ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645
Indian Journal of Science and Technology, *Vol 9(48), DOI: 10.17485/ijst/2016/v9i48/107986, December 2016*

Detection and Prediction of Abnormal Users in Cloud Network

G. Vinodhini* and V. Meena

School of Computing, SASTRA University, Tirumalaisamudram, Thanjavur – 613401, Tamilnadu, India;
viniyaeee@gmail.com, menna@cse.sastra.edu

Abstract

Objectives: Cloud computing refers to an internet based computing that allows sharing data and resources, which provides create, configure and customize the applications online. **Methods**: To overcome this issue, we proposed an analyzing behavior of user's traffic. First, construct a whole feature set by collecting features from users and feature selection is helps to choose set of accurate features where necessary features are selected and predicted for an abnormal user is done by Naive Bayes classification. **Findings**: Support Vector Machine (SVM) produces high accuracy and over fitting of data is applicable. This new method improves the efficiency and detection rate of the system in the analysis of traffic behavior of abnormal users. **Applications**: Email communication, banking, e-commerce and social networks.

Keywords: Data Mining Techniques, Feature Set, Feature Selection, Network Traffic Analysis, Network Traffic Prediction, SVM

1. Introduction

Banking, e-commerce and business corresponding are among the highly privileged activities and information are communicated within the network are considered important and there is a need to maintain the significant of network traffic analysis to ensure perfect information security. Analysis and prediction of network traffic in cloud is a proactive approach, where the network is monitored to protect that security violation which doesn't occur within the network. In[1] Traffic anomalies are the result that changes the normal flow of data in a network. Some occurrences may be triggered are Denial-of-Service (DoS), Distributed Denial-of-Service (DDoS) attacks or flash crowds. Though the traffic anomalies occur at any case of internet, the unpredictable behaviors may vary between single attacks of network failure to a complex attack of security. Now a days, Cloud computing plays a major role in vast areas to upload and download files/data and network traffic will occur from massive users. Cloud computing is a movement of computing from local platforms and physical hardware to virtualized services which hosted in cloud[2]. Cloud network is classified into Public cloud, Private cloud and the Hybrid cloud. Public cloud does not specify to particular organizations, can be access from anywhere over the internet as Microsoft Azure, Amazon web services, Salesforce.com and Google App Engine. Private cloud is specially designed for particular/specific organizations which have susceptible information[3]. The Eucalyptus and Open stack are the software developed for managing private cloud. Hybrid cloud is a combination of public cloud and private cloud systems which afford both the services. Cloud computing[4] involves three level of services are Infrastructure as a Service (IaaS), Software as a Service (SaaS), Platform as a Service (PaaS). These three services provides reliable, con-

*Author for correspondence

venient and safe. To monitor network traffic, the analysis of network traffic plays an important role. In the earlier period, a small amount of the network devices or a less than a thousand computers is monitored by administrators. In the bandwidth of network is less than 100 Mbps (Megabits per second). Presently, administrators have to treat with high speed network that is more than 1 Gbps (Gigabits per second). To manage the network and solve network problems fastly and to overcome the network failure and network security, some additional current tools for analysis of network traffic is required. In present days, analysis of network traffic presents an amount of challenges. Network is monitored and analyzed at various levels are network level, packet level and flow level for the security management. Researchers used different methods for analysis of network traffic. A standard structure for analysis of network traffic involves pre-processing which is followed by the actual analysis and some observations to expose patterns from network data. In[5] the analysis of network traffic involves three phases are Pre-processing, Analysis and Evaluation. Pre-processing technique is a major phase which manipulates real data that is often incomplete and noisy in particular behavior. In other prose, analyze the data by using techniques of data mining which are inconsistent and incomplete. Therefore, the methods of pre-processing were required to increase the data quality for boosting the efficiency and accuracy of upcoming data. Data mining techniques is used to classify abnormal user and normal user in analyzing traffic behaviors of massive user's. Data mining performs a major role in analyzing the network traffic. Data mining is a concept of extracting information from wide amount of data where large data are stored and may apply some intelligence methods to support the process of decision making. The classification methods in data mining are Naive Bayesian classification, Logistic regression, Decision tree, SVM. All these techniques are used to classify traffic behavior of abnormal user and normal user in the network. In this work, focused on detection of network traffic that is identifying abnormal user in the entire network or identifying the user who creating abnormal moment. To monitor network traffic in cloud, two major techniques are analyzed, by User's traffic behavior and Pattern analysis. Some challenges occurs in analyzing massive user's traffic behaviors are (a) it is complex to explain behaviors of user's traffic due to the difficulty of cloud traffic, (b) high time consumption for analyzing traffic behaviors of massive users, (c) to distinguish and recognize traffic user's behaviors of individual network from a user's data of massive network. Some techniques and efforts are developed to overcome these challenges in collection of pattern analysis of traffic network, it remnants dispute to analyze the traffic behaviors of massive user's in terms of efficiency and accuracy.

In[6] the Adaptive Energy Conversation (ADEC) protocol, the lifetime of the network is improved by saving power of each node. An ADEC protocol thoroughly examines the traffic pattern and assigns sleeping time for each node. The nodes in the network accept the sleep duration based on the similarity of traffic. This method increases sleep time duration and provide continuous connectivity to the network, which results in efficient conservation of energy. In[7] for detecting traffic anomalies, they proposed a method of anomaly independent which is combined with an analysis of multi-criteria tomographic. The criteria includes bytes, flow rate, address distribution. The algorithm consists of three phases which is used to detect and identify the traffic anomalies. The algorithm has two applications, where the first application enhances with improvement of traffic engineering and the second application deals with an Intrusion Prevention System (IPS). In[8] described about Three State Disk Models (3DSM) which is mainly concentrated on improving the quality, consumption of storage system. This method is based on reducing the workload without losing the quality in order to conserve the energy. The efficiency is mainly based on predicting the information of the blocks which can be accessed and also the blocks which are not able to be accessed. High amount of energy can be saved in the prediction method which is used for storage systems. In[9] the concept is basis on the flow data which displays the current network status and by changing the data flow, the anomalies are found on network. The concept of Deep Packet Inspection (DPI)[10] provides to check anomalies using security rules. The DPI is used to change the matching of target feature which is based on the finite state machine beside with the explicit string. In this proposed system[11], a high accuracy traffic classifier is used in a range

of applications without the information of host address or port for source and destination. Bayesian Neural Network is used as a supervised learning for this issue. By providing the classification without accessing the packets, this method enables greater application than the concepts which require all packets or classification payloads. It's a dominant advantage, by means of the samples to permit the traffic categorization based on the available information. In[12] the challenge for models of network deals with concept drifting that depends on the statistical heuristics educated from data stream, for example, Anomaly Based Prevention or Detection Systems. In this work, to quantify the drift of concept in the network traffic uses adaptive learning strategies with the window of fixed training to evolve the method constantly. The classifier of **Naive Bayes** is used for data classification. ROC curve is used for identifying the concept drift which is generated from Naive Bayes Classifier.

2. System Model

The user is analyzed based on monitoring the network and the architecture is shown in Figure 1. The proposed system consists of three steps are: (1) Construct a full feature set which is based on the network traffic characteristics to illustrate the behavior of user's traffic. (2) Based

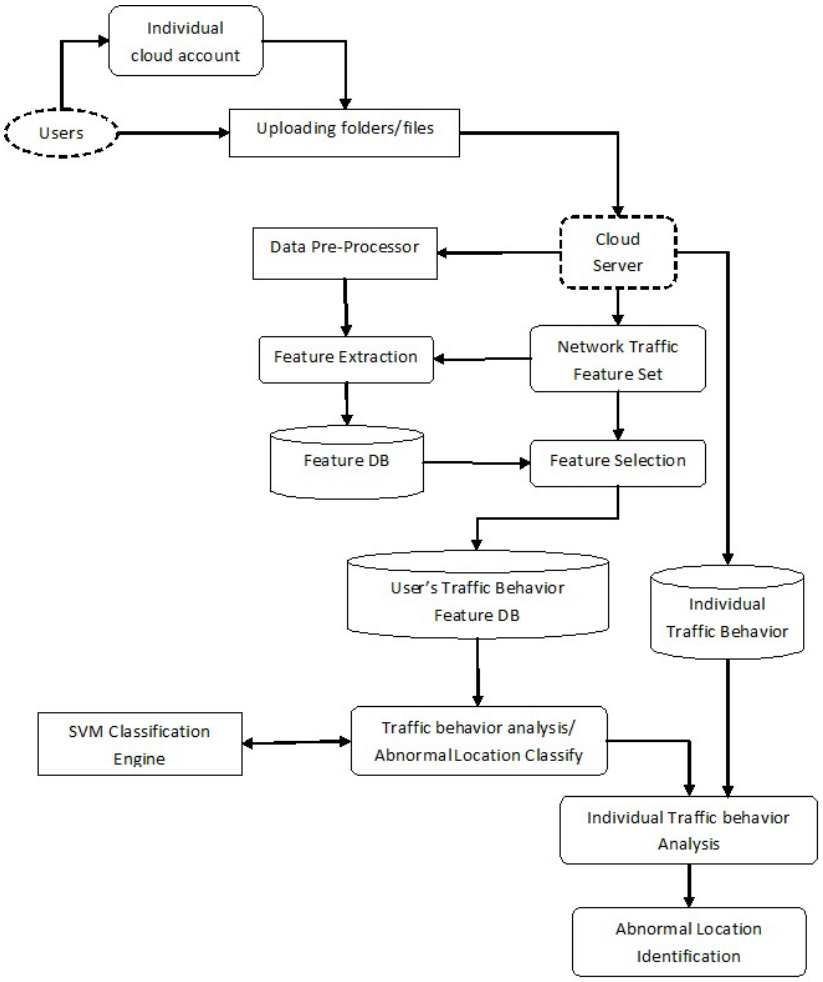

Figure 1. Architecture of user's traffic behavior.

on relative deviate distance, best feature sets are selected using feature selection rule. (3) Finally, an analysis of traffic behavior system depends on the predictive method is implemented to increase the efficiency of system and accuracy[1]. The experimental outcome results in high rate of detection and high efficiency while comparing to the existing systems.

2.1 Cloud Storage Module

New users register to the Microsoft Azure cloud to access the cloud network for uploading/downloading folders, files and service access. The Microsoft Azure storage emulator offers a local platform which emulates the Azure storage as Blob, Queue, and Table storage for the development purposes. This emulator is used to test the application locally without creating the Azure subscription, or incurring any other costs. Azure users login to the Microsoft cloud and upload their own data's/files/folders to their individual cloud account. Azure Blob storage can store any type of text or binary data as document, media files, or application installer.

2.2 Data Access Module

New files or folders will automatically sink to the cloud storage. A new folder with some data is created in system and browses the same folder for uploading in to the cloud storage. Now, additional files or data can be added to the folder in the system and there is no need to browse the new data to upload, which is uploaded automatically in to the cloud storage. The easiness -of access and time required is depends on type and format of file/data that a user upload and service levels for some specific storage type.

2.3 Feature Selection Module

Feature set is provided by the analysis of the IP packets are upload time, source ip, source port, destination ip, destination port, packet length, number of byte, upload type, speed, protocol, direction. The feature behavior of network traffic consists of static features and the dynamic features. Static features as port, packet length, special packet and IP address and the dynamic features as connection, speed, integration and distribution. The rule for selecting the feature is depends on set of features which is needed to improve the efficiency. The tabular column shows the collected features from the massive users with user name, file name, file size, upload time, speed. Feature extraction includes entropy and gain calculation for the features file size, upload time and speed. The entropy is the average amount of information which is needs to classify the abnormal user and normal user.

Table 1. Collection of feature set from the users

File name	User name	Source Ip	Destin. Ip	Upload time	File size	Packet length	Speed
Img1.Jpg	Xxx	127.0.0.1	127.0.0.1	00:00:01.0483	40002	400	1.22
Vid. AVI	Xxx	127.0.0.1	127.0.0.1	00:00:00.6272	18167867	181	554.44
Img2.Jpg	Yyy	192.168.0.18	127.0.0.1	00:00:00.1040	40699	406	3.05
Doc.pdf	Zzz	172.22.44.103	127.0.0.1	00:00:00.2610	99904	999	13.95

$$I=-\sum_{c} p(c)\log_2 p(c) \quad (1)$$

By using this formula, entropy is calculated for features to classify the massive user's data of normal user and abnormal user. The features upload time, file size, and speed is calculated. The Table 1 represents the collection of feature set from the users are tabulated by the following features as user name, file name, source ip, destination ip, upload time, packet length, file size and speed. The file types are jpg, pdf, doc, etc., can be uploaded. The source ip and destination ip are sender port and receiver port. The packet length represents the files are transfer in packets and the file size is measured in bytes.

2.4 Classification and Prediction

A classification process is used in the analysis of network traffic is to classify all user's traffics either as normal or abnormal. The effort of classification technique is to reduce the amount of false positives as detection of normal traffic behavior as abnormal and false negatives as detection of malicious traffic behavior as abnormal. The optimized features are selected from user's specifications fields for classifying abnormal users and normal users.

2.4.1 Support Vector Machine (SVM)

SVM is a superintended learning machine which is used for classification and regression. SVM is designed to test huge amount of high dimensional data or n – dimensional data. SVM performs better experimental results than Naive Bayes classification. The SVM is used to identify hyper plane which is best between the two classes of data, the positive and negative values are separated through solid line in the middle which is called as decision line. The SVM applications are machine vision, analysis of time series, text categorization, bioinformatics and handwritten character recognition.

2.4.1 Kernels in SVM

SVM uses different kernels for machine creation are Gaussian kernel, Polynomial kernel, Laplacian kernel, Sigmoid kernel. Gaussian kernel is an universal kernel which reduces the estimation error and approximation errors. The kernel methods are used for pattern analysis with a class of algorithms. The different types of relations are to be found and studied is the main work of pattern analysis as classifications, clusters, principal components, rankings and correlations in data sets. The properties of SVMs are kernels, duality, sparseness, convexity and margin.

3. Result and Discussion

The Figure 2 shows the performance measure for SVM which includes sensitivity, specificity, efficiency and accuracy. Comparing to Naïve Bayes classification, the SVM shows better accuracy and the efficiency for network

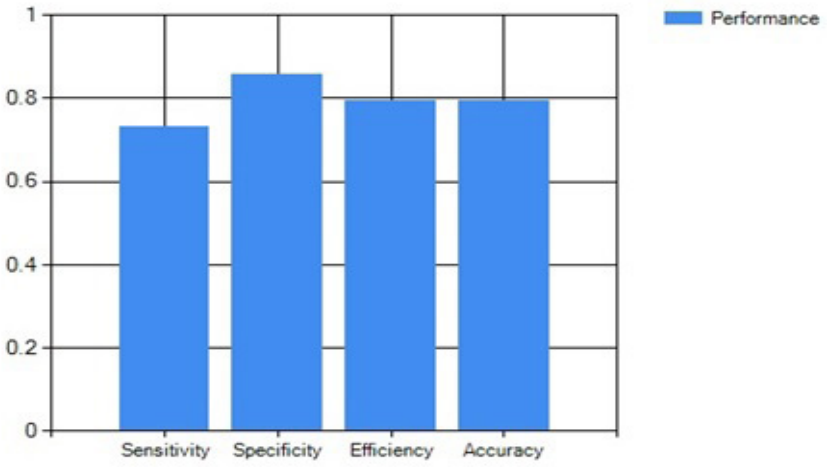

Figure 2. Performance measure for SVM.

Ab-Normal Data Detection Result Area

	SenderPort	SenderIP	FileName	Protocol	FileSize
▶	5495	127.0.0.1	Complete BMW i3 Produ...	IPv4	414
	1260	172.22.44.103	Neeya Naana Boys Vs Gi...	IPv4	714
	1260	172.22.44.103	YouTube - NN 2.flv	IPv4	474
	1260	172.22.44.103	YouTube - NN 5.flv	IPv4	565
	1260	172.22.44.103	YouTube - NN 7.flv	IPv4	458
	1260	172.22.44.103	YouTube - NN 8.flv	IPv4	631

Figure 3. Results of abnormal user location.

users. The Figure 3 shows the detection of abnormal data and lists the abnormal data with sender IP, sender port, file name, file size and protocol. By using the sender IP and sender port, can prejudge the traffic behavior of massive user's that is prediction of abnormal data is performed. The detection and prediction of analysis of the massive behavior of user's traffic is identified and abnormal data with abnormal user is detected.

4. Conclusion

An analysis of traffic and abnormality of traffic detection plays as competent method for detecting various attacks in cloud network. The SVM classification is used to classify abnormal data and normal data, and predict abnormal location by using the specifications of users. The SVM provides high accuracy and over fitting of data than Naive Bayes classifier. The experimental results and performance chart shows increased rate of detection and increased efficiency for the examination behavior of users in the enormous data of traffic behaviors. For upcoming work, exploit this proposed system to detect additional types of abnormal traffic behaviors and use Multinomial Naive Bayes (MNB) algorithm for classifying the abnormal users and normal users.

5. References

1. Lai Y, Chen Y, Liu Z, Yang Z, Li X. On monitoring and predicting mobile network traffic abnormality. Simulation Modelling and Practice Theory. 2015 Jan; 50:176–88.
2. Rutka G. Some aspects of traffic analysis used for internet traffic prediction. Research Journal of Electronics and Electrical Engineering. 2009; 5(93):1–4.
3. On the investigation of cloud-based mobile media environments with service-populating and Qos-aware mechanisms [Internet]. 2012 [cited 2012 Nov 13]. Available from: http://www.mdx.ac.uk/__data/assets/pdf_file/0028/49384/IEEE-TM-camera-ready.pdf.
4. Kirubakaramoorthi R, Arivazhagan D, Helen D. Analysis of cloud computing technology. Indian Journal of Science and Technology. 2015 Sep; 8(21):1–3.
5. Zhani MF, Elbiaze H. Analysis and prediction of real network traffic. Journal of Networks. 2009 Nov; 4(9):855–65.
6. Du Z, Fan W, Chai Y, Chen Y. Priori information and sliding window based prediction algorithm for energy-efficient storage systems in cloud. Simulation Modelling Practice and Theory. 2013 Dec; 39:3–19.
7. Mavromoustakis CX, Karatza HD. Adaptive traffic-based control method for energy conservation in wireless devices. Simulation Modelling Practice and Theory. 2005 Apr; 13(3):213–32.
8. Farraposo S, Owezarski P, Montreiro E. A multi-scale tomographic algorithm for detecting and classifying traf-

fic anomalies. Proceedings of Institute of Electrical and Electronics Engineers (IEEE) International Conference on Communications; 2007 Jul. p. 1–5.
9. Zhang L, Wang J, Lin S. Design of the network traffic anomaly detection system in cloud computing environment. In the proceedings of Fourth International Symposium on Information Science and Engineering (ISISE), Institute of Electrical and Electronics Engineers (IEEE) Xplore Digital Library; 2012 Dec 14. p. 16–9.
10. Prasad SNSE, Srinath MV, Basha MS. Intrusion detection systems, tools and techniques – an overview. Indian Journal of Science and Technology. 2015 Dec; 8(35):1–7.
11. Singh MP. Quantifying concept drifting in network traffic using ROC curves from naive bayes classifier. In the Proceedings of Nirma University International Conference on Engineering (NUiCONE), Institute of Electrical and Electronics Engineers (IEEE) Xplore Digital Library; 2013 Nov 28. p. 1–5.
12. Auld T, Moore AW, Gull SF. Bayesian neural networks for internet traffic classification. Institute of Electrical and Electronics Engineers (IEEE) Transactions on neural networks. 2007 Jan; 18(1):223–39.