Research Article Impact Factor: 0.621 ISSN: 2319-507X
Niraj Telrandhe, IJPRET, 2014; Volume 2 (9): 396-403 IJPRET

INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

DETECTING ANOMALOUS NETWORK TRAFFIC BASED ON CLASSIFYING FREQUENT TRAFFIC PATTERNS METHOD & TOOLS

NIRAJ TELRANDHE[1], VIMAL PAL[2], MANGESH WANJARI[3]

1. M. tech Scholar, Wainganaga College of Engineering& Management, Nagpur.
2. Asst. Prof, Wainganaga College of Engineering& Management, Nagpur.
3. Asst. Prof, Ramdevbaba Kamla Nehru Engineering College, Nagpur.

Accepted Date: 27/02/2014 ; Published Date: 01/05/2014

Abstract: Network traffic anomalies stand for a large fraction of the Internet traffic and compromise the performance of the network resources. Detecting and diagnosing these threats is a laborious and time consuming task that network operators face daily. During the last decade researchers have concentrated their efforts on this problem and proposed several tools to automate this task. Thereby, recent advances in anomaly detection have permitted to detect new or unknown anomalies by taking advantage of statistical analysis of the traffic. In spite of the advantages of these detection methods, researchers have reported several common drawbacks discrediting their use in practice. Indeed, the challenge of understanding the relation between the theory underlying these methods and the actual Internet traffic raises several issues. This paper discusses a statistical approach to analysis the distribution of network traffic to recognize the normal network traffic behavior also discusses a method to recognize anomalies in network traffic, based on a non-restricted α-stable first -order model and statistical hypothesis testing.

Keywords: Change-Point Detection, Wavelet Analysis, Principal Component Analysis

Corresponding Author: MR. NIRAJ TELRANDHE

Access Online On:

www.ijpret.com

How to Cite This Article:

Niraj Telrandhe, IJPRET, 2014; Volume 2 (9): 396-403

PAPER-QR CODE

INTRODUCTION

Due to advancements in Internet technologies and the concomitant rise in the number of network attacks, network intrusion detection has become a significant research issue. In spite of remarkable progress and a large body of work, there are still many opportunities to advance the state-of-the-art in detecting and thwarting network-based attacks [1].

According to Anderson [2], an intrusion attempt or a threat is a deliberate and

Unauthorized attempt to (i) access information, (ii) manipulate information, or (iii) render a system unreliable or unusable. For example, (a)Denial of Service (DoS) attack attempts to starve a host of its resources, which are needed to function correctly during processing; (b)Worms and viruses exploit other hosts through the network; and (c)Compromises obtain privileged access to a host by taking advantages of known vulnerabilities.

The termanomaly-based intrusion detection in networks refers to the problem of finding exceptional patterns in network traffic that do not conform to the expected normal behavior. These nonconforming patterns are often referred to as anomalies, outliers, exceptions, aberrations, surprises, peculiarities or discordant observations in various application domains . Out of these, anomalies and outliers are two of the most commonly used terms in the context of anomaly-based intrusion detection in networks.

Anomaly detection has extensive applications in areas such as fraud detection for credit cards, intrusion detection for cyber security, and military surveillance for enemy activities. For example, an anomalous traffic pattern in a computer network may mean that a hacked computer is sending out sensitive data to an unauthorized host.

II. Statistical Approaches for Network Anomaly Detection:

statistical approach uses some steps for detecting network anomaly. The first step is to pre-process or filter the specified data inputs. This is a significant step as the types ofdata presented and the time scales in which these data are measured can significantly distress the detection performance [4]. In the second step, statistical examination and/or data transforms are performed to take apart normal network behaviors from anomalous behaviors and noise. A diversityof techniques can be applied here, like Covariance Matrix analysis, Wavelet Analysis, and Principal Component Analysis. The primary challenge here is to find computationally proficient techniques for anomaly detectionwith low false alarm rate. At last in final step, decision theories for instance Generalized Likelihood Ratio (GLR) test can be used to determine whether there is a network anomaly based on the deviations observed.

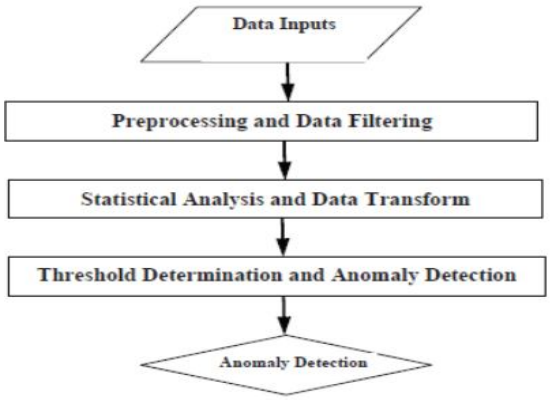

Fig: 1 Statistical Approaches for Network Anomaly Detection

In a larger context, statistical anomaly detection can also be inspected from the machine learning point of view, where the objective is to find appropriate discriminant functions that can be accessed to classify any new input data vector into the normal or anomalous region with excellent accuracy for anomaly detection. One restrained difference among statistical anomaly detection and machine learning based methods is that statistical approaches generally focus on statistical analysis of the collected data, whereas machine learning methods focuses on the "learning" part.

2.1 Change-Point Detection: Change-point detection is the difficulty of discovering time points at which properties of time-series data change. This includes a broad range of real world problems and has been vigorously conversed in the community of statistics and data mining. A representative statistical formulation of change-point detection is to consider probability distributions from which data in the past and present intervals are generated, and observe the intention time point as a change point if two distributions are significantly different. In these methods, the logarithm of the probability ratio between two successive intervals in time series data is observed for detecting change points.

2.2 Wavelet Analysis: Wavelet analysis has been applied to modeling of non-stationary data series because it can characterize the scaling properties in the temporal and frequency domains. The wavelet transform can get arbitrary signal characteristic of time-frequency domain, which can help to explore the transient abnormal phenomenon from normal signals and demonstrate its components. Researchers used wavelet analysis to detect anomaly just based on the differences between the normal and anomalous traffic signals in the frequency domain.

sequences. The norm profile of the normal traffic can then be described by the mathematical expectation of all covariance matrices constructed from samples of the

normal class in the training dataset. The covariance matrix method is extended, where the sign of the 2.3 Principal Component Analysis: Principal Component Analysis (PCA) is a dimensionality-reduction method of mapping a set of data points onto new coordinates. The spirit of PCA based anomaly detection is to separate the normal behavior from anomalies through dimensionality-reduction. The basic idea of using PCA for traffic anomaly detection is that: the k-subspace obtained through PCA corresponds to the normal behavior of the traffic, whereas the remaining (n-k) subspace corresponds to either the anomalies or the anomalies and the noise. Each new traffic measurement vector is projected on to the normal subspace and the anomalous subspace. Afterwards, different thresholds can be set to classify the traffic measurement as normal or anomalous. Later on the source of the inconsistent traffic can then be pinpointed by determining the ingress and egress points of different traffic flows.

III. Discrete Algorithms for Network Anomaly Detection:

In numerous cases, network anomaly detection involves tracking noteworthy changes in traffic patterns such as traffic amount or the number of traffic connections. Due to the high link speed and the large volume of the Internet, it is generally not scalable to trace the per-flow status of traffic. By limiting the number of flows that require to be monitored, sampling can incompletely solve the scalability problem at the cost of anomaly detection performance. However, simple sampling cannot completely solve the scalability problem as any packets or flows that are not sampled may contain important information about anomalies. Furthermore, it is expected that this information can only be recovered if these packets or flows are sampled and stored. Specifically, using streaming techniques, the anomaly detection trouble can be formulated as a heavy-hitter detection problem or a heavy change detection problem. In the heavy-hitter detection problem, the main target is to recognize the set of flows that represent a significantly large proportion of the ongoing traffic or the capacity of the link.

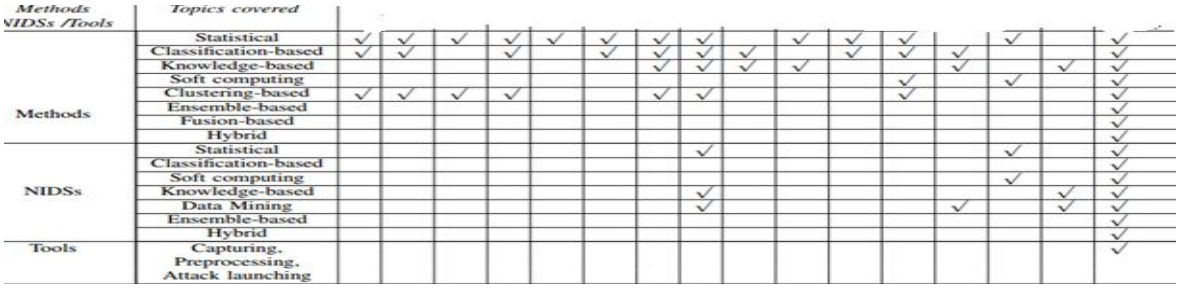

Fig: 2 Comparison of Survey.

IV. OVERVIEW OF NETWORK ANOMALY DETECTION:

Anomaly detection attempts to find patterns in data, which do not conform to expected normal behavior. The importance of anomaly detection is due to the fact that anomalies in data translate to significant (and often critical) actionable information in a wide variety of application domains [10]. For example, an anomalous traffic pattern in a computer network could mean that a hacked computer is sending out sensitive data to an unauthorized host. However, anomalies in a network may be caused by several different reasons.

As stated in [11], there are two broad categories of network anomalies: (a) performance related anomalies and (b) security related anomalies. Various examples of performance related anomalies are: broadcast storms, transient congestion, babbling node, paging across the network, and file server failure.

Security related network.

Technique	*Characteristics*
Misuse-based	(i) Detection is based on a set of rules or signatures for known attacks. (ii) Can detect all known attack patterns based on the reference data. (iii) How to write a signature that encompasses all possible variations of the pertinent attack is a challenging task.
Anomaly-based	(i) Principal assumption: All intrusive activities are necessarily anomalous. (ii) Such a method builds a *normal activity profile* and checks whether the system state varies from the established profile by a statistically significant amount to report intrusion attempts. (iii) Anomalous activities that are not intrusive may be flagged as intrusive. These are false positives. (iv) One should select threshold levels so that neither of the above two problems is unreasonably magnified nor the selection of features to monitor is optimized. (v) Computationally expensive because of overhead and possibly updating several system profile matrices.
Hybrid	(i) Exploits benefits of both misuse and anomaly-based detection techniques. (ii) Attempts to detect known as well as unknown attacks.

V. RELATED WORK

Ignasi Paredes-Oliva et al [12] proposed Practical Anomaly Detection based on Classifying Frequent Traffic Patterns. They introduce a novel scheme and build a system to sense and classify anomalies that are based on an elegant blend of frequent item-set mining with decision tree learning. This scheme automatically identifies and classifies anomalies in high-speed networks using traffic flow data, like Net Flow. They combines normally two techniques on is from data mining and another is machine learning .

In this method frequent item-set mining (FIM) is used to find a set of frequent item-sets (FIs) firstly. A frequent item-set is a great set of flows that have one or more flow features in common. Secondly builds a decision tree to categorize frequent item-sets as anomalous or benevolent and to determine their specific type in anomalous case. Instinctively, they decompose observed traffic into distinct groups (FIs) of related traffic flows that allows us to categorize each FI with high accuracy [12].Frequent Item-set Mining: Frequent item-set mining (FIM) is a well-known data mining method that focuses on finding that arise frequently together in a certain dataset. A set of items will be measured frequent if they emerge together at least as many times as a specified threshold, which is described as minimum support. Concerning FIM to network traffic permits us to decide groups of numerous flows sharing a certain combination of features [12].

Yingjie Zhou Guangmin Hu Weisong He recommended Using Graph to Detect Network Traffic Anomaly [13]. In this a network traffic anomaly detection method based on time-series graph mining. It perfectly and completely describes the relationships among multi-time series which are used in traffic anomaly detection by timeseries graph, and can efficiently detect the network traffic anomaly; especially DDos attacks .

Curtis Storlie et al [14] proposed Graph Based Statistical Analysis of Network Traffic. They suggest a graph-based method for analyzing traffic patterns in a huge computer network in permutation with novel statistical methods for determining time-related anomalies in data with diurnal trends. They model the traffic as a graph and extort the sub graphs corresponding to individual sessions and use them to develop a statistical model for the network traffic. The aim of this analysis is to find out patterns in the network traffic data that might indicate intrusion activity or other malicious behavior. They also described a statistical method for analyzing TSG decompositions that obtains into account the diurnal patterns of the communications and computes on that basis a predictive model for prospect traffic that can be used to perceive anomalies [7].

In year 2012, Iwan Syarif et al [15] presented unsupervised clustering approach for network anomaly detection. They describe the advantages of using the anomaly detection scheme over the misuse detection technique in detecting unidentified network intrusions or attacks. It also examines the performance of a variety of clustering algorithms when applied to anomaly detection. They implement and compare the performance of five different clustering algorithms in our anomaly detection module which are k-Mean, improved k-Mean, k-Medoids, EM clustering and distance-based outlier detection algorithms [8]. Their results shows that the misuse detection procedure achieves a very good performance result with more than 99% correctness when detecting identified intrusion but it fails to

perfectly detect data set with a large number of unidentified intrusions where the highest accuracy result is only 63.97%.

Again in same year Anup Bhange and Sumit Utareja proposed Anomaly Detection and Prevention in Network Traffic based on Statistical approach and α-Stable Model[16]. As per their research proposals in anomaly detection characteristically follow a four-stage approach. In this scheme the first three stages define the detection mechanism, while the last stage is dedicated to authenticate the scheme. Consequently, in the first stage, traffic data are collected from the network (known as data collection). Second, data are analyzed to extort its most relevant features (i.e. data analysis). Third, traffic is classified as normal or abnormal (inference); and fourth, the entire approach is validated with different types of traffic anomalies [10].

VII. CONCLUSION

Network traffic anomaly refers to the status that the traffic behaviors deviated from its normal behaviors. It can bring great damage to networks and network equipments in a short time. Existing traffic anomaly detection methods usually treat time-varying traffic information as a one-dimensional signal, and detect traffic anomaly through a variety of signal analysis methods. In this paper we give general review of such techniques that are used to detect and classify Anomalies. But it still required improvements.

REFERENCES:

1. A. Sundaram, “An introduction to intrusion detection,” Crossroads, vol. 2, no. 4, pp. 3–7, April 1996.

2. J. P. Anderson, “Computer Security Threat Monitoring and Surveillance,” James P Anderson Co, Fort Washington, Pennsylvania, Tech. Rep., April 1980

3. V. Chandola, A. Banerjee, and V. Kumar, “Anomaly Detection: A Survey,” ACM Computing Surveys, vol. 41, no. 3, pp. 15:1–15:58, September 2009

4. Thottan M., Ji C. “Anomaly Detection in IP Networks”,IEEE Trans. Signal Processing, Special Issue of SignalProcessing in Networking, Vol. 51, No. 8, pp. 2191-2204,2003

5. R. Heady, G. Luger, A. Maccabe, and M. Servilla, “The Architecture of a Network Level Intrusion Detection System,” Computer Science Department, University of New Mexico, Tech. Rep. TR-90, 1990.

6. J. P. Anderson, “Computer Security Threat Monitoring and Surveillance,” James P Anderson Co, Fort Washington, Pennsylvania, Tech. Rep., April 1980.

7. P. Ning and S. Jajodia, Intrusion Detection Techniques. H Bidgoli (Ed.), The Internet Encyclopedia, 2003.

8. F. Wikimedia, "Intrusion detection system,"http://en.

9. M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "Surveying Port Scans wikipedia.org/wiki/Intrusion-detection system,Feb 2009.and Their Detection Methodologies," The Computer Journal, vol. 54, no. 10, pp. 1565–1581, October 2011

10. V. Kumar, "Parallel and distributed computing for cybersecurity," IEEE Distributed Systems Online, vol. 6, no. 10, 2009

11. M. Thottan and C. Ji, "Anomaly detection in IP networks,"IEEE Transactions on Signal Processing, vol. 51, no. 8, pp. 2191–2204, 2003.

12. I. Ignasi Paredes-Oliva, Ismael Castell-Uroz, Pere Barlet-Ros, Xenofontas Dimitropoulos and Josep Sol´e-Pareta"Practical Anomaly Detection based on Classifying Frequent Traffic Patterns", IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS-2012), pp. 49-54, 2012.

13. Yingjie Zhou Guangmin Hu Weisong He "Using Graph to Detect Network Traffic Anomaly", International Conference on Communications, Circuits and Systems (ICCCAS-2009), pp. 341 – 345, 2009

14. Hristo Djidjev, Gary Sandine, and Curtis Storlie "Graph Based Statistical Analysis of Network Traffic", Los Alamos National Lab, MLG '11 San Diego, CA, USA-2011.Onlineavailableat: http://users.cis.fiu.edu/~lzhen001/activities/KDD2011Program/workshops/MLG/doc/paper_10.pdf

15. Iwan Syarif, Adam Prugel-Bennett, Gary Wills"Unsupervised clustering approach for network anomaly detection", Networked Digital Technologies Communications in Computer and Information Science, Volume 293, pp 135-145, 2012.

16. Anup Bhange and Sumit Utareja "Anomaly Detection and Prevention in Network Traffic based on Statistical approach and α-Stable Model", International Journal of Advanced Research in Computer Engineering & Technology, ISSN: 2278 – 1323, Volume 1, Issue 4, pp. 690 – 698, June 2012.