Volume 7, Issue 1, January 2017 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

# A Survey on Attack Detection System using Unsupervised Learning Method

**Nisha A. Bhalke, Rahul L. Paikrao**

Department of Computer Engineering, Amrutvahini College of Engineering, Sangamner, University of Pune, Maharashtra, India

**DOI: 10.23956/ijarcsse/V6I12/0218**

***Abstract— Responsibility and availability of network services area unit being vulnerable by the growing variety of Denial-of-Service (DoS) attacks. Effective mechanisms for DoS attack detection area unit demanded. Completely different systems were planned for detection Dos attacks exploitation machine learning, applied mathematics analysis, data mining, etc. This system is improvement over earlier system in which k- means clustering will be applied over a training samples so that it can categorize the samples into completely different clusters and then it applies statistical methods to seek out the correlation between features to gain information standard deviation, mean and covariance matrix. Applying multivariate correlation analysis on every cluster will help us to get profile parameters according to their cluster and therefore allowing us to understand sharp boundary of characterizing a sample packet. This will help us to reduce false positive rate and to enhance efficiency and accuracy.***

***Keywords— DoS attacks, Intrusion Detection, K-means clustering, Machine Learning, MCA, and Triangle Area Map.***

## I. INTRODUCTION

Network security [1] system is taking part in important role in modern society because of the recent advancement in network connected system. Increasing speed and capability in computational and communication resources notably the recognition on interconnected system of just about all computers has created it vital to secure these systems with high level of security. The network attack intruders have found a big platform to attack on network. Thus it is important to secure interconnected system that contains confidential and sensitive information and is flowing through the network that might be manipulated. Intruders can be developed in various algorithm that build effective use of current oversimplified and networking model of internet. Network security [1] has become additional vital than ever. However, just in case of a complex interconnected network system, it is tough to make sure a secure networking environment Intruders risk system security by crashing services, dynamic or stealing important data. To address these issues more correct intrusion detection system [3] should be developed so as to safeguard this confidential information.

Among the detection strategies employed by IDPS [2], which are:

- Signature Based Detection: This methodology relies on the comparison of the units of activities (Package, Log Entry) to an inventory of models by victimization the operators of comparison. A model corresponds to a known attack.
- Anomaly Based Detection: It's a way basing itself on applied mathematics calculations and it is a "Profile" that represents the normal behavior. Thus this methodology consists of creating comparison between the events and also the definition of the events thought of normal to detect deviations.
- Stateful Protocol Analysis: This methodology compares the protocols and their profiles. Additionally, it exploits the mixture of the request and its answer to be ready to measure the state.

Anomaly based detection system have the advantage of detecting new attacks but their false positive rates are high. [1], [7] Intrusion detection systems have undergone various developments in both power and scope. Machine learning [6] is the ability of a machine that mechanically improves its performance through learning from expertise. Machine learning [6] techniques are used to study normal computer activities and determine abnormal behaviors that deviate from normal as intrusions. Though these anomaly-based IDSs are able to observe narrative attacks, most of them suffer from a high error rate as a result of a deficiency in their discrimination ability. Therefore these detection approaches can only distinguish abnormal attacks, they miss several attacks that don't seem to be considerably completely different from normal behaviors. Detecting such abusive not only provides information on damage assessment, but additionally helps to prevent from new attacks. These all attacks are typically detected by tools said as intrusion detection system.

Intrusion detection is the method of monitoring the events occurring in a computer system or network and analyzing them for doable incidents, that square measure violations or close threats of violation of computer/laptop security policies, acceptable use policies, or commonplace security practices [2]. Intrusion detection by using K-means clustering algorithm [3], [11], [12] is going to be developed to discriminate packets by victimization preformed protocol wise clusters.

## II. LITERATURE SURVEY

Memoona Khanum, Tahira Mahboob, Warda Imtiaz, Humaraia Abdul Ghafoor, Rabeea Sehar [1] describes the survey of methodologies and techniques used for Unsupervised Machine Learning that are used for learn advanced, extremely non-linear models with millions parameters to used great amount of unlabeled information. Deep belief networks (DBNs) and sparse coding are 2 well-known methods of unsupervised learning models. Data clustering distinguishes by the absence of class information. Primarily structure in information is finding in clustering and it has long history in scientific field. K-means is that the most preferred and simple clustering algorithm. This Algorithm was published in 1955. Hierarchical matching pursuit (HMP) for RGB-D information is mentioned. Sparse coding learns hierarchical feature representations from raw RGB-D data in an unsupervised machine learning by using hierarchical matching pursuit. The formal study of learning systems is deduced from Machine learning; that is a field of research or analysis. It has found to be extremely knowledge domain field that acquires and constructs upon ideas from statistics, computer science, optimization itheory, and several other disciplines of science and mathematics.

Alireza Osareh, Bita Shadgar [3] illustrate that network security technology has become crucial in protective government and business computing infrastructure. Modern intrusion detection applications face complicated requirements; they have to be reliable, extensible, straightforward to manage, and have low maintenance amount. In recent years, machine learning-based intrusion detection systems have high accuracy, smart generalization to novel sorts of intrusion, and robust behavior in very ever-changing environment. This work aims to check potential of machine learning strategies in intrusion detection system, as well as artificial neural networks and support vector machine, with the hope of giving reference for establishing intrusion detection system in future. Compared with different connected works in machine learning-based intrusion detectors, we have a tendency to propose to calculate the average value via sampling totally different ratios of normal data for every measurement, that lead us to achieve a far better accuracy rate for observation knowledge in real world. We have tendency to compare the accuracy, detection rate, false alarm rate for four attack types. The in depth experimental results on the KDD-cup intrusion detection benchmark dataset demonstrate that the planned approach produces higher performance than KDD Winner, particularly for U2R and U2L kind attacks.

D. E. Denning, SRI International [4] describes a model of a real-time intrusion-detection knowledgeable system capable of detecting break-ins, penetrations, and alternative forms of computer abuse which is described. The model relies on the hypothesis that security violations is detected by monitoring a system's audit records for abnormal patterns of system usage. The model includes profiles for representing the behavior of subjects with relation to objects in terms of metrics and applied mathematics models, and rules for acquiring knowledge about this behavior from audit records and for detecting abnormal behavior. The model is independent of any specific system, application surroundings, system vulnerability, or form of intrusion, thereby providing a framework for a general intrusion detection knowledgeable system.

Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu, and Ali A. Ghorbani [5] describes that during the last decade, anomaly detection has attracted the eye of the many researchers to beat the weakness of signature-based IDSs in detecting novel attacks and KDD CUP 99 is the largely wide used knowledge set for the analysis of the systems. Having conducted a statistical analysis on this knowledge set, we found two necessary issues that extremely affects the performance of evaluated systems and results in a poor analysis of anomaly detection approaches. To resolve these issues, we have proposed a new data set, NSL-KDD that consists of selected records of the whole KDD data set and doesn't suffer from any of mentioned shortcomings.

P. Garca-Teodoro, J. Daz-Verdejo, G. Maci-Fernndez, and E.Vzquez [6] illustrate that the Internet and computer networks are exposed to an increasing veriety of security threats. With new forms of attacks showing continually, developing versatile and adaptive security familiarized approaches is a severe challenge. During this context, anomaly-based network intrusion detection techniques are a valuable technology to safeguard target systems and networks against malicious activities. However, despite the range of such ways described within the literature in recent years, security tools incorporating anomaly detection functionalities are simply starting to appear and number of important issues stay to be solved.

Roshan Chitrakar, Huang Chuanhe [7] describes the role of Intrusion Detection System (IDS) has been inevitable within the area of data and Network Security - particularly for building an honest network defense infrastructure. Anomaly based intrusion detection technique is one among the building blocks of such a foundation. The attempt has been created to use hybrid learning approach by combining k-Medoids based clustering method followed by Naïve Bayes classification technique. As a result of the fact that k-Medoids clustering techniques represent the real world situation of information distribution, the projected increased approach will group the complete information into corresponding clusters more accurately than k-Means such that it results in a higher classification. An experiment is administrated in order to evaluate performance, accuracy, detection rate and false positive rate of the classification scheme. Results and analyses show that the projected approach has increased.

Shi Na, Liu Xumin, Guan yong [8] explain the clustering analysis technique which is one of the main analytical techniques in data mining, the technique of clustering algorithm will influence the clustering results directly. It discusses the standard k-means clustering algorithm and analyzes the shortcomings of standard k-means algorithm, like the k-means clustering algorithm has to calculate the distance between every information object and all cluster centers in each iteration that makes the efficiency of clustering isn't high. It proposes an improved k-means algorithm in order to solve this question, requiring an easy data structure to store some data in each iteration that is to be employed in the next iteration. The improved technique avoids computing the distance of every information object to the cluster centers repeatedly, saving the period of time. Experimental results show that the improved technique can effectively improve the speed of clustering and accuracy, reducing the computational complexity of the k-means.

Zhiyuan Tan, Aruna Jamdagni, Xiangjian, Priyadarsi Nanda, Ren Ping Liu [9] place associate degree interconnected systems, like web servers, database servers, and cloud computing servers and so on, are currently under threads from network attackers. Collectively of commonest and aggressive suggests that, denial-of-service (DoS) attacks cause serious impact on these computing systems. A DoS attack detection system that uses multivariate correlation analysis (MCA) technique for accurate network traffic characterization by selecting the geometrical correlations between network traffic choices. MCA-based DoS attack detection system uses the principle of anomaly based detection in attack recognition. This makes answer capable of detecting known and unknown DoS attacks effectively by learning the patterns of legitimate network traffic. A triangle-area-based technique is planned to enhance and to speed up the process of MCA. The effectiveness of planned detection system is evaluated using KDD Cup 99 data set and the influences of each non-normalized data and normalized data on the performance of the planned detection system are examined. The results show that system outperforms two different previously developed state-of-the-art approaches in context of detection accuracy.

K. Lee, J. Kim, K. H. Kwon, Y. Han, and S. Kim [10] describes the Distributed Denial of Service (DDoS) attacks that generates enormous packets by a large number of agents and can simply exhaust the computing and communication resources of a victim in a brief amount of time. Proposed method for proactive detection of DDoS attack by exploiting its design that consists of the choice of handlers and agents, the communication and compromise, and attack. First it will check into the procedures of DDoS attack and then select variables based on these options. After that, cluster analysis is being performed for proactive detection of the attack. Experiment with 2000 DARPA Intrusion Detection Scenario Specific Data Set in order to use the technique. The results show that every section of the attack scenario is divided well and that can detect precursors of DDoS attack and the attack itself.

K. Krishna and M. Narasimha Murty [11] plans a unique hybrid genetic algorithm (GA) that finds a globally best partition of a given data into a specified number of clusters. GA's used earlier in clustering employ either an upscale crossover operator to create valid child chromosomes from parent chromosomes or a costly fitness function or both. To avoid these costly operations, use of GA with a classical gradient descent algorithm which is used in clustering, viz. K-means algorithm. Thus, the name genetic K-means algorithm (GKA). They defined K-means operator, one-step of K-means algorithm and use of it in GKA as a search operator rather than crossover. Also defined a biased mutation operator specific to clustering called distance-based-mutation. It is observed within the simulations that GKA converges to the simplest known optimum like the given information in concurrence with the convergence result. It is also observed that GKA searches quicker than number of the opposite evolutionary algorithms used for clustering.

L. Ertoz, M. Steinbach, and V. Kumar [12] illustrate the clustering technique which depends on density and distance, however these ideas become tougher to define as dimensionality increases. Definitions of density and similarity which work well for high dimensional data (actually, for information of any dimensionality). Especially, they use a similarity measure which is supported amount of neighbors that two points share and define the density of a point because the total of the similarities of a point's nearest neighbors. Then define a new clustering algorithm which relies on these ideas. This algorithm eliminates noise (low density points) and creates clusters by associating non-noise points with representative or core points (high density points). This approach handles several issues that historically plague clustering algorithms, e.g., finding clusters within the presence of noise and outliers and finding clusters within the data that has clusters of various shapes, sizes, and density. They used clustering algorithm on a variety of high and low dimensional data sets with better results.

## III. CONCLUSIONS

In this means we've studied the existing or different approaches for detecting attack system. The multivariate correlation analysis based denial of service attack finding system that is powered by a triangle area based MCA method and anomaly-based finding strategies. We are going to use k-means clustering to cluster the packets according to their protocols. Here we are going to be used KDD 99 Dataset that will be used as input to our system that is Standard Dataset which provides effectiveness and performance of intrusion detection system. The latter technique facilitates our system to be able to differentiate each known and unknown attacks from correct network traffic.

## ACKNOWLEDGMENT

I place a deep sense of appreciation to my Project Guide Mr. Rahul Paikrao giving me all possible facilities and suggestions to convey this survey paper a perfect shape. We would like to thank all the authors of various research papers referred during study of this paper. It was absolutely knowledge gaining and useful for the further analysis to be done in future.

## REFERENCES

[1] Memoona Khanum, Tahira Mahboob, Warda Imtiaz, Humaraia Abdul Ghafoor, Rabeea Sehar, "*A Survey on Unsupervised Machine Learning Algorithms for Automation, Classification and Maintenance*", International Journal of Computer Applications 119(13):34-39, June 2015.

[2] Youssef Senhaji, Hicham Medromi, "*Network Security:Hybrid IDPS*", International Journal of Applied Information Systems (IJAIS) ISSN : 2249-0868 Foundation of Computer Science FCS, New York, USA.

[3] Alireza Osareh, Bita Shadgar, "*Intrusion Detection in Computer Networks based on Machine Learning Algorithms*", IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.11, November 2008.

[4] D. E. Denning, "*An Intrusion-detection Model*", IEEE Transactions on Software Engineering, pp. 222-232, 1987.

[5] Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu, and Ali A. Ghorbani, "*A Detailed Analysis of the KDD CUP 99 Data Set*", Proceedings of the 2009 IEEE Symposium on Computational Intelligence in Security and Defense Applications (CISDA 2009).

[6] P. Garca-Teodoro, J. Daz-Verdejo, G. Maci-Fernndez, and E.Vzquez, "*Anomaly-based Network Intrusion Detection: Techniques, Systems and Challenges*", Computers Security, vol. 28,pp. 18-28, 2009.

[7] Roshan Chitrakar, Huang Chuanhe, "*Anomaly based Intrusion Detection using Hybrid Learning Approach of combining k-Medoids Clustering and Nave Bayes Classification*".

[8] Shi Na, Liu Xumin, Guan yong, "*Research on k-means Clustering Algorithm*", Third International Symposium on Intelligent Information Technology and Security Informatics.

[9] Zhiyuan Tan, Aruna Jamdagni, Xiangjian, Priyadarsi Nanda, Ren Ping Liu, "*A System for Denial-of-Service Attack Detection Based on Multi- variate Correlation Analysis*", IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS VOL:25 NO:2 YEAR 2014.

[10] K. Lee, J. Kim, K. H. Kwon, Y. Han, and S. Kim, "*DDoS attack detection method using cluster analysis*", Expert Systems with Applications, vol. 34, no. 3, pp. 1659-1665, 2008.

[11] K. Krishna and M. Narasimha Murty, "*Genetic K-means algorithm*", IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 29, no. 3, pp. 433439, Jun. 1999.

[12] L. Ertoz, M. Steinbach, and V. Kumar, "*A new shared nearest neighbor clustering algorithm and its applications*", in Proc. Workshop Clustering High Dimensional Data Appl., 2002, pp. 105115.