DOI: http://dx.doi.org/10.26483/ijarcs.v9i4.6279

ISSN No. 0976-5697

Volume 9, No. 4, July – August 2018

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

INTRUSION DETECTION SYSTEM USING CLASSIFICATION APPROACH IN DATA MINING

Parminder Singh
Department of Computer Applications
National Institute of Technology, Jamshedpur
Jamshedpur, India

Chandrashekhar Azad
Department of Computer Applications
Jamshedpur, India
National Institute of Technology, Jamshedpur

Ashok Kumar Mehta
Department of Computer Applications
National Institute of Technology, Jamshedpur
Jamshedpur, India

Abstract: Data on a network comprises of interrelated units. As an illustration, web pages on the World Wide Web are connected by hyperlinks and research papers are associated by references also phone accounts are linked by calls, and conceivable terrorists are linked by communications. Networks have turned out to be pervasive. Correspondence networks, financial transaction networks, networks portraying physical systems, and social networks are all ending up noticeably progressively important in our everyday life. Regularly, we are interested in models of how nodes in the system influence each other (for example, who taints whom in an epidemiological system), models for predicting an attribute of intrigue in light of observed attributes of objects in the system. The technique of SVM is applied which will classify the data into malicious and non-malicious.

Keywords:Data mining, classification, VoIP, SVM, KNN, Security, IDS

I. INTRODUCTION

The connection of more than one computer systems that provide benefits to each other is considered as a network. The computers connected to communicate and provide exchange of information to each other. The collections of computer devices that facilitate communication amongst each other are gathered here within this setup. The scenario in which numerous computers are gathered and connected with each other to exchange information and provide facilities to other resources is called a network [1]. The information such as data communication is provided with the help of networking technology. There are software and hardware types of resources present within the sharing devices. VoIP stands for Voice over Internet Protocol that uses internet or other data network rather than using conventional Public Switched Telephone Network (PSTN). A rapid growth has been seen in use of internet for voice communications that results in reduce cost of equipment, operation and maintenance [2]. The VoIP is a solid technology that allows people to communicate through voice using IP protocol instead of telephone lines. The property standards, high price tag, limited integration with existing telephony environments are some of the factors that have assigned this technology in a niche market. Now a day's situation has been changed due to advent of asterisk as well as low-cost VoIP telephone adapters open source tools [3]. This has become easy and common for internet providers to provide their customers VoIP calls at very low cost, if any in addition to standard xDSL connectivity.The goal for developers is relatively simple: add telephone calling capabilities (both voice transfer and signaling) to IP-based networks and interconnect these to the public telephone network and to private voice networks in such as way as to maintain current voice quality standards and preserve the features everyone expects from the telephone. Data Analysis can be defined as the process of reviewing and evaluating the data that is gathered from different sources. Data cleaning is very important as this will help in eliminating the redundant information and reaching to the accurate conclusions. Data analysis is the systematic process of cleaning, inspecting and transforming data with the help of various tools and techniques [4]. The objective of data analysis is to identify the useful information which will support the decision-making process. There are various methods for data analysis which includes data mining, data visualization and Business Intelligence. Analysis of data will help in summarizing the results through examination and interpretation of the useful information. Data analysis helps in determining the quality of data and developing the answers to the questions which are of use to the researcher [5].Various attacks or threats in VoIP have given in this section and their impact on the overall network security. Denial of Service (DoS) attack id the attack in which attacker's main target is to make resources unavailable to users. In this process, attackers full the server with so many fake requests so that it cannot process genuine requests. For example, observe that a server can take 100 users at a time. Attacker sends 100 fake messages to the server continuously

due to which server is filled up or exhausted in processing and replying to these messages. Therefore, this attack prevents the legal user to take services from the server as resources are fully used by the attack [6]. Man in the Middle Attackis the attack in which attacker listens the private data between the two parties. The attacker sit in the middle of the two communicating vehicle and launch this attack. In this attacker control all the communication between the sender and the receiver but communicating vehicles assume they are directly communicating with each other. In this attack, attacker listen the communication between the vehicles and inject false or modified message between the vehicles. Registration Hijacking is an attack in which an attacker registers himself as one of the already existing legal users. Attacker too received the call when a call is forwarded to the legal user. Spam over Internet Telephony (SPIT) is the attack in which spam calls are transferred by an attacker to users connected to the internet [7]. There are a number of protocols that are employed in order to provide for VoIP communication services. They can be implemented in using both the proprietary and open protocols and standards. In order to be able to communicate using a VoIP system, there are two types of protocol that must be used.Literature survey is done where different and various approaches along with the unique researches that have been done by numerous writers are analyzed and written the findings of the literature.Research method and classification using SVM and KNN is discussed along with the flowchart of the work.

Results shown that how accuracy and execution time varies depending upon the method of classifier and dataset proportion. A table is shown from where former parameters can be analyzed and compared.

Concluded the result which favored our method to achieve better network traffic classification.

II. LITERATURE REVIEW

Mamadou Alpha Barry, et.al (2018) presented a study related to the work that has been previously done to obtain results that clearly show the importance of QoS to ensure optimal transmission of traffic and media streams on IP/MPLS network. In short, this approach is able to install and configure the IP/MPLS network through the GNS3 simulator and install the multimedia services that were tested and emulated in a virtual lab [8]. This makes it possible to rely on a dimensioning phase and to scale up. The DiffServ architecture is used to deploy QoS. In this work, the influence of the diffserv approach is analyzed in terms of the delay and the jitter applied to the services and tested in the absence and presence of QoS..

Ahmed Fawzy Gad, et.al (2018) presented a review about spam over internet telephony attack on voice over IP networks. This paper starts by explaining why IP networks became the most dominant type of information networks and how it is better than the legacy PSTN for connecting users all over the world. Requirements for carrying voice over IP networks are discussed in terms of both devices and protocols [9]. Each approach is discussed by showing its characteristics, how it works in addition to its pros and cons. A virtual VoIP network is created to conduct an experiment to compare these presented approaches.

Mario A. Ramirez-Reyna, et.al (2017) proposed a differentiated call admission control (CAC) strategy for VoIP traffic-based wireless networks using different codecs and/or codec mode-sets and mathematically analyzed [10]. The aim of this strategy is to regulate and restrain the admission of most resource demanding VoIP sessions (those with a larger packet size requirement). A joint connection and packet level analysis is formulated to assess the performance of the proposed CAC strategy. Maximum achieved Erlang capacity for different data rate transmission requirement ratios and proportion of users using each codec and/or codec mode-set is evaluated. Numerical results show that system performance is improved with the proposed CAC strategy.

Murizah Kassim, et.al (2017) presented that wireless mobile telecommunication has evolved from the Third Generation (3G) to Fourth Generation (4G) network. This paper presents the comparison analysis on 3G and 4G of VoIP network performance [11]. Standard Quality Management Scale recommended by ITU-T P.862 is used for the measurement analysis. A test bed experiments on voice Skype application is done and data is collected. The traffic is analyzed using Jperf software to display the network performance and measurement which is tested in 30 second per session. Result presents performance of bandwidth availability, jitter performance, latency, VoIP, and current LTE4G analysis versus previous WiMAX 4G. It is identified that an average rating of 4 for both 3G and 4G LTE network. This shows performance of the VoIP is achieved. Three elements which are bandwidth, latency and jitter need to be in a good order to get a good connectivity for both connections.

Jan Holu, et.al (2018) analyzed call detail records of 16 million live calls over Internet-Protocol-based telecommunications networks. The objective is to examine the dependency between average call duration and call quality as perceived by the user. Surprisingly, the analysis suggests that the connection between quality and duration is non-monotonic [12]. This contradicts the common assumption that higher call quality leads to longer calls. In light of this new finding, the use of average call duration as indicator for (aggregated) user experience must be reconsidered. The results also impact modeling of user behavior. Based on the finding, such models must account for quality since user behavior is not fully inherent, but also depends on external factors like codec choice and network performance.

Eko Ramadhan, et.al (2017) presented that computer network technology as a medium of communication between devices has made significant progress in terms of communication media [13]. One benefit is VoIP can be used as a communication network implemented with Asterisk applications as a server to a Private Automatic Branch eXchange (PABX) in a system simulation using GNS3 emulator. In this research the routing used is BGP routing protocol to get optimal QoS value with different bandwidth. From the simulation results of testing using the bandwidth of 64 kbps, 128 kbps and 256 kbps are performed each test three times as much bandwidth as QoS values obtained on average better than the results of delay, jitter, packet loss

and throughput obtained from the VoIP network based on a standard ITU- T G.114.

III. RESEARCH METHODOLOGY

This work is based on the network traffic classification to classify the traffic into malicious, non-malicious. The network traffic analysis is the technique which is applied to predict the malicious activities of the users which are active on the network. To classify the network traffic three steps has been followed in the methodology, in the first step technique of k-mean clustering is been applied in which similar and dissimilar type of data will clustered. The dataset which is taken as input will be refined by removing redundancy and missing values. In the second step, technique of k-mean clustering is applied in which arithmetic mean of the whole dataset is calculated which will be the central point of the dataset. The Euclidian distance from the central point is calculated which define the similarity and dissimilarity of the points. The points which are similar will be clustered in one cluster and other in the second cluster. In the last step of classification technique, SVM classifier will be applied which classify the data into two classes. To improve the performance of the existing system technique of Knn classifier will be applied which will cluster the uncluttered points and increase accuracy of classification. The KNN classifier the nearest neighbor classifier in which Euclidian distance is calculated and points which have similar distance will be clustered in one class and other in the second class.

A. PROPOSED ALGORITHM

Step 1] Input dataset;

Step 2] Pre-process dataset;

Step3] Divide and Input dataset for training and testing;

Step 4] Apply the KNN;

Step 5] Print the output;

Step 6] Close;

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

The proposed algorithm is implemented in Python and the results are evaluated by making comparison with existing algorithm in terms of accuracy and execution time.

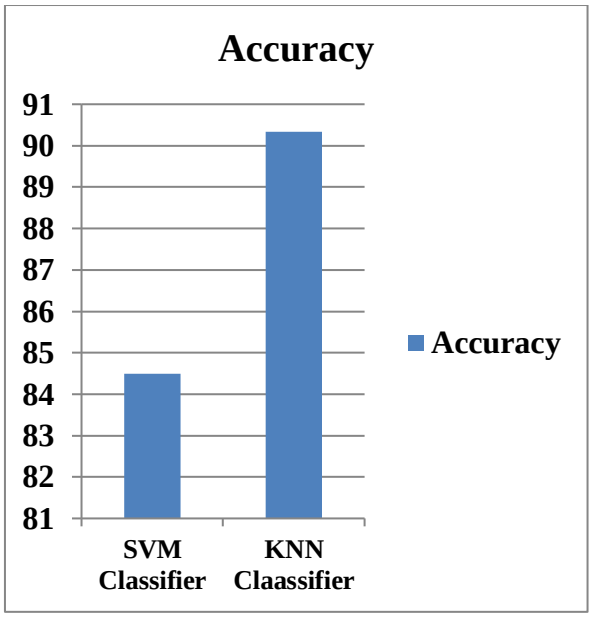

Fig 1: Accuracy Comparison

As shown in figure 2, the value of accuracy of SVM classifier is compared with the KNN classifier for the network traffic classification. It is been analyzed that accuracy of KNN classifier is high as compared to SVM classifier.

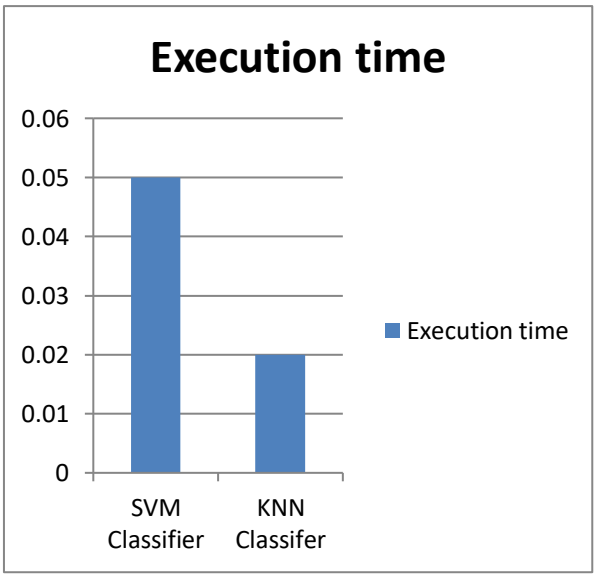

Fig 2: Execution Time

As shown in figure 3, the execution time of the proposed algorithm is compared with the existing algorithm. It is analyzed that execution time of KNN classifier is less as compared to SVM classifier.

We have discussed the results by dividing dataset into various proportions. After the execution, we analyzed the results i.e. accuracy for our approaches SVM and KNN. Below are the 4 different test cases that we have applied on the KNN and SVM to check the variations in accuracy.

Table 5.2: Variations in accuracy using different dataset proportion using KNN and SVM

Learning method	Training data (%)	Testing data (%)	Accuracy (%)
KNN	85	15	97.07
SVM	85	15	84.00
KNN	20	80	91.65
SVM	20	80	79.50
KNN	60	40	90.35
SVM	60	40	84.00

V. CONCLUSION

One major task which is vital in the machine learning is to classify the data. Because of the vast number of applications, numerous data classification systems have been developed. A portion of the well-known ones are decision trees, instance-based learning, e.g., the K-nearest neighbor's algorithm (KNN), artificial neural networks, Naive-Bayes, and support vector machines (SVM). All things considered, the greater part of them is highly dependent of appropriate parameter tuning. One illustration is that minimum number of cases that are required for partition set in C4.5 decision tree; the K value in KNN, the number of hidden layers, and others in artificial neural networks; and the soft margin, the piece function, the bit parameters, the stopping criterion, and others in SVM.

VI. REFERENCES

[1] D. Rodrigues, E. Cerqueira, and E. Monteiro, "QoE Assessment of VoIP in Next Generation Networks," MMNS 2009, LNCS 5842, International Federation for Information Processing, pp. 94-105, 2009.

[2] James Yu, Imad Al Ajarmeh, "Design and Traffic Engineering of VoIP for Enterprise and Carrier Networks", International Journal on Advances in Telecommunications, vol. 1, No. 1, 2008.

[3] O. Hersent, J.P. Petit, and D. Gurle, "Beyond VoIP Protocols. Understanding Voice Technology and Networking Techniques for IP Telephony," John Wiley & Sons Ltd, 2005.

[4] C. Olariu, J. Fitzpatrick, P. Perry, and L. Murphy, "A QoS based call admission control and resource allocation mechanism for LTE femtocell deployment," in Consumer Communications and Networking Conference (CCNC), 2012 IEEE. IEEE, 2012, pp. 884–888.

[5] M. Afaq, S. U. Rehman, and W. C. Song, "Visualization of elephant flows and qos provisioning in sdn-based networks," in Network Operations and Management Symposium (APNOMS), 2015 17th Asia-Pacific, Aug 2015, pp. 444–447.

[6] C. Xu, B. Chen, and H. Qian, "Quality of service guaranteed resource management dynamically in software defined network," Journal of Communications, vol. 10, no. 11, 2015.

[7] M. F. Bari, S. R. Chowdhury, R. Ahmed, and R. Boutaba, "Policycop: an autonomic qos policy enforcement framework for software defined networks," in Future Networks and Services (SDN4FNS), 2013 IEEE SDN for. IEEE, 2013, pp. 1–7.

[8] Mamadou Alpha Barry, James K. Tamgno, Claude Lishou, Modou Bamba Cissé, "QoS Impact on Multimedia Traffic Load (IPTV, RoIP, VoIP) in Best Effort Mode", International Conference on Advanced Communications Technology(ICACT), 2018

[9] Ahmed Fawzy Gad, "Comparison of Signaling and Media Approaches to Detect VoIP SPIT Attack", IEEE, 2018

[10] Mario A. Ramirez-Reyna, S. Lirio Castellanos-Lopez, Mario E. Rivero-Angeles, "Connection Admission Control Strategy for Wireless VoIP Networks Using Different Codecs and/or Codec Mode-sets", The 20th International Symposium on Wireless Personal Multimedia Communications (WPMC2017)

[11] Murizah Kassim, Ruhani Ab. Rahman, Mohamad Azrai A.Aziz, Azlina Idris, Mat Ikram Yusof, "Performance Analysis of VoIP over 3G and 4G LTE Network", IEEE, 2017

[12] Jan Holu, Michael Wallbaumy, Noah Smithy and Hakob Avetisyan, "Analysis of the Dependency of Call Duration on the Quality of VoIP Calls", IEEE, 2018

[13] Eko Ramadhan, Ahmad Firdausi,3Setiyo Budiyanto, "Design and Analysis QoS VoIP using Routing Border Gateway Protocol (BGP)", IEEE, 2017