Implementation of Classification Approach for the VoIP Network Traffic Classification into Malicious and Non Malicious Traffic

[1]Nidhi, [2]Gaganpreet Kaur, [3]G.N. Verma
[1]Research Scholar, Sri Sukhmani Institute of Engineering and Technology, Derabassi, India
[2]Assistant Professor, Sri Sukhmani Institute of Engineering and Technology, Derabassi, India
[3]Principal, Sri Sukhmani Institute of Engineering and Technology, Derabassi, India
([1]nidhimankoo@gmail.com, [2]gagan682@gmail.com)

Abstract-Networked data contain interconnected entities for which inferences are to be made. For example, web pages are interconnected by hyperlinks, research papers are associated by references, phone accounts are linked by calls, and conceivable terrorists are linked by communications. Networks have turned out to be ubiquitous. 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-*VoIP;SVM; KNN.*

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. The information such as data communication is provided with the help of networking technology [1]. 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 [3]. Now a day's situation has been changed due to advent of asterisk as well as low-cost VoIP telephone adapters open source tools. 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]. It is 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 [7]. 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.

II. LITERATURE REVIEW

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. There are a number of challenges for voice over IP networks and security attacks are at the top. This paper concentrates on SPIT attack and its detection methods which are signaling and media [8]. 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 [9]. 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 [10]. 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. 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 [11]. 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. Currently we can communicate through this network by utilizing technology called Voice Over Internet Protocol (VoIP). It is one of the fastest growing internet applications now [12]. 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.

Mohammad Tariq Meeran, et.al (2017) proposed two techniques that can contribute to the VoIP quality improvement in WMNs. Firstly, the best choices for integration of standards, protocols and voice codecs that can produce the best VoIP quality are proposed in the specific scenarios in consideration of the standards, protocols and codecs used in our experiments. Secondly, the inclusion of stationary and mobile supportive mesh nodes to the mesh topology is proposed [13]. It is seen that the addition of supportive mesh nodes to the topology for offloading the packets forwarding process from VoIP communicators can result in improving the VoIP quality based on 5-point MOS rating- scale by 0.2 in no mobility, 2.2 in partial mobility and 0.9 in full mobility scenarios. It can be concluded that minor gains are achieved in the no-mobility scenarios, significant and considerable gains are achieved for the full mobility and partial mobility scenarios, respectively.

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 unclutered 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. During the performance of multiclass categorization, there can be tie in case when k is an odd whole number. The classification of samples on the basis of majority class of its nearest neighbor is the major task of KNN algorithms.

$$Class = arg_v max \sum_{(X_i, y_i) \in D_z} I(v = y_i) \quad \ldots (1)$$

Here, the class label is represented by v. The class label for ith nearest neighbors is denoted by y_i.

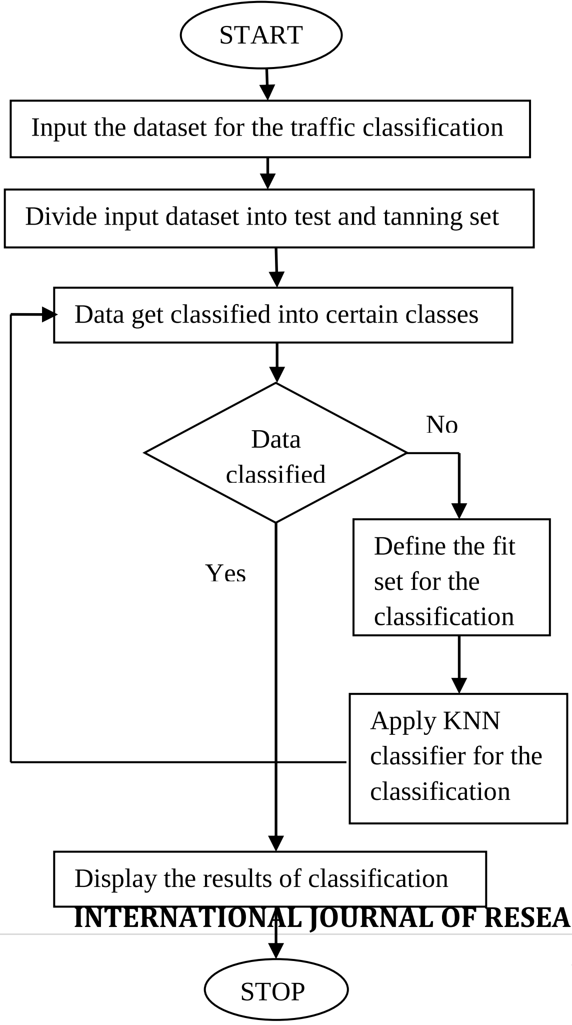
START
Input the dataset for the traffic classification
Divide input dataset into test and tanning set
Data get classified into certain classes
Data classified
No
Yes
Define the fit set for the classification
Apply KNN classifier for the classification
Display the results of classification
STOP

Fig.1: Proposed Flowchart

IV. EXPERIMENTAL RESULTS

The proposed algorithm is implemented in Python and the results are compared with existing approach in terms of accuracy and execution time.

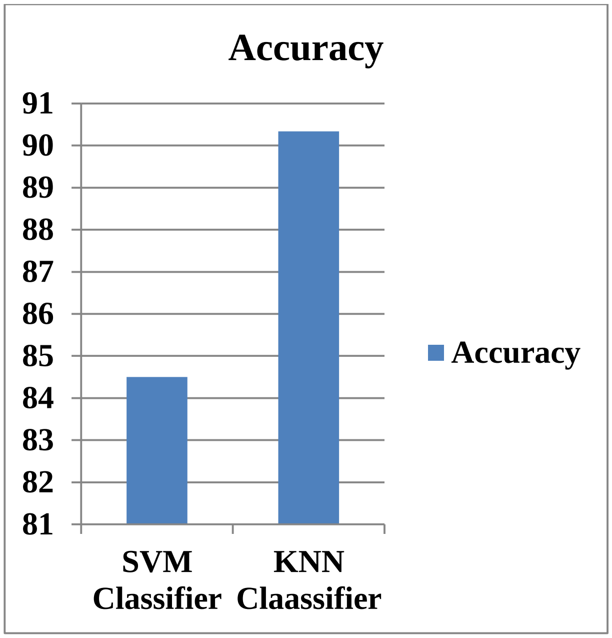

Fig.2: Accuracy Comparison

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

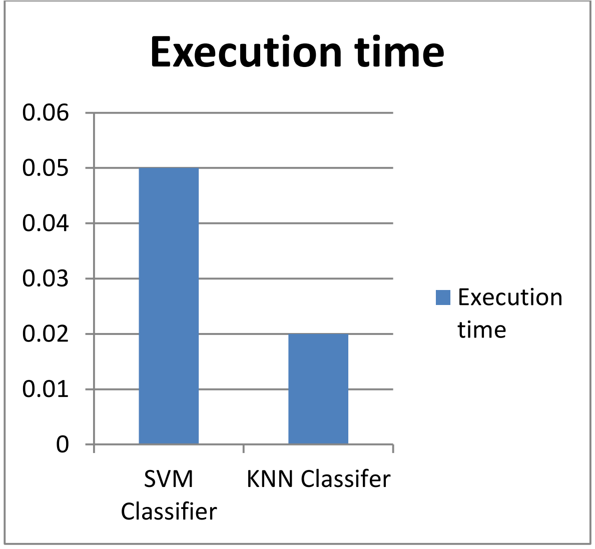

Fig.3: Execution Time

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.

V. CONCLUSION

Data classification is an important task in machine learning. It is identified with develop computer programs ready to gain from labeled data sets and, in this way, to predict unlabeled instances. 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 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. Examples include the confidence factor and the minimum number of cases to partition a set in C4.5 decision tree; the K value in KNN; the stop criterion, the number of neurons, 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.

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

[9]. 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)

[10]. 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

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

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

[13]. Mohammad Tariq Meeran, Paul Annus, Yannick Le Moullec, "Approaches for Improving VoIP QoS in WMNs", IEEE, 2017