



A Review on Hybrid Fingerprinting on Internet Traffic

Amit kumar Sahu

M.Tech Scholar,

Department of Computer Science
and EngineeringRKDF Institute of Science and
Technology (R.G.P.V),
Bhopal, India**Chinmay Bhatt**

Asst. Professor

Department of Computer Science
and EngineeringRKDF Institute of Science and
Technology (R.G.P.V),
Bhopal, India**Shrikant Lade**

Asst. Professor

Department of Computer Science
and EngineeringRKDF Institute of Science and
Technology (R.G.P.V),
Bhopal, India

Abstract: *Countless have been proposed on intrusion detection system, which prompts the execution of executor based intelligent IDS (IIDS), Non – intelligent IDS (NIDS), signature based IDS and so on. While building such IDS models, taking in calculations from stream of system movement assume essential part in precision of IDS frameworks. The proposed work concentrates on actualizing the novel strategy to group system activity which wipes out the restrictions in existing internet grouping calculations and demonstrates the heartiness and correctness over expansive stream of system movement touching base at to a great degree high rate. We contrast the current calculation and novel routines to examine the correctness and unpredictability.*

Keywords— *Fingerprinting, Data Stream Mining, Clustering algorithm*

I. INTRODUCTION

Quick development being used of systems administration and web makes security essential in late decades. The latest subject in system security is Network Intrusion Detection System (NIDS) which keeps the security at the most abnormal amount. Numerous differing methodologies have been proposed and actualized, which minimizes the assaults and helplessness in the system and makes it secure. Most broadly utilized NIDS are signature based models [1]. Such models identify just known assaults, thus discovering obscure assaults without former information about particular interruption remains a test. To adapt to these difficulties, intelligent IDS frameworks have advanced [2].

The IIDS framework concentrate on particular example of known assaults, which uncovers the underlying driver of interruption by always gaining from system activity, and if such examples are distinguished and scholarly, they can create the characterization model for potential interruption. Such frameworks are packaged with two layers, the first layer is preparing or learning layer, which takes in the examples of interruption in the stream of system activity. An alternate layer is trying, which applies educated standards to distinguish interruptions in obscure activity information. As gaining from online information is trying than gaining from static information, it got to be crucial to give consideration towards exactness of stream grouping calculations [3][4].

The proposed model concentrate on gaining from system movement by applying inventive stream grouping calculations and afterward make utilization of created bunches to fabricate characterization models for IIDS. Also the methodology advocate the proficiency and straightforwardness of new calculation by contrasting it and existing RAH grouping calculation.

II. LITERATURE SURVEY

The expansion of World Wide Web and increased use of internet has increased the risk of harmful intrusion every day. To cope with potential harmful intrusions, many diverse techniques have evolved. The diverse approaches include histogram based anomaly detection models [5], Hidden Markow for IDS [6], IDS using Neural Networks [7], IDS using Genetic Algorithms [8] and Signature Based IDS [1].

In [11], the authors discussed Profound Packet Inspection (DPI) module in Intrusion Detection Systems (Idses) comprises of two segments: Pre-channel and Rule Verification (RV). Prefilter receives Multi-Pattern Matching (MPM) motor to channel out the greater part of generous parcels and after that leave a couple of suspicious bundles with false positives into RV segment. These false positives are because of the filtering process in the prefilter: it recognizes the movement in a solitary pass against a set of fingerprints, which are concentrated from the given ruleset by selecting just a little divide of the examples in every signature. RV segment absolutely checks the suspicious parcels and wipes out these false positives. The execution of DPI module is identified with the concentrated unique mark set. A productive finger impression set ought to enhance the prefilter throughput, and in the meantime diminish the include of checking exercises RV part. We indicate in this paper that these two prerequisites can't be at the same time fulfilled in the current unique mark extraction techniques. Prefilter execution enormously profits from more diminutive unique finger impression set in light of the more smaller MPM motor. However RV segment experiences the higher rate of false positives brought about by the more modest finger impression set. We ideally exchange off these two necessities with another extraction technique in this work. Through investigating a little measure of preparing activity in the beginning stage, our method

gives each one unique mark applicant an exact weight for the ensuing extraction. Trial results acquired by incorporating our proposed system into the Snort IDS demonstrate that our method enhances the IDS normal throughput by no less than 69% over the most recent genuine ruleset and true activity.

NIDS using neural network introduces two layered architecture [7]. The first layer is training of neural network by either feed forward network or recurrent network and second layer introduces testing of network traffic by diverting it towards trained neural network.

NIDS using data mining is most diverse among all approaches. The basic model introduces training and testing phases. The training phase learns the flow of network. To do so, it can use either online network stream or offline batch of network traffic data. To learn from network stream various stream classification algorithms are used, for e.g. CluStream [4], Hoeffding Tree and VFDT [3].

The signature based IDS systems uses attack signatures to classify unknown traffic, and updates signature data whenever new signatures are found. The data mining approach for NIDS also uses clustering approaches to group the network traffic in specific classes which can be further used by classification modules to classify the data with high accuracy. However the traffic is online and arriving at extremely high rate, which is to be clustered immediately when it arrives. This is concerned with online clustering algorithms and various online clustering approaches can be used to cluster this online data, but the issue remains is about the time complexity of online clustering algorithm. The time complexity is crucial part of such algorithm because the samples arrive so fast, and in large number so we would not have enough resources to store them before analysis. Moreover the clustering output is demanded very quickly by classification algorithms, where we cannot wait for batch of network packets to arrive which would be processed later.

In [9], the authors employed several supervised machine learning algorithms, namely, J48, Boosted J48, Naive Bayesian, Boosted Naive Bayesian and SVM in order to classify malicious and non malicious traffic. The aforementioned learning algorithms were used to build classification models. So far, results show that J48 and Boosted J48 performed better than other algorithms. There future works will fall into classify the malicious traffic accordingly to malware types and families, and deploying the model on a network in order to test its performance on realtime traffic.

III. DATASET

The dataset for the proposed research is created by using the packet sniffer, which sniffs the incoming packets through the network adapter. The sniffer is designed such that user can configure the attributes of packet that are traced by the sniffer. Many studies have revealed that the attributes such as Source IP Address, Destination IP Address, Source Port Number, Destination Port Number, TCP Window Size, and TCP Data Length are most promising fields associated with different types of attacks [10], hence the above fields are considered while applying clustering algorithm on the stream of packets.

Input to the clustering algorithm is mostly one or two dimensional numeric data points. By having single dimensional data, the similarity between two samples can be computed by taking direct difference between two values. For two dimensional sample data, the similarity between two samples is computed by Euclidian or Manhattan distance measures. But for multi-dimensional categorical data, the difference measure is very challenging. To calculate distance among network packets there is no standard measure.

The packet is set of attributes and every attribute may have numeric or categorical values. To compute the similarity among packets, first we need to focus on specific attribute values. If such selected attribute values for both packets are equal, then we can state that two packets are similar. But predicting such similarity on the basis of single attribute would not give accuracy, so we require multiple attributes and their priorities. The following algorithm explains the similarity measure between two packets based on attribute priority technique.

IV. PROPOSED METHOD

In this research work, a general research methodology has been adopted which is shown in following figure.

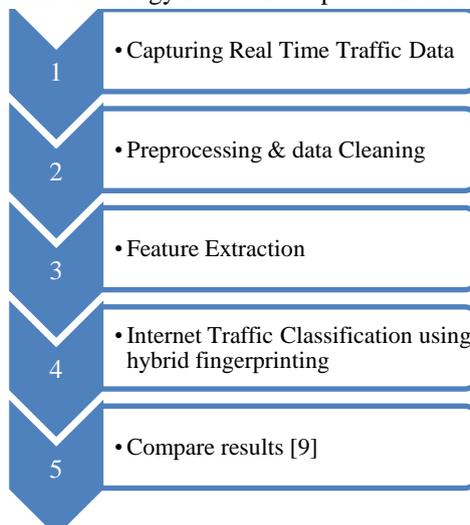


Fig 1. Proposed Methodology

The proposed approach will show the network traffic captured by packet sniffer. The packet sniffer is configured to capture packets with attributes – source port, destination port, source IP address, destination IP address, TCP length, TCP checksum. The module of proposed hybrid tool will cluster every network packet into specific category of cluster using hybrid clustering algorithm. For calculating similarity measurement among packet, three attributes are selected in their incremental priority order. These three attributes are source IP address, destination port number and TCP header length.

V. CONCLUSION

Traffic classification encounters more critical problems in current advanced network and system, especially in cloud computing environment. In this paper, we proposed a novel traffic classification method to address the problem associated with past researches. The proposed method uses hybrid clustering technique to cluster the network according to the nature and type of incoming packets. A large number of experiments will be carried out on two real-world traffic datasets to evaluate the efficiency of proposed method. Also it is expected that the proposed method will display more robust ability to various parameters and superior unknown detection performance especially on false detection.

REFERENCES

- [1] Farooq Anjum, Dhanant Subhadrabandhu and Saswati Sarkar,” —Signature based Intrusion Detection for Wireless Ad-Hoc Networks: A Comparative study of various routing protocols”, Telcordia. Tech Inc. Morristown NJ 07960J.
- [2] N.Jaisankar and R.Saravanan K. Durai Swamy,”Intelligent Intrusion Detection System Framework Using Mobile Agents”, International Journal of Network Security & Its Applications (IJNSA), Vol 1, No 2, July 2009R.
- [3] Pedro Domingos, Geoff Hulten, —Mining High Speed Data Streams”.
- [4] Charu C. Aggarwal, Jiawei Han, Jianyong Wang, Philip S. Yu,” A Framework for Clustering Evolving Data Streams”, Proceedings of the 29th VLDB Conference, Berlin, Germany, 2003.
- [5] Andreas Kind, Marc Ph. Stoecklin, and Xenofontas Dimitropoulos,” Histogram-Based Traffic Anomaly Detection”, IEEE Transactions On Network Service Management, Vol. 6, No. 2, June 2009
- [6] Jiankun Hu and Xinghuo Yu,” A Simple and Efficient Hidden Markov Model Scheme for Host-Based Anomaly Intrusion Detection”
- [7] Jake Ryan, Meng-Jang Lin, —Intrusion Detection with Neural Networks”, Advances in Neural Information Processing Systems 10, Cambridge,MA:MIT Press, 1998.
- [8] Vivek K. Kshirsagar, Sonali M. Tidke & Swati Vishnu,” Intrusion Detection System using Genetic Algorithm and Data Mining: An Overview”, International Journal of Computer Science and Informatics ISSN (PRINT): 2231 –5292, Vol-1, Iss-4, 2012.
- [9] Amine Boukhtouta, Nour-Eddine Lakhdari, Serguei A. Mokhov, Mourad Debbabi, “Towards Fingerprinting Malicious Traffic”, The 4th International Conference on Ambient Systems, Networks and Technologies (ANT 2013), Procedia Computer Science 19 (2013) 548 – 555.
- [10] Anna Sperotto, Gregor Schaffrath, Ramin Sadre, Cristian Morariu, Aiko Pras and Burkhard Stiller,” An Overview of IP Flow-Based Intrusion Detection”, IEEE Communications Surveys & Tutorials, Vol. 12, No. 3, Third Quarter 2010.
- [11] Haiyang Jiang ; Inst. of Comput. Technol., Beijing, China ; Gaogang Xie ; Salamatian, K, “Efficient Fingerprint Extraction for High Performance Intrusion Detection System”, Computers and Communications (ISCC), 2013, pp. 000179 - 000184.