INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

(Volume2, Issue5)

Available online at: www.ijariit.com

A Review on Traffic Classification Methods in WSN

Jaskirat Singh*

*Department of Electronics and Communication Engineering, Doaba Institute of Engineering and Technology, Ghataur (Punjab)

hunjan.jaskirat@gmail.com

Harpreet Kaur Saini**

**Department of Electronics and Communication Engineering, Doaba Institute of Engineering and Technology, Ghataur (Punjab)

saini.kim3@gmail.com

Abstract— In a wireless network it is very important to provide the network security and quality of service. To achieve these parameters there must be proper traffic classification in the wireless network. There are many algorithms used such as port number, deep packet inspection as the earlier methods and now days KISS, nearest cluster based classifier (NCC), SVM method and used to classify the traffic and improve the network security and quality of service of a network.

Keywords— Traffic classification, Network security, SVM, Deep packet inspection.

I. INTRODUCTION

In a wireless network traffic classification and unknown flow detection methods are used to solve the networking issues such as security, congestion, intrusion detection and quality of service issues [3]. A number of supervised classification algorithms and unsupervised clustering algorithms have been applied to network traffic classification. In supervised traffic classification the flow classification model is learned from the labeled training samples of each predefined traffic class. But sometimes existing traffic classification methods suffer from poor performance in the crucial situation where more and more new/unknown applications are emerging in the cloud computing based environment [5]. In this paper, we aim to tackle the problem of unknown flows in a WSN.This work considers very few labeled training samples and investigates flow correlation in real world network environment, which makes it better to the previous works.

Another approach that is presented is KISS algorithm. It a novel classifier explicitly targeting UDP traffic that couples the stochastic description of application protocols with the discrimination power of Support Vector Machines. Signatures are extracted from a traffic stream by the means of Chi-square like test that allows application protocol format to emerge, while ignoring protocol synchronization and semantic rules. A decision process based on Support Vector Machine is then used to classify the extracted signatures, leading to exceptional performance. Performance of KISS has been tested in different scenarios, considering both data, VoIP, traditional P2P applications and novel P2PTV systems [3]. This work considers very few labeled training samples and investigates flow correlation in real world network environment, which makes it better to the previous works.

II. RELATED WORK

A. K-MEANS CLUSTERING ALGORITHM

The paper on Mining Unclassified Traffic using Automatic Clustering Techniques discusses the, several traffic classification techniques have been proposed in the last years. In the beginning port-based approaches were mainly used; however, the characteristics of many nowadays applications that employ randomly chosen ports, significantly reduce the effectiveness of these approaches. Those are today abandoned in favor of deep packet inspection (DPI) or behavioral techniques. This helped us revealing 40% of the traffic that we could not classify with the previous used classifiers. Second, the algorithm can reveal the born of new applications, as well as the changes of existing ones.

B. KISS: STOCHASTIC PACKET INSPECTION CLASSIFIER

The paper on KISS: Stochastic Packet Inspection Classifier [3] discuss about the method KISS, a novel Internet classification engine.

Motivated by the expected raise of UDP traffic, which stems from the momentum of P2P streaming applications, a novel classification framework which leverages on statistical characterization of payload. Statistical signatures are derived by the means of a Chi-Square like test, which extracts the protocol "format", but ignores the protocol "semantic" and "synchronization" rules. The signatures feed a decision process based either on the geometric distance among samples, or on Support Vector Machines. KISS is very accurate, and its signatures are intrinsically robust to packet sampling, reordering, and flow asymmetry, so that this algorithm can be used on almost any network. KISS is tested in different scenarios, considering traditional client-server protocols, VoIP and both traditional and new P2PInternet applications. The average True Positive percentage is 99.6%, and in the worst case equal 98.1%.

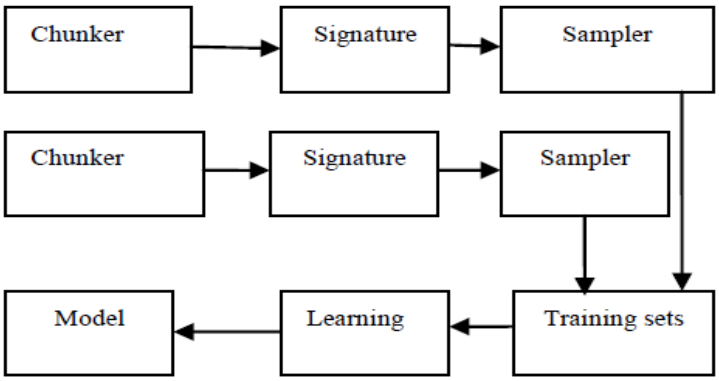

Fig. 1 KISS algorithm

C. SUPPORT VECTOR MACHINES (SVM) BASED TRAFFIC CLASSIFIER

The paper A Survey on Recent Traffic Classification Techniques Using Machine Learning Methods [4], an approach to traffic classification is proposed which is based on SVM, that approach is used to solve the mutli-class problem that arises in SVM, classify network traffic and apply optimization algorithm to make classifier perform properly even with a bad training set for hundreds of samples. It is one of the most promising ML tools, a binary classifier suitable for solving high dimensional feature space and small training set size problems. This approach uses flow representation that describes the statistical characteristics of application protocol through monitoring node whose duty is to assign flows to the concerned application classes it was trained with or with unknown class. It considers bidirectional flows only, which follows proper TCP three way handshakes and proper termination. After packets are captured, each flow is mapped to feature values which are based on packet's length to determine which application the packet belongs to. The main feature of this proposed classifier is based on packet size as it is being captured on the application layer. By the use of this classifier accuracy is very good with True Positive of 90%.

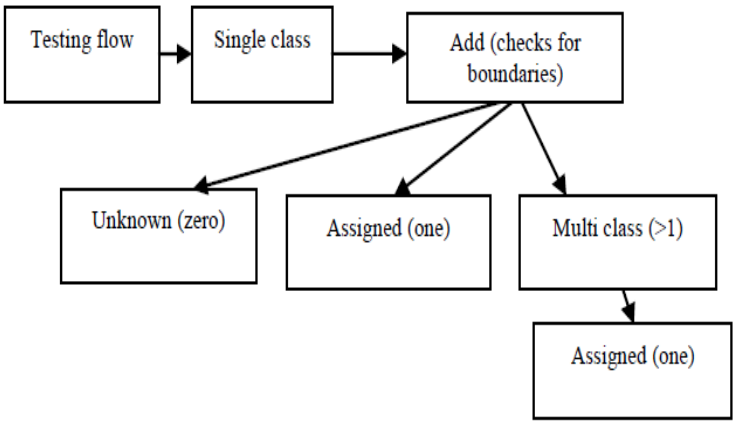

Fig 2 Support Vector Machines (SVM) based traffic classifier

D. NAIVE BAYESIAN AND BAYESIAN NEURAL NETWORK BASED TRAFFIC CLASSIFICATION

It is a traffic classifier that can achieve a high accuracy across a range of application types without any source or destination Host-address or port information can be designed using supervised machine learning based on a Bayesian trained neural network. Bayesian neural network (NN) based traffic classifier [6] can produce more accurate results compared to naïve Bayesian traffic classifier. The Bayesian framework using a neural network model allows identification of traffic without using any port or host information. A

classification accuracy of over 99% can be achieved when training and testing on homogeneous traffic from the same site on the same day.

E. EFFICIENT FLOW BASED NETWORK TRAFFIC CLASSIFICATION USING MACHINE LEARNING

Traffic classification based on their generation applications has a very important role to play in network security and management. The port-based prediction methods and payload-based deep inspection methods comes under Traditional methods. The standard strategies in current network environment suffer from variety of privacy issues, dynamic ports and encrypted applications. Recent research efforts are focused on traffic classification and Machine Learning Techniques are used for classification. This paper conducts a flow based traffic classification and comparison on the various Machine Learning (ML) techniques such as C4.5, Naïve Bayes, Nearest Neighbor, and RBF for IP traffic classification. From this C4.5 Decision Tree gives 93.33% accuracy compare with other algorithms. The two methods are used Full Feature selection and reduced feature set for classification. From this classification the reduced feature selection gives good result.

III. OBJECTIVES

We are aiming to find a novel approach for traffic classification, which can improve the classification performance effectively. The problem at hand is to detect the malicious traffic flow in the network and work out a new route which is free from the faulty nodes. In the beginning we have to design a network that contains nodes (some good nodes and some bad nodes). Then we have to make a flow from one point to the point. In this way with the help of previous used methods we will try to get better results and more accuracy then the earlier used methods.

IV. PLANNING OF WORK

In the planning firstly we have to design a network that contains the sensor nodes in the particular described area. Then we have to find that how many among then are proper and how many of them have some issues due to which they will be unable to participate during the transmission/ requirement time. Then to develop a path that follows the shortest path and ignore the sensor nodes that are not participating. As well as shortest path we have to achieve the maximum efficiency during this traffic classification.

V. ACKNOWLEGEMENT

The authors are thankful to Mrs. Maninder kaur, M.tech coordinator, DIET and her staff for providing the necessary facilities for the preparation of the paper. Without the proper guidance of them it is not possible to learn about latest technologies and research works.

REFERENCES

[1]J. Zhang, Y. Xiang, Y. Wang,W. Zhou, Y. Xiang, and Y. Guan, "Network traffic classification using correlation information," *IEEE Trans. Parallel Distributed. Syst.*, vol. 24, no. 1, pp. 104–117, Jan. 2013.
[2] A. Finamore, M. Mellia, and M. Meo, "Mining unclassified traffic using automatic clustering techniques," in *Proc. 2011 TMA International Workshop on Traffic Monitoring and Analysis*, pp. 150–163.
[3] A. Finamore, M. Mellia, M. Meo, and D. Rossi, "KISS: stochastic packet inspection classifier for UDP traffic," *IEEE/ACM Trans. Netw.*, vol. 18, no. 5, pp. 1505–1515, Oct. 2010.
[4] T. T. Nguyen and G. Armitage, "A survey of techniques for internet traffic classification using machine learning," *IEEE Common. Surveys& Tutorials*, vol. 10, no. 4, pp. 56–76, Fourth Quarter 2008.
[5] J. Erman, A. Mahanti, and M. Arlitt, "Internet traffic identification using machine learning," in *Proc. 2006 IEEE Global Telecommunications Conference*, pp. 1–6.
[6] A. W. Moore and D. Zuev, "Internet traffic classification using Bayesian analysis techniques," *SIGMETRICS Perform. Eval. Rev.*, vol. 33, pp. 50–60, June 2005.

