg)	IMN(ip.u.bg)	IAN(ip.d.sm)	IMN(ip.d.sm)
9	TAN(ip.spt)	TMN(ip.spt)	UAN(ip.spt)	UMN(ip.spt)	TAN(ip.dpt)	TMN(ip,dpt)	UAN(ip.dpt)	UMN(ip.dpt)	TAN(c.u)	TMN(c.u)
10	TAN(c.d)	TMN(c.d)	TAB(c.u)	TMB(c.u)	TAB(c.d)	TMB(c.d)	TAN(c.u.bg)	TMN(c.u.bg)	TAN(c.d.sm)	TMN(c.d.sm)

Color code: speed packet length distribution IP address port connection special packet integration

Abbreviations: I: IP protocol; C: ICMP protocol; T: TCP protocol; U: UDP protocol; A: average; M: maximum; N: number; B: the number of bytes; v: variance; r: the ratio of two numbers; R: the ratio of the total; ip: independent ip; spt: independent source port; dpt: independent destination port; c: independent tcp connection; H: header; u: upload; d: download; rq: request; bd: the establishment; kp: maintained; bg: packet length greater than 1000 bytes; sm: packet length less than 100 bytes; ov: packet header length greater than 20 bytes; erq: echo request packet; erp: echo reply packet.

2.2. Feature selection rule

Based on the characteristics of network traffic, we denote the feature set as $F = \{f_1, f_2, ..., f_n\}$, where, each feature is described by f_i, where, $i=1,2,..., n$ and $n = 110$. The feature selection based on the feature set is necessary for the improvement of analysis efficiency. Denning [25] pointed out that the network traffic can be described by normal distribution statistic model. According to the Central Limit Theorem, if the random variable X can be expressed as the sum of many independent random variables $x_1, x_2, ..., x_n$, then, as long as each variable x_i ($i = 1,2, ..., n$) has a slight influence on X and n is large enough, X is considered to follow the normal distribution, no matter what distribution x_i is. Because the massive users' traffic composed of many independent user's traffic, each of which has a slight influence on overall traffic, thus, we approximately model the overall network traffic to follow the normal distribution. Definition 1 presents the method to calculate the relative deviate distance of feature f_i.

Definition 1: The relative deviate distance (D_i) of feature f_i is defined:

$$D_i = \frac{|f_i - E(f_i)|}{\sigma_i} \tag{1}$$

Where $E(f_i)$ and σ_i are the expectation and standard deviation values of feature f_i respectively. Then based on definition 1, the feature's deviation is defined:

Definition 2: Feature's deviation is defined as $\frac{(D_i)_A}{(D_i)_N}$, where $(D_i)_A$ and $(D_i)_N$ are the relative deviate distance of feature f_i with abnormal and normal user's training set respectively, $i=1,2,...,110$.

Finally, we define a feature selection rule to reduce the feature dimension based on definition 2, the rule is defined as:

Rule 1: Feature selection rule of network user's traffic behavior.

Features which meet equation (2) will be selected.

$$\frac{\frac{(D_i)_A}{(D_i)_N}}{\sum_{i=1}^{110}\frac{(D_i)_A}{(D_i)_N}} \geq threshold \tag{2}$$

where, $threshold \in (0,1]$.

3 USER'S TRAFFIC BEHAVIOR ANALYSIS BASED ON TRAFFIC MONITORING

3.1. User's traffic behavior analysis system

In this section, we identify the abnormality moment from the analysis of the massive users' traffic, and the abnormal users emerged at the abnormality moment will be located by the analysis of an individual user's traffic behaviors. Because the time consumption of locating the abnormal user is higher than locating abnormality moment, we first analyze the massive users' traffic behaviors, and then analyze individual user's behaviors only when the massive users' traffic is abnormal. The structure of the network user's traffic behavior analysis system is shown in Figure 1. The system includes four modules: data preprocessing, feature selection, massive users' traffic behavior analysis, and abnormality user's locating. Data preprocessing module is mainly

responsible for data processing of the network traffic. We extract the feature in the feature selection module, where the proposed feature selection rule is used to select feature set that is suitable for user's traffic behavior analysis. The massive users' traffic behavior analysis module is responsible for the massive users' traffic behavior analysis. The massive users' Bayesian classifier is used to locate the abnormality moment. Abnormal user's locating module is responsible for the location of abnormal users who emerged at the abnormality moment.

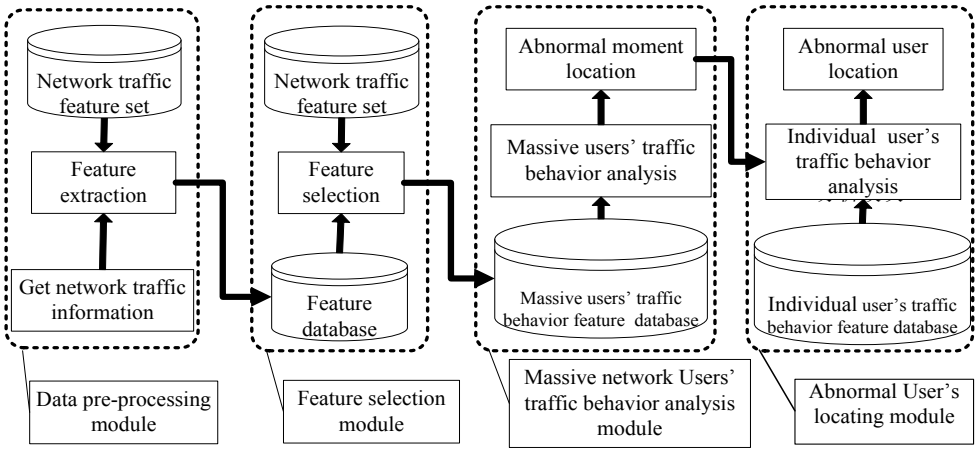


Figure 1. User's traffic behavior analysis system based on network traffic monitoring

3.2. Traffic behavior analysis for massive users

We first analyze the massive users' traffic behaviors to locate the moment when users' abnormal behaviors emerged. The process is as follows:

First, we build a time window which is used to extract feature. Within the time window, we extract 110 features from normal and abnormal traffic, respectively. Then we select features based on Rule 1, a Bayesian classifier is then constructed with the selected features used for the analysis of the massive network users' traffic behaviors. Finally, we use the Bayesian classifier to locate the moment quickly when abnormal behaviors emerged.

3.3. Abnormal user's locating

Abnormal user's locating is necessary when the abnormality moment is located in sub-section 3.2. In this sub-section, features selection in sub-section 3.2 will be used to analyze individual user's traffic behaviors. We use the Bayesian classifier to identify an abnormal user when abnormal behavior emerged.

3.4. Naive Bayesian classification algorithm

Network traffic has statistical characteristic, and follows a specific probability distribution [23]. Naive Bayesian classification algorithm [27] is suitable for the data which follows a probability distribution. So we use Naive Bayesian classification algorithm to classify abnormal traffic behaviors. Assuming F to be an unknown traffic sample, the Bayes' Theorem can be described in equation (3).

$$P(C_i \mid F) = \frac{P(C_i)P(F \mid C_i)}{P(F)} \tag{3}$$

Assuming the traffic sample F is described by the combination of features $F = \{f_1, f_2, \ldots, f_n\}$, then, equation (4) can be obtained by the assumption of independence of Bayesian.

$$P(F \mid C_i) = P(f_1, f_2, \ldots, f_n \mid C_i) = \prod_{k=1}^{n} P(f_k \mid C_i) \tag{4}$$

Here, C is the number of categories. $(C_i \mid F)$ is the probability of any random sample having category C_i. $(C_i \mid F)$ is the probability of the sample having the features that it does assuming that the sample comes from category C_i.

3.5. *Experimental results analysis*

We use two kinds of experiments to prove our method. Experiment 1 is for mobile devices, and experiment II is for mobile devices and PC.

3.5.1. Data sets

Normal data sets of mobile devices used in experiment I are the wireless network data collected by 275 freshmen with HP Jornada PDAs at the University of California, San Diego [28]. These PDAs are equipped with symbol 802.11b compact Flash card, which use 802.11b protocol to communicate with others.We use the network layer data of this wireless data set.

Normal data sets of mingled users in experiment II consist of five parts: flow trace data collected by the WIDE project team in the backbone network links [29], flow trace data collected by the University of Waikato in campus network [30], flow trace data collected by the University of Auckland in campus network [31], flow trace data collected by the CAIDA organization in the Equinix-sanjose backbone network link [29], the NLANR PMA dataset provided by RIPE NCC organization [33].

Abnormal data is consisted of two parts, one part is the Conficker worm dataset, the ICMP DDOS 2007 dataset [32] and the Witty worm dataset published by CAIDA organization; Another part is the LLDOS1.0 and LLDOS2.0.2 datasets [34] published by the MIT Lincoln Laboratory in 2000.

3.5.2. Evaluation parameters

We use quantitative indicators TP, TN, FP, FN to have an effective evaluation of the experimental results, where TP and TN indicate the number of abnormal and normal samples which are classified correctly respectively, FP and FN indicate the number of abnormal and normal samples which are classified wrongly respectively. Test indicators include detection rate TPR and false alarm rate FNR, it is calculated as equation (5) and equation (6).

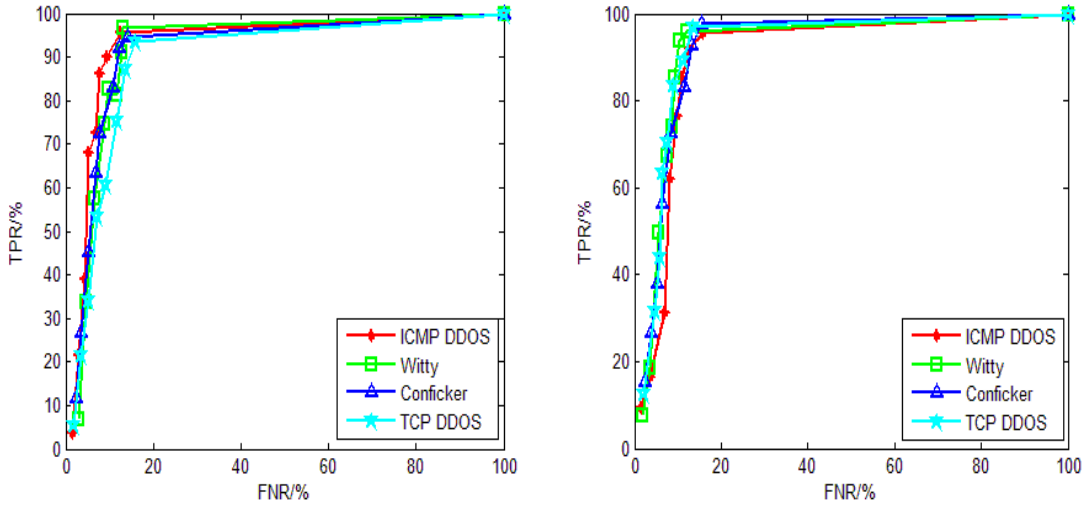
$$TPR = \frac{TP}{TP + FP} \tag{5}$$

$$FNR=\frac{FN}{TN+FN} \tag{6}$$

In order to verify the effectiveness of the proposed feature selection rules, we select classic network attacks of ICMP DDOS, Witty, Conficker and TCP DDOS. Samples of normal traffic are 10 minutes of normal flow; Samples of abnormal traffic are 5 minutes of ICMP DDOS, Witty, Conficker and LLDOS1.0 traffic datasets. The abnormal test samples are the four attack traffic and normal traffic mixed as ratio 1:9, respectively.

3.5.3. Experimental I: Mobile Users

Based on the model and the parameters discussed in the previous sections, the mobile users' traffic behavior analysis results are shown in Figure 2. In this paper we use the receiver operating characteristic (ROC) curve as a performance measure. In general, better decision or detection performance is indicated by an ROC curve that is higher and to the left in the ROC space. Figure 2.a is ROC curves of abnormality moment location, Figure 2.b is ROC curves of abnormal users' location. As can be seen from Figure 2's ROC curves, the system has a high detection rate for the analysis of mobile user's abnormal traffic behaviors and a low false alarm rate, which verifies the effectiveness of the proposed method for the analysis of mobile user's traffic behaviors.



a. Locate for abnormality moment b. Locate for abnormal users

Figure 2. ROC curve t of mobile users' traffic behavior

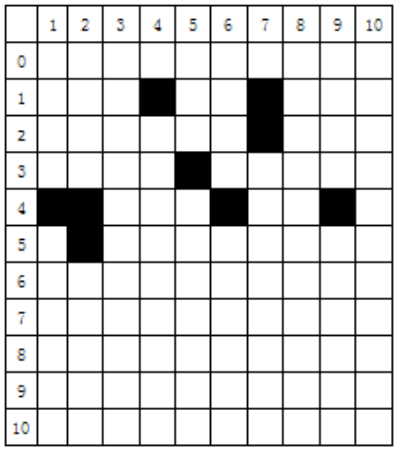
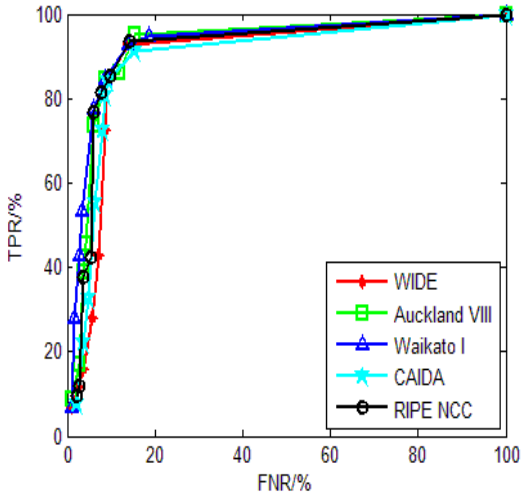
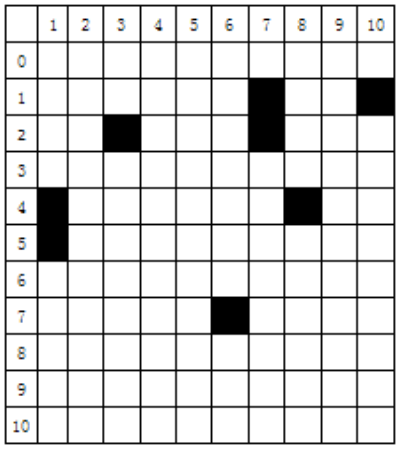
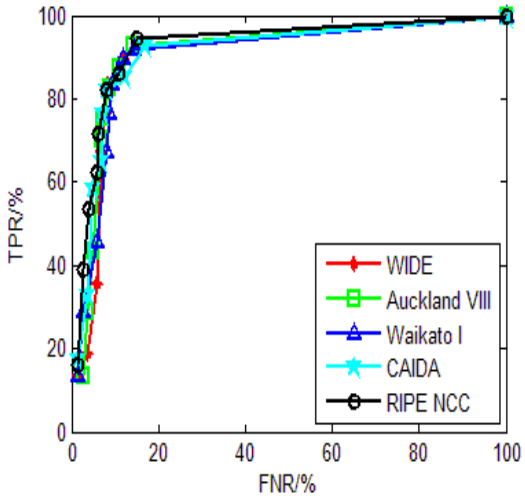
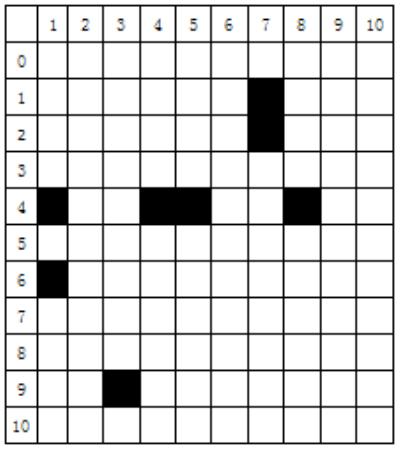
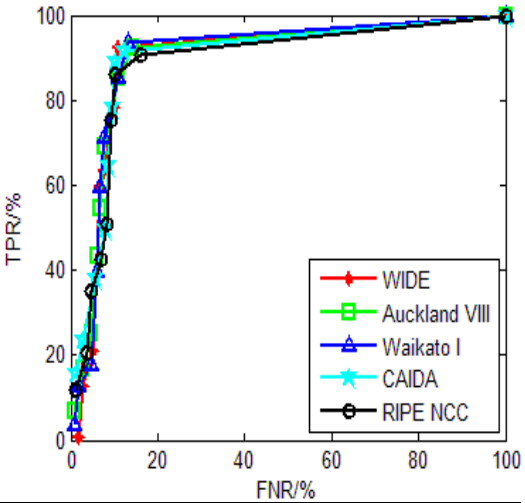
3.5.4. Experimental II: mingled users

1. Results of analyzing massive users' traffic behaviors

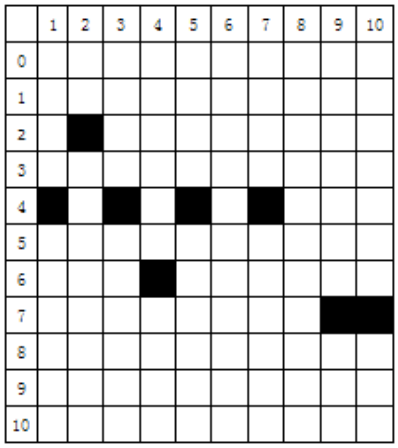
(1) Feature selection and abnormal traffic behavior analysis

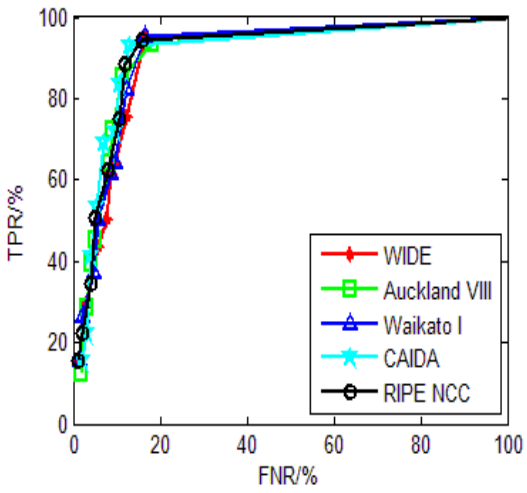
The massive users' traffic behavior analysis results are shown in Table 2. The left column is feature selection result, the shadow parts are selected features, such as in ICMP DDOS attack, the selected features are 14, 17, 27, 35, 41,42,46,49 and 52. The right column the ROC curve by the classifier based on selected features.

Table 2. The massive network users' behavior analysis result

Type of attack	Feature selection results(WIDE dataset)	ROC curve
ICMP DDOS		
Witty worm		
Conficker worm		

TCP DDOS





We can find the feature subset for the analysis of each abnormal traffic behavior in Table 2. ICMP DDOS is a distributed denial of service attack by ICMP protocol, the attacker uses multiple hosts to send a large number of ICMP echo request packets to the target host in order to exhaustits resources. Feature subset of ICMP DDOS mainly includes features such as speed, packet length and distribution. Witty worm will randomly generate a large number of IP addresses and scans these IP addresses when it enters the network. Feature subset of Witty worm mainly includes feature type of packet length, distribution and special packet. Conficker worm is a typical hidden worm. It will reduce the scan rate in order to avoid system's detection. However, in order to spread, it will send a large number of UDP packets to scan the target host's system. Feature subset of Conficker worm mainly includes feature type of packet length, distribution, port and integration. TCP DDOS is a distributed denial of service attack which used TCP protocol. Feature subset of TCP DDOS mainly includes feature type of packet length, distribution and connection. Feature subset contains a few of features, which effectively reduced the complexity. If we contrast the attack characteristics and the selected features, we can find that the features selected by Rule 1 have practical significance. The experimental results show the effectiveness of the feature selection rule.

As can be seen from Table 2's ROC curves, the system has a high detection rate for the location of abnormality moment with a low false alarm rate.

(2) Analysis of the efficiency

The massive users' analysis efficiency results are shown in Table 3.

Table 3. The efficiency of the massive network users' traffic behavior analysis

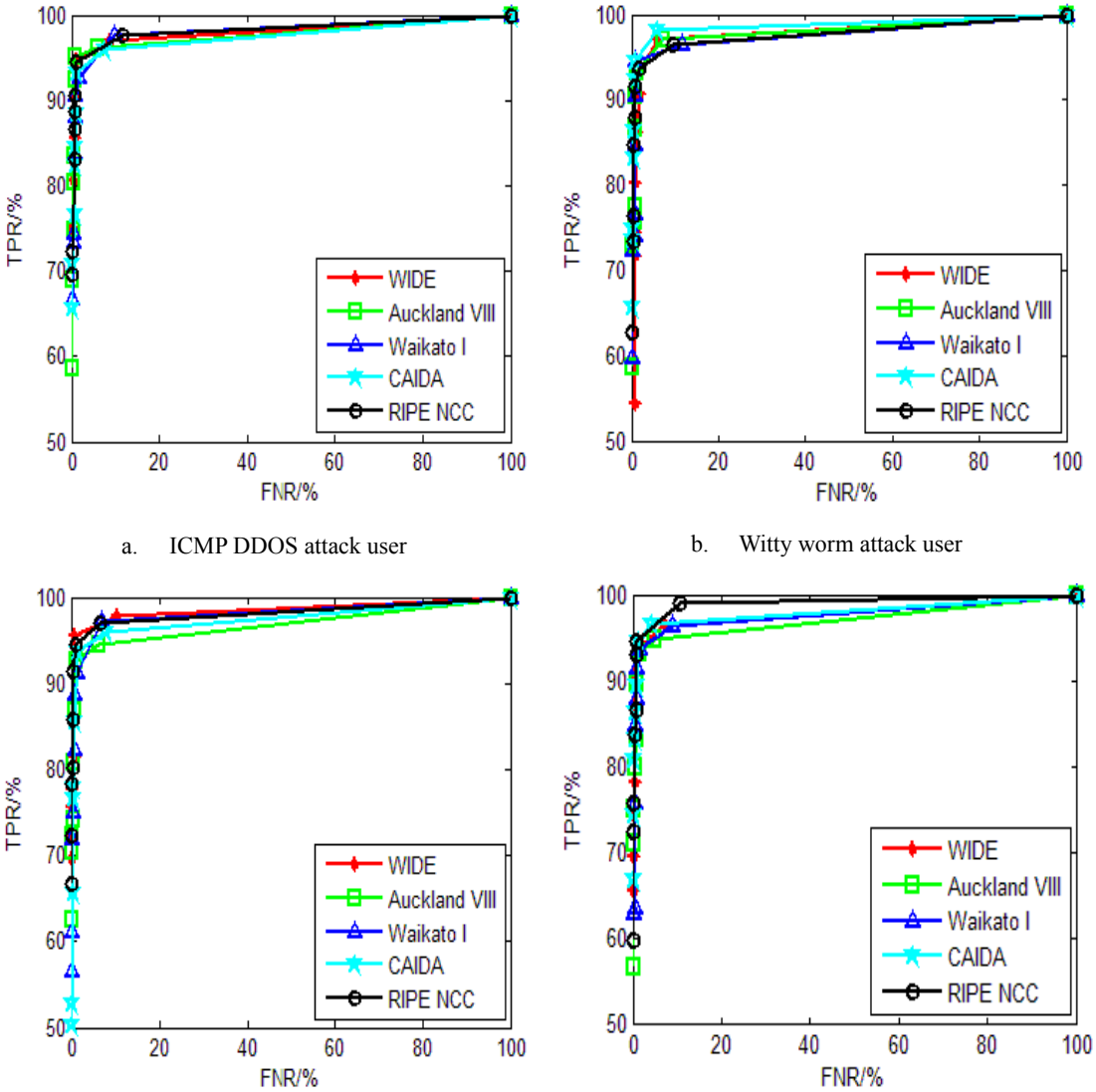
Number of training samples	Training time (ms)	Number of test samples	Test time (ms)
600	164.318	600	37.645

As can be seen from Table 3, the system has a high efficiency for the analysis of massive network users' traffic behaviors. The abnormality moment can be located rapidly. Therefore, the system can be used for massive users' real-time analysis.

2. The analysis of abnormal user's location

(1) Abnormal user's location

After the abnormality moment has been located, the users who product traffic at that moment will be analyzed. We use the Bayesian classifier to identify an abnormal user or multiple abnormal users. Features are same as those are used to analyze traffic behaviors. In our test dataset, at the abnormality moment, there are nearly 5000 users to product traffic. In these users, there are about 10 abnormal users. The results of abnormal user's location are shown in Figure 3.



a. ICMP DDOS attack user

b. Witty worm attack user

c. Conficker worm attack user

d. TCP DDOS attack user

Figure 3. Results of abnormal user's location

As can be seen from ROC curves in Figure 3, almost abnormal users have been identified. The system has a high detection rate for the location of abnormal user's with a low false alarm rate.

(2) Analysis of the efficiency

If we contrast the efficiency of the massive users' analysis and the individual user's analysis, we can find that the efficiency of the former is much higher than that of the latter, as shown in Table 4. This is why we first analyze the massive users' traffic behaviors.

Table 4. The efficiency of abnormal user'slocation

Number of training samples	Training time (min)	Number of test samples	Test time (min)
178881	22.251	211718	1.162

In summary, the system has a high detection rate for the location of abnormality moment and abnormal users with a low false alarm rate. However, the method to monitor the massive network traffic will lead to longer delay. To address this problem, we propose a new method in section 4 to improve the efficiency of the massive traffic users' analysis.

4. ANALYSIS OF MASSIVE USER'S TRAFFIC BEHAVIORS BASED ON PREDICTION

In this section, we propose a new analysis method that is based on the prediction to obtain the future traffic behaviors and to reduce analysis delay. We first smooth the time sequence of traffic behavior features that are selected in section 3.2. Then we construct ARMA model for these features sequence. Finally, we predict the future behavior's feature value and prejudge the future behaviors.

4.1. Prediction model

We construct feature sequence model based on the rule of the minimum sum of squares of the remaining residual, the process to construct the model is as follows:

1. In general, the network user's traffic behavior feature sequence is not stable [35], which will have an influence on constructing of the model. Thus, we use the nonlinear preprocessing method to smooth the sequence. The method of logarithmic transformation $x_t = \lg x_t$ will be used in this approach.

2. Begin with p=1, gradually increasing the model order to fit the ARMA (p, p-1) model, and the model parameters are evaluated by the method of least squares. The model that has minimum sum of squares of the remaining residual will be selected.

4.2. Prejudgment for abnormal traffic behaviors

We propose a prejudgment method for traffic behaviors in this section in order to reduce the delay caused by the traffic analysis.

We construct the ARMA model for features which are selected in sub-section 4.2, and then we use these models to predict the future features value. The model average error is calculated as equation (7):

$$MAPE = \frac{\sum_{i=t+1}^{t+l} \left| \frac{x_i - \hat{x}_i}{x_i} \right|}{l} \times 100\% \tag{7}$$

Where, x_i is the actual feature value at the moment of i, $\hat{x}_i$ is the predicted value at the moment of i, l is the length of the prediction interval.

This method is based on the assumption that the time sequence will fluctuate within a certain range in a short term. So we believe that the deviation between the predictive value and the true

value should be within a certain range. If the deviation exceeds this range, the abnormal traffic behaviors may emerge.

The proposed prejudgment method consists of the following steps. We first obtain the maximum deviation V_{max}, which is calculated by equation (7) among these features sequence. And then the abnormal deviation threshold value is set as m $(m \geq 1)$ times of V_{max}. The moment will be prejudged as abnormal as long as there is one feature sequence meet equation (8). Then we analyze the massive users' actual traffic behaviors with the method in sub-section 4.2.

$$V_{newi} \geq m \times V_{\max} \tag{8}$$

4.3. Experimental analysis

Traditional method used the features of the massive users traffic's byte number or packets number [35] to build the time sequence model. However, this method can only detect the abnormal behaviors that have a huge impact on network traffic. We compare the performance between our method and the method proposed by reference [35] in the experiment.

4.3.1. Experimental dataset

Normal dataset used in the experiment is the flow trace data collected by the WIDE project team in the backbone network link [29]. Abnormal datasets are the four abnormal datasets in section 4. Samples of normal traffic are 10 minutes of backbone flow of the WIDE dataset. Samples of abnormal traffic are 5 minutes of ICMP DDOS, Witty, Conficker and LLDOS1.0 traffic datasets. The test sample sets are the four attack traffic and normal traffic mixed by 1:9 separately.

4.3.2. Experimental process

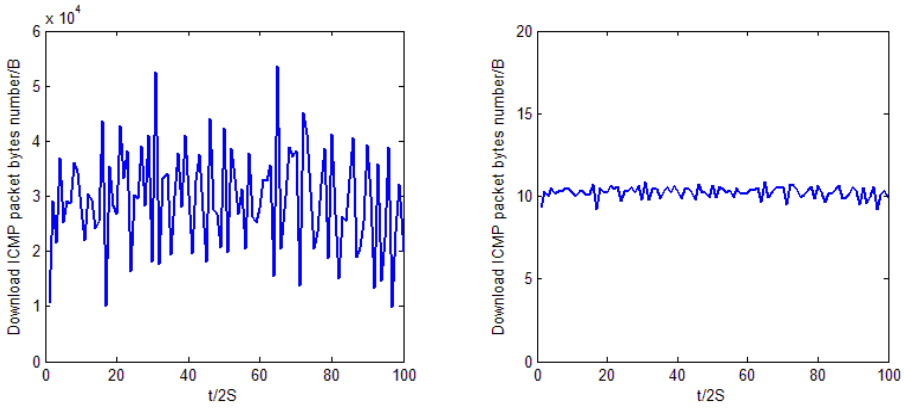
We set up the time interval as 2 seconds and the observation sequence length as 100. Finally, we set up the predicted interval length l = 10 and l = 20 in this experiment separately. The process of the experiment is as follows:

1. Smooth the time sequence of traffic behaviors.
2. Build the ARMA model for the time sequence of traffic behaviors.
3. Obtain the abnormal threshold.
4. Using our method and method proposed in reference [35] to prejudge the traffic behaviors. Experimental results are given in 4.3.3.

4.3.3. Experimental results and analysis

(1) Smooth processing

Figure 4 gives the smooth process results of feature 14 (Bytes number of download ICMP packet). From the original time sequence we can find that the sequence has a larger variance and not smooth, so we use the method proposed in sub-section 4.1 to smooth the sequence.



a. Original time sequence b. Time sequence with smooth processing

Figure 4. Smooth processing result

(2) Sequence modeling and model error analysis

We built ARMA models on the features selected in section 3.2. Table 5 gives the optimal model structure and error.

Table 5. Optimal model structure and model error

feature id	Model Structure	Model parameters		Model error	
		φ_i	θ_j	10-step prediction	20-step prediction
5	ARMA(4,3)	0.9027,-0.6089 -0.1799,-0.1306	-1.6135,1.4014 -0.4153	5.763%	11.723%
13	ARMA(3,2)	0.7944,-0.3467 -0.4954	-1.2507,1.0000	3.294%	8.679%
14	ARMA(4,3)	1.1052,-0.3901 -0.1490 ,-0.1348	-2.2045,1.9361 -0.5754	3.287%	9.158%
17	ARMA(5,4)	0.4925,0.6327 0.7242,0.6233 0.3156	0.4844,0.4962 0.4006,0.4034	4.135%	10.679%
20	ARMA(4,3)	0.1365,0.3075 0.3506,0.1657	0.1191,0.1701 0.1271	2.976%	7.062%
22	ARMA(3,2)	0.6104,0.7931 0.5567	0.2101,0.2188	3.651%	6.613%
23	ARMA(2,1)	0.1240,0.1300	0.0812	6.694%	10.264%
27	ARMA(3,2)	0.1647,0.2111 0.1744	0.1037,0.0743	3.681%	8.293%
35	ARMA(4,3)	0.1514,0.1484 0.1574,0.1450	0.0547,0.0393 0.0532	5.675%	8.186%
41	ARMA(6,5)	-0.8886,-0.2142 -0.5132,0.2166	0.4925,-0.3291 0.3701,-0.6390	7.592%	10.137%
42	ARMA(2,1)	0.5613,-0.0248 0.1883,0.1595	-0.8946 0.1396	6.454%	14.057%
43	ARMA(5,4)	-0.0772,-0.2090 -0.2900,0.5211 0.1527	-0.3737,-1.298e-11 0.3737,-1.0000	7.694%	12.618%
44	ARMA(4,3)	-0.3812,-1.0573 -0.4233,-0.1657	-0.5184,0.9827 -0.5498	5.721%	9.613%
45	ARMA(6,5)	-1.0881,-0.3213 -0.4510,0.2179	0.6459,-0.4290 0.1967,-0.5310	4.915%	8.164%
46	ARMA(2,1)	0.6071,0.0558 0.2086,0.1731	-0.8825 0.1567	7.359%	16.394%
47	ARMA(5,4)	-0.5045,-0.3838 -0.0908,0.6359 0.2775	-0.0038,-0.0439 -0.0719,-0.8803	5.786%	10.169%

48	ARMA(5,4)	-1.5165,-0.4624 0.2071,-0.0181 -0.0130	0.4942,-1.0123 -0.4891,0.4658	7.237%	11.483%
49	ARMA(2,1)	0.1809,0.1600	0.1522	3.167%	8.786%
51	ARMA(2,1)	0.1033,0.1162	0.0626	0.694%	0.907%
52	ARMA(2,1)	0.1720,0.1494	0.1432	4.497%	9.357%
61	ARMA(4,3)	0.1478,0.1577 0.1662	0.1154,0.1143	3.916%	7.281%
64	ARMA(3,2)	-1.8134,-0.9963 -0.0215	1.7964,1.0000	7.579%	10.823%
76	ARMA(4,3)	1.2008,1.0665 0.1551,0.1507	1.2122,2.1371 1.0103	6.837%	9.682%
79	ARMA(2,1)	0.1210,0.1031	-0.8759	8.339%	13.845%
80	ARMA(2,1)	0.1028,0.0935	0.0280	7.585%	13.103%
93	ARMA(2,1)	0.1345,0.1261	0.0700	4.137%	6.055%

As can be seen from Table 5, we can build the model in real time. Our method has a maximum model error of 16.394% in 20 step prediction for the feature 46, and 8.339% in the 10-step prediction for the feature 79. So we set up the prediction interval length is 10 and V_{max} = 8.339%. The reference [35] built the ARMA model based on feature 5 and feature 13. The maximum model error was11.723% with 20 step prediction and 3.294% with 10 step prediction. To contrast with our study, we set up the prediction interval length is 10 and V_{max} = 3.294% for the method in reference [35].

(3) The analysis of prejudgment

We set the abnormal threshold to 2 and 3 times of the maximum deviation respectively. Then we give the results for the prejudgment of single abnormal and multiple anomalies below respectively.

(1) Prejudgment analysis for an individual abnormal

The results of prejudgment for the analysis of individual abnormal traffic behaviors are shown in Table 6. As can be seen from the results of Table 6, the system can has a 100% accuracy for the prejudgment of the future abnormal traffic behaviors when m=2, and has a low false alarm rate. The time consuming is low because of the ARMA model is a linear model.

Table 6. Single abnormal prediction results

Type of attack	Value of m	TPR (%)	FNR (%)	Time consumption(ms)
ICMP DDOS	2	100	10	1.379
	3	100	0	1.656
Conficker	2	100	0	1.640
	3	100	0	1.285
TCP DDOS	2	100	0	1.318
	3	90	0	1.245
Witty	2	100	0	1.962
	3	100	0	1.574

(2) Prejudgment analysis for multiple anomalies

Prejudgment results for the analysis of multiple abnormal traffic behaviors are shown in Table 7. We can find that the proposed method has a 100% accuracy for prejudgment of the future

abnormal traffic behaviors when m=2, and has a low false alarm rate. However, reference [35] has a lower accuracy and a higher false alarm rate. This is because reference [35] use the upload or download byte number to detect abnormal traffic. Thus, it can't detect abnormal behaviors that have less impact on the network traffic. In the actual environment, abnormal traffic has a small percent in the overall traffic. Because the abnormal test samples are the attack traffic and normal traffic mixed in the packet number ratio 1:9, so the abnormal traffic only has a small percent in the test sample, which will lead to a low detection rate by the method in reference [35].

Table 7. A variety of anomalies prediction results

Type of attack	Value of m	TPR (%)		FNR (%)	
		The proposed method	Method of reference [35]	The proposed method	Method of reference [35]
ICMP DDOS	2	100	80	0	30
	3	90	70	0	20
Conficker	2	100	80	0	40
	3	100	60	0	20
TCP DDOS	2	100	70	0	30
	3	90	60	0	10
Witty	2	100	90	0	20
	3	100	70	0	10

The features used in this experiment are the union of the four abnormal features subset. Because the ARMA model is a linear model, so the prejudgment analysis can still keep a high efficiency.

If we compare the time consumption of the traffic monitoring and traffic prediction method, we can see that the former is larger than the latter, thus, we first prejudge the future behaviors, only when the future behaviors are abnormal, then we analyze the actual behaviors. We believe that this combined method is more efficiency than merely traffic monitoring.

4. CONCLUSIONS AND FUTURE WORK

As traffic analysis and abnormal traffic behaviors detection become an efficient way of detecting network attacks in recent years, applying such analysis to mobile network is a necessity. We identified the problems of existing traffic analysis method, and developed a feature set to describe traffic behaviors, and then we defined a feature selection rule to reduce the dimensions of features. Because the time delay caused by traffic behavior analysis, we introduced a method to prejudge the future behaviors. We applied the method in mobile network and mobile cloud. Experimental results showed that the method has a high detection rate and high efficiency in the analysis of abnormal user's traffic behaviors within massive traffic behaviors.

For future research, we will utilize the proposed method to detection more types of abnormal traffic behaviors, we believe that we can detect more abnormal behaviors in the future with a more comprehensive traffic feature set.

ACKNOWLEDGEMENT

This research was partly supported by the National Natural Science Foundation of China (61001178), Scientific Research Common Program of Beijing Municipal Commission of Education (KM201210858003), Funding Project for Academic Human Resources Development in Institutions of Higher Learning under the Jurisdiction of Beijing Municipality (PHR201108016).

REFERENCES

[1] http://en.wikipedia.org/wiki/Mobile_cloud_computing

[2] F. Sardis, G. Mapp, J. Loo, M. Aiash, A. Vinel, On the Investigation of Cloud-Based Mobile Media Environments with Service-Populating and QoS-Aware Mechanisms, IEEE Transactions on Multimedia 15(4)(2013) 769-777.

[3] J.H.Christensen, Using Restful web-services and cloud computing to create next generation mobile applications, Proceedings of the 24th ACM SIGPLAN conference companion on Object oriented programming systems languages and applications (OOPSLA), 2009,Orlando,USA.pp.627-634.

[4] Lookout Mobile Security, “Lookout Mobile Threat Report”, August 2011.<https://www.lookout.com/static/ee_images/lookout-mobile-threat-report-2011.pdf>.

[5] X. Zhang, J. Schiffman, S. Gibbs, A. Kunjithapatham, and S. Jeong, Securing Elastic Applications on Mobile Devices, Proceedings of the 2009 ACM workshop on Cloud computing security, 2009, Chicago, USA. pp. 127-134.

[6] D. Popa, M. Cremene, M. Borda, K. Boudaoud, A security framework for mobile cloud applications, Roedunet International Conference (RoEduNet), 2013, Sinaia, Romania. pp. 1-4,

[7] S. Roschke, C. Feng, C. Meinel, Intrusion Detection in the Cloud, Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing, DASC '09, 2009, Chendu, China. pp. 729-734.

[8] K. Vieira, A.Schulter, C.B. Westphall, C.M. Westphall, Intrusion Detection Techniques in Grid and Cloud Computing Environment, IT Professional, 12(2010) 38-43.

[9] A.V. Dastjerdi, K.A. Bakar, S.G.H. Tabatabaei, Distributed Intrusion Detection in Clouds Using Mobile Agents". Third International Conference on Advanced Engineering Computing and Applications in Sciences, ADVCOMP '09. 2009, Sliema, Malta.pp.175-180.

[10] S. Farraposo, P. Owezarski, E. Monteiro, A Multi-Scale Tomographic Algorithm for Detecting and Classifying Traffic Anomalies, Proceedings of IEEE ICC’07, 2007,Glasgow, Scotland. pp.24-28.

[11] N.Cheng, G. Ni, J. Luo, Z. Pan, Method of detecting the satellite communication network abnormality based on the normal behavior of clustering, Journal of PLA University of Science and Technology, 9(2008) 497-501.

[12] W. He, Research on Backbone Network Traffic Abnormal Behavior Awareness, Ph.D. Thesis, University of Electronic and Technology of China, 2011.

[13] A. Lakhina, M. Crovella, C. Diot, Mining anomalies using traffic feature distributions, Proceedings of SIGCOMM, 2005, Philadelphia, USA. pp. 217-228.

[14] A. Lakhina, M. Crovella, C. Diot, Diagnosing network-wide traffic anomalies, Proceedings of SIGCOMM, 2004, Portland, USA. pp.219-230.

[15] H. Ringberg, A. Soule, J. Rexford, Sensitivity of PCA for traffic anomaly detection, Proceedings of ACM SIGMETRICS’07, 2007, Kyoto, Japan. pp.109-120.

[16] W.E. Leland, M.S. Taqqu, W. Willinger, D.V. Wilson, On the self-similar nature of ethernet traffic, SIGCOMM '93 Conference proceedings on Communications architectures, protocols and applications, 1993, San Francisco, USA.pp.183-193.

[17] R.C. Garcia, M.N.O. Sadiku, J.D. Cannady, WAID: wavelet analysis intrusion detection, circuits and systems, Proceedings of the 2002 45th Midwest Symposium, 2002, Tulsa, Oklahoma. pp. 688-691.

[18] P. Barford, J. Kline, D. Plonka, A signal analysis of network traffic anomalies,Proceedings of IMW, 2002, Marseille, France.pp.71-82.

[19] V. Paxson, S. Floyd, Wide-area traffic: the failure of poisson modeling, IEEE/ACM Transactions on Networking, 1(3) (1995) 226-244.

[20] X. Yang, S. Yang, J. Li, A Flooding-Based DDoS Detection Algorithm Based on Non-Linear Preprocessing Network Traffic Predicted Method, Chinese journal of computers, 34(2) (2011) 395-405.

[21] D. Park, J. Yu,J. Park, MS. Kim, A comprehensive network traffic analysis model based on multidimensional OLAP data cube, International Journal of Network Management, 23(2013) 101-118.

[22] C.X. Mavromoustakis, H.D. Karatza, Real-time performance evaluation of asynchronous time division traffic-aware and delay-tolerant scheme in ad hoc sensor networks, International Journal of Communication Systems, 23(2) (2010) 167-186.

[23] C.X. Mavromoustakis, H.D. Karatza, Adaptive traffic-based control method for energy conservation in wireless devices, Simulation Modelling Practice and Theory, 13(3) (2005) 213-232.

[24] Z. Du, W. Fan, Y. Chai, Y. Chen, Priori information and sliding window based prediction algorithm for energy-efficient storage systems in cloud, Simulation Modelling Practice and Theory. 39(2013) 3-19.

[25] D.E. Denning, An intrusion-detection model, IEEE Transactions on Software Engineering, SE-13(1987) 222-232.

[26] T. Auld, A.W.Moore, S.F.Gull, Bayesian neural networks for internet traffic classification, IEEE Transactions on Neural Networks, 18(1) (2007) 223-239.

[27] H. Zeng, Machine Learning, China Machine Press, China, 2003.

[28] Wireless Topology Discovery, wtd_data_release.<http://sysnet.ucsd.edu/wtd/>.

[29] MAWI Working Group. Traffic traces. <http://mawi.wide.ad.jp/mawi/>.

[30] The University of Waikato, Traffic traces. <http://wand.net.nz/wits/waikato/1/20040507-233830-64.php>.

[31] The University of Auckland, Traffic traces. <http://wand.net.nz/wits/auck/8/20031202-090000.php>.

[32] CAIDA. CAIDA Data.<http://www.caida.org/data/>.

[33] RIPE NETWORK COORDINATION CENTRE.Traffic traces. <https://labs.ripe.net/datarepository/data-sets/nlanr-pma-data>.

[34] MIT Lincoln Laboratory. DARPA Intrusion Detection DataSets.<http://www.ll.mit.edu/mission/communications/cyber/CSTcorpora/ideval/data/2000data.html>.

[35] B. Zou. Detection and prediction of network traffic anomaly, Ph.D. Thesis, Graduate School of Chinese Academy of Sciences, 2003.