Volume 5, Issue 1, January 2015 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

An Approach to Determine Malicious Activity by Fingerprinting Network Level Traffic

Amit Kumar Sahu
M Tech Scholor
Department of CSE
RKDFIST Bhopal, India

Chinmay Bhatt
Assistant Professor
Department of CSE
RKDFIST Bhopal, India

Shrikant Lade
Assistant Professor
Department of CSE
RKDFST Bhopal, India

Abstract: ***This paper works depicts a far reaching study on the current techniques and strategies for the Network traffic fingerprinting. In this thesis, we exhibit a novel methodology called Hybrid Fingerprinting Network Traffic. The proposed work focuses on realizing the novel system to gathering framework movement which wipes out the limitations in existing web gathering estimations and shows the generosity and accuracy over an extensive stream of framework development connecting to an incredible degree high rate.***

Fingerprinting the malignant: Packet Symmetry concentrates on system level structure and powers the instinct that decently acted applications don't transmit boundlessly a bigger number of bundles than they get. Traffic examination confirms the possibility of utilizing bundle symmetry for edge-based, entrance centered anticipation methodologies for volume-based assaults. We differentiate the current count and novel schedules to analyze the rightness and unconventionality of the proposed one, in which we are producing computerized system based malware signature technique for fingerprinting the web traf

Keywords— Fingerprinting, Data Stream Mining, Signature based 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].

II. PROBLEM STATEMENT

The key issues on building a packet filter are:

- Real-time performance: the packet filter should be able to quickly capture a raw packet from the data link layer and process it in a short period of time.
- No packet dropping: no packet dropping is allowed, especially for a network intrusion detection system. The information missed from dropped packets can make the whole detection scheme fail.
- Flexibility: the specification of packet patterns can be modified easily to support different communication protocols.
- Scalability: in terms of a system for network intrusion detection, new intrusion signatures can be added into the packet filter without degrading performance.

III. LITERATURE REVIEW

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 [10], 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 thatJ48 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 real-time traffic.

Classification according to the kinds of knowledge mined: Data mining systems can be categorized according to the kinds of knowledge they mine, i.e., based on data mining functionalities, such as characterization, discrimination, association, classification, clustering, trend and evolution analysis, deviation analysis, similarity analysis, etc. A comprehensive data mining system usually provides multiple and/or integrated data mining functionalities.

Moreover, data mining systems can also be distinguished based on the granularity or levels of abstraction of the knowledge mined, including generalized knowledge (at a high level of abstraction), primitive-level knowledge(at a raw data level), or knowledge at multiple levels (considering several levels of abstraction). An advanced data mining system should facilitate the discovery of knowledge at multiple levels of abstraction.

Classification according to the kinds of techniques utilized: Data mining systems can also be categorized according to the underlying data mining techniques employed. These techniques can be described according to the degree of user interaction involved (e.g., autonomous systems, interactive exploratory systems, query-driven systems), or the methods of data analysis employed (e.g., database-oriented or data warehouse-oriented techniques, machine learning, statistics, visualization, pattern recognition, neural networks, and so on). A sophisticated data mining system will often adopt multiple data mining techniques or work out an effective, integrated technique which combines the merits of a few individual approaches.

IV. PROPOSED TECHNIQUE

In this research work, a general research methodology has been adopted which is shown in following figure 1.

The proposed approach will shows 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 clusters every network packet into specific category of cluster using hybrid Fingerprinting with Signature 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.

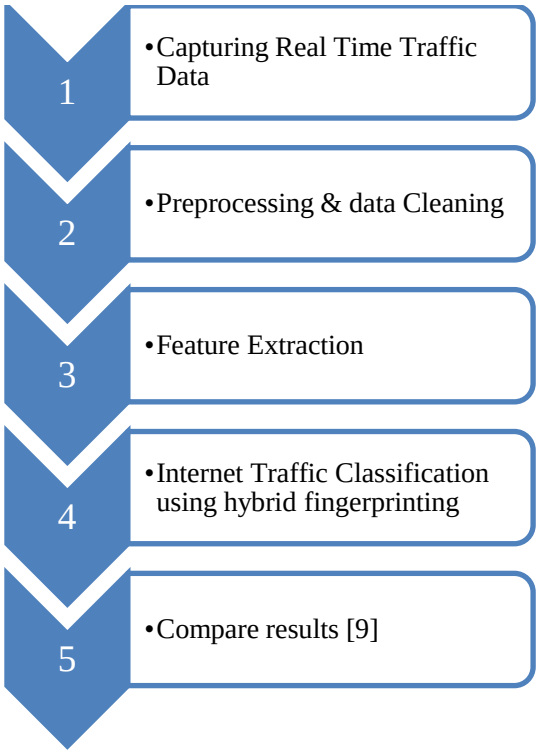

Figure 1: Proposed Methodology

V. IMPLEMENTATION DETAILS

The proposed scheme implemented by developing a web application, using Microsoft .Net Framework 4.0, Visual Studio 2010. Which is tested on windows environment. Which is one of the best combination for creating data-driven sites. Since both efforts are collaborative in nature, there's always plenty of support from documentation and mailing lists. Bugs are fixed rapidly, and requests for features are always heard, evaluated, and if feasible, implemented.

VI. RESULT ANALYSIS

In this section, we evaluate the performance of our proposed method under the network traffic caught by the network sniffer. The experimental results demonstrate that the proposed algorithm perform well in terms of the accuracy, sensitivity. Significant signatures that are useful for predicting the vulnerability were extracted from the dataset.

Table 1: Test scenario-1

TEST-1	**Actual Scan**	**Actual Attack**
Predict Scan	44	0
Predict Attack	1	1

Table 2: Test scenario-2

TEST-2	**Actual Scan**	**Actual Attack**
Predict Scan	325	4
predict Attack	2	98

Table 3: Test scenario-3

TEST-3	**Actual Scan**	**Actual Attack**
Predict Scan	458	23
Predict Attack	12	2

Table 4: Test scenario-4

TEST - 4	**Actual Scan**	**Actual Attack**
Predict Scan	674	0
Predict Attack	5	87

Table 5: Decision parameters – metrics, basis for fingerprinting the traffic

Result	**Sensitivity**	**Accuracy**
Test-1	0.977778	0.978261

Test-2	0.993884	0.986014
Test-3	0.974468	0.929293
Test-4	0.992636	0.993473

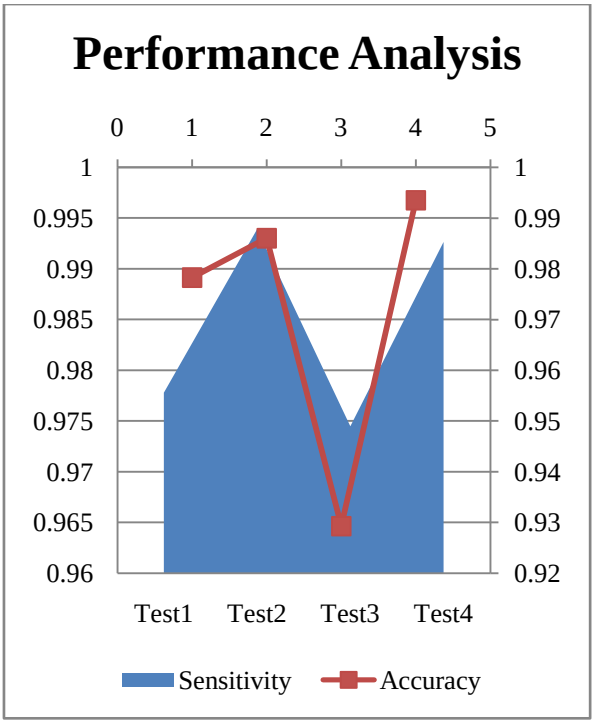

Figure 2: Results

VII. CONCLUSION

Traffic classification experiences more discriminating issues in present propelled system and framework, particularly in distributed computing environment. In this paper, we proposed a novel traffic classification strategy to address the issue connected with past inquires about. The hidden perception is that decently carried on applications don't send a lot of parcels without accepting any, transforming bundle asymmetry exceedingly skewed to transmitting side into a fingerprint of pernicious movement. I displayed an estimation study confirming that decently carried on traffic is to be sure exceptionally symmetric at the bundle level. Rather than the substance based methodologies exhibited prior, the concentrate here is not on computerized era of such a fingerprint however on the practicality of finding general disagreeableness fingerprints. The proposed technique utilizes mixture Fingerprinting with Signature method to bunch the system as indicated by the nature and sort of approaching parcels. An extensive number of examinations will be completed on two certifiable traffic datasets to assess the productivity of proposed technique. Additionally it is normal that the proposed system will showed more hearty capacity to different parameters and prevalent obscure recognition execution particularly on false detection.

VIII. FUTURE WORK

The proposed work can be further upgraded and stretched for the mechanization of vulnerabilities location in a period successful way with high precision. We additionally want to accomplish more formal correlation of the execution of different mixes of machine learning calculations and gimmick choice routines alongside some heuristic strategy with add expense to each bundle of traffic. At last, would it say it is conceivable to tune surmised matching calculations further, to get better result.

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] 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.

