Integration of Signature Based and Anomaly Based Detection

[1] Prerika Agarwal, [2] Sangita Rani Satapathy

[1] M.Tech Scholar, Dept. of Computer Science
Ajay Kumar Garg Engineering College, Ghaziabad, India

[2]Assistant Professor, Dept. of Computer Science
Ajay Kumar Garg Engineering College, Ghaziabad, India

Abstract- As the technology is advancing, there are risks of information to be available to the malicious users while providing it to normal users and the possibility of attack is also increasing in that ratio. An intrusion detection system is required for securing network. Signature-based detection is used for detecting known attacks as many attacks have distinct signatures. An anomaly-based IDS tries to find suspicious activity on the system. Clustering is suitable for anomaly detection, since no knowledge of the attack classes is needed whilst training. In this paper a survey has been done on anomaly detection techniques and clustering. It also consists idea to our research of integrating Snort with Clustering Algorithm for anomaly detection.

Key words- *Intrusion Detection System, Signature based Detection, Anomaly based Detection, Hierarchical Clustering, Snort.*

1. Introduction

As the technology is advancing, there are risks of information to be available to the malicious users while providing it to normal users. Therefore, there needs to be some kind of security to the organization's private resources from the Internet as well as from inside users. Most of the attacks happen from inside users for the very fact that they know the systems much more than an outsider knows and access to information is easier for an insider. An intrusion detection system is required for securing network.

An Intrusion detection system (IDS) is a security system that monitors computer systems and network traffic and analyzes that traffic for possible hostile attacks originating from outside the organization and also for system misuse or attacks originating from inside the organization.

Intrusion detection technology can be divided into two categories:

- Signature based detection
- Anomaly detection

Signature-based Intrusion Detection Systems (IDS) references a database of previous attack signatures and known system vulnerabilities. It is mainly used of most commercial intrusion detection systems, by matching the current data and signature-known type of attack found.
While **Anomaly-based Intrusion Detection Systems (IDS)** references a baseline or learned pattern of normal system activity to identify active intrusion attempts. Deviations from this baseline or pattern cause an alarm to be triggered. Anomaly detection system compared with the current activities of history normal activities profile, which can detect unknown types of attacks.

IDSs based on the anomaly detection method use normal patterns to detect abnormal activities from observed data. They typically attempt to identify deviations from predefined normal patterns, and regard them as potential attacks. The former can detect well-known attacks at relatively higher accuracy than the latter, but they have a fatal weakness in that they cannot detect unknown attacks that are not matched to any predefined signatures. In anomaly detection normal behavior of users or the protected system is modeled, often using machine learning or data mining techniques. During detection new data is matched against the normality model, and deviations are marked as anomalies. Since no knowledge of attacks is needed to train the normality model, anomaly detection may detect previously unknown attacks.

Clustering is a technique for training of the normality model, where similar data points are grouped together into clusters using a distance function. Clustering is suitable for anomaly detection, since no knowledge of the attack classes is needed whilst training. Clustering is available in flavors of i) Hierarchical, and ii) Partition. In hierarchical clustering the data are not partitioned into a particular cluster in a single step. Instead, a series of partitions takes place, which may run from a single cluster containing all objects to n clusters each containing a single object. Hierarchical Clustering is subdivided into agglomerative methods, which proceed by series of fusions of the n objects into groups, and divisive methods, which separate n objects successively into finer groupings.

Traditional signature-based intrusion detection techniques use patterns of well known attacks to match and identify known intrusions. The main drawback of these techniques is the inability to detect the newly invented attacks. Limitation of signature-based IDS is failure to identify novel attacks, and sometimes even minor variations of known patterns.

To obtain sufficient information about complex network traffic and compensate for the weaknesses of traditional *Intrusion Detection Systems* (IDS), *Anomaly Detection Algorithms* (ADA) were used. Theses algorithms can be employed as a useful mechanism to analyze network anomalies and detect misbehaviors issued by users, or even unknown signature viruses and worms. Anomaly detection has an advantage over signature-based detection in that a new attack for which a signature does not exist can be detected if it falls out of the normal traffic patterns.
But there is also drawback of anomaly detection as it suffer high false detection rate. Thus both algorithms which are signature based and anomaly based, to be combined, in order to improve the detection of new malicious packet and reduce excessive false alarm rate.
This paper provides a survey on various anomaly detection techniques. It is also introducing idea of integration of signature based and anomaly based detection using Snort and Clustering Algorithm.

2. Motivation

Toward a more practical unsupervised anomaly detection system

J. Song et al. [10] have proposed a new anomaly detection method by which they are able to optimize the values of two parameters, i.e., *a* (no. of attack data) and *k* (no. of normal pattern), without predefining them. Among the three parameters, they have focused on only two *a* and *k*, because they have proven that parameter b barely affects the performance of the previous anomaly detection method as long as a reasonable range of value is chosen. The proposed method let them construct intrusion detection models based on anomaly detection without tuning the parameters, and this advantage can thus contribute to more practical operations in the real environment.

A clustering-based method for unsupervised intrusion detections

S. Jiang et al. [8] proposes a novel method for calculating cluster radius threshold. It defines the outlier factor of cluster to measure the degree of a cluster deviating from the whole where anomalous classes can be distinguished from normal ones. It obtained an improved nearest neighbor (INN) method for classifying data and a novel strategy for detecting intrusion. INN considers not only the candidate classified object and its nearest neighbor in model, but also the distance between them. It is capable of detecting unknown intrusions.

A novel intrusion detection system based on hierarchical clustering and support vector machines

S.J. Horng et al.[7] proposed an SVM-based intrusion detection system based on a hierarchical clustering algorithm to preprocess the KDD Cup 1999 dataset before SVM training. The hierarchical clustering algorithm was used to provide a high quality, abstracted, and reduced dataset for the SVM training, instead of the originally enormous dataset. Thus, the system could greatly shorten the training time, and also achieve better detection performance in the resultant SVM classifier.

Adaptive Real-Time Anomaly Detection with Fast Indexing and Ability to Forget

K. Burbek et al.[4] presented Anomaly Detection With fast Incremental Clustering (ADWICE), a novel adaptive anomaly detection scheme, inspired by the BIRCH clustering algorithm, and extended with new capabilities. It is based on an earlier presented version of the algorithm but complements the original technique with a novel search index that increases detection quality, and provides new means to handle adaptation (forgetting). It showed the application of the new mechanisms in two settings: (1) an emulated test network at a major Telecom operator in Europe for evaluating the scalability and timeliness of the algorithm, and (2) the comparative analysis of the detection quality of the algorithm based on the only common (open) data source available – the KDD99 attack data.

Adaptive real-time anomaly detection with incremental clustering

K. Burbek et al.[3] explained the role of anomaly detection in a distributed architecture for agents that has been developed within the European Safeguard project. It applies clustering as the technique for training of the normality model, where similar data points are grouped together into clusters using a distance function. Clustering is suitable for anomaly detection, since no knowledge of the attack classes is needed whilst training. Contrast this to other learning approaches, e.g. classification, where the classification algorithm needs to be presented with both normal and known attack data to be able to separate those classes during detection.

ADWICE – Anomaly Detection with Real-Time Incremental Clustering

K. Burbek et al.[2] developed a new indexing mechanism for ADWICE (Anomaly Detection With fast Incremental Clustering), an adaptive anomaly detection scheme inspired by the BIRCH clustering algorithm. This paper does not attempt to justify the quality of clustering within anomaly detection or compare performance with other machine learning work.

Y-MEANS: a clustering method for intrusion detection

Y. Guan et al. proposed a K-means based clustering algorithm, named Y-means, for intrusion detection. Y-means overcomes two shortcomings of K-means: *number of clusters dependency* and *degeneracy*. A data set can be partitioned into an appropriate number of clusters automatically. This is one of the advantages of the Y-means algorithm for intrusion detection.
Another advantage is that the raw log data of information systems can directly be used as training data without being manually labeled.

3. Our Contribution

The ensemble behavior of the network can be characterized by the means of two main approaches: the first is inference of the overall network behavior and the second is to analyze behavior of the individual entities or nodes. The approaches used to address the anomaly detection problem depend on the nature of the data that is available for the analysis. Network data can be obtained at multiple levels of granularity such as network-level or end-user-level.

The main objective is to combine signature-based algorithm and anomaly detection algorithm to improve the detection of new malicious packet and reduce excessive false alarm rate. This objective will be achieved by integrating Snort (Signature-based tool) with hierarchical clustering algorithm (for Anomaly Detection).A signature-based IDS analyzes the network traffic looking for patterns that match a library of known signatures. The signatures are composed by many elements that identify traffic. They usually examine the network traffic with predefined signatures and each time database is updated. An example of Signature based Intrusion Detection System is SNORT.

An anomaly-based IDS tries to find suspicious activity on the system. These systems are not unlike virus detection systems -- they can detect many or all *known* attack patterns, but they are of little use for as yet unknown attack methods. Anomaly detection commonly used data mining or machine learning methods. Clustering, one method of data mining is paid attention to in the study which is based on unsupervised.

3.1 SNORT

Snort is a signature-based IDS that allows to monitor the status of a network. It is operated in various aspects with sniffers, because Snort analyzes all the network traffic looking for any type of intrusion. Snort is an open source network intrusion prevention and detection system. It is available under GPL, is free and runs under Windows and GNU/Linux. It implements a detection engine that allows registering, warning and responding to predefined attack.

3.2 Anomaly Detection

Anomaly detection techniques assume that all intrusive activities are necessarily anomalous. This means that if we could establish a "normal activity profile" for a system, we could flag all system states varying from the established profile by statistically significant amounts as intrusion attempts. The concept behind misuse detection schemes is that there are ways to represent attacks in the form of a pattern or a signature so that even variations of the same attack can be detected.

Anomaly detection systems are also computationally expensive because of the overhead of keeping track of, and possibly updating several system profile metrics. ADWICE is a hierarchical clustering to deal effectively with noise of intrusion detection, which can dynamic clusters profile, and stores compressed information using tree structure, thereby reducing the time and space consumption. The algorithm generation system normal action profile by the network data source, with the profile of the monitoring system has achieved a higher detection rate. However, there is a relatively high rate of false positives due to too much noise in the network data.

4. Conclusion

Security is a big concern for all networks in today's enterprise environment. Intruders have made many successful attempts in bringing down networks and web services. Signature based IDS are reliable when receives pattern matching with library of predefined signatures. Anomaly based IDS are able to detect unknown attacks, but producing number of false alarms. Snort is a powerful tool, capable of performing real time traffic analysis and packet logging. Clustering Algorithm forms normal behavior profile on the audit records and adjust the profile timely as the program behavior changed.

References

[1] Burbeck K, Nadjm-Tehrani S.: ADWICE – anomaly detection with fast incremental clustering. In: Proceedings of the seventh international conference on security and

cryptology (ICICS'04). Springer Verlag; December 2004.

[2] Burbeck. K, Nadjm-Tehrani S.: ADWICE – Anomaly Detection with Real-Time Incremental Clustering ICISC 2004, LNCS 3506, pp. 407–424, 2005. Springer-Verlag Berlin Heidelberg 2005

[3] Burbeck K. Adaptive real-time anomaly detection for safeguarding critical networks. Linko¨ping University, ISBN 91-85497-23-1; February 2006

[4] Burbeck. K, Nadjm-Tehrani S.: Adaptive real-time anomaly detection with incremental clustering, Information Security Technical Report, 1363-4127 Elsevier 2007 Ltd.

[5] Chen.Z, Zhu.D.*:* Hierarchic Clustering Algorithm used for Anomaly Detecting, Advanced in Control Engineering and Information Science, Procedia Engineering 15 (2011) 3401-3405, Elsevier 2011

[6] Guan, Y., Ghorbani, A.A., Belacel, N.: Y-means: A clustering method for intrusion detection. In: Canadian Conference on AI. Volume 2671 of Lecture Notes in Computer Science., Montreal, Canada, 616- 617, Springer (2003)

[7] Horng.S.J, Su.M.Y,Chen.Y.H, Kao.T.W, Chen.R.J, La.J.J,Perkasa.C.D.: A novel intrusion detection system based on hierarchical clustering and support vector machines Expert Systems with Applications 38 (2011) 306–313, Elsevier 2011

[8] Jiang. S.Y, Song.X, Wang.H, Han.J.J, Li.Q.H.: A clustering-based method for unsupervised intrusion detections. Pattern Recognition Letters 27 802–810, Elsevier 2006

[9] Portnoy L, Eskin E, Stolfo S.: Intrusion detection with unlabeled data using clustering. In: ACM workshop on data mining applied to security; November 2001.

[10] S.Jungsuk, Takakura.H, Okab.Y, Nakao.K.: Toward a more practical unsupervised anomaly detection system, Inform. Sci. 1345-1356 Elsevier 2011 Ltd.

[11] Zhang T, Ramakrishnan R, Livny M.: BIRCH: an efficient data clustering method for very large databases. In: SIGMOD record 1996 ACM SIGMOD international conference on management of data, vol. 25(2); 4–6 June 1996. p. 103–14.

[12] C. Chang and C. Lin, "LIBSVM: a library for support vector machines," 2001. Available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[13] KDD Cup 1999. Available on: http://kdd.ics.uci.edu/databases/kddcup 99/kddcup99.html, Ocotber 2012

[14] Nsl-kdd data set for network-based intrusion detection systems." Available on: http://nsl.cs.unb.ca/NSL-KDD/, March 2009.

Prerika Agarwal did her schooling from Seventh Day Adventist Sr. Sec. School, Hapur. She completed her B.Tech in Information Technology from Shobhit Universty. She is perusing M.Tech in Computer Science from AKGEC. She has published four research papers in international journal and international and national conferences. Her areas of interest are Network Security, Data Mining.

Sangita Rani Satapahty has been working as a Asst. Professor, department of CSE, AKGEC, Ghaziabad. She has 6 years of teaching experience. She has published/ presented several papers in journals/ conference of repute. His research interest includes Data Mining and Algorithm.