International Journal of Computer and Internet Security.
ISSN 0974-2247 Volume 5, Number 1 (2013), pp. 1-9

http://www.irphouse.com

Adaptive Anomaly Detection for Network Security

Kamini Nalavade[#1] and B.B. Meshram[*2]

[1]Research Scholar, [2]Professor
[#]Computer Technology Department, V.J.T.I., Matunga, Mumbai
[1]kamini.nalavade@live.com, [2]bbmeshram@vjti.org.in

Abstract

Intrusion detection is an integral part of computer security. It improves the security of information systems by allowing the review of patterns of access in order to discover abnormal activity of users and serving as a deterrent to users attempts to bypass system privilege or protection mechanisms. Anomaly detection systems, a subset of intrusion detection systems, model the normal system/network behavior which enables them to be extremely effective in finding and foiling both known as well as unknown or "zero day" attacks. Anomaly detection is an important problem that has been researched within diverse application domains and many anomaly detection techniques have been specifically developed in past years. This paper is an attempt to provide a structured and comprehensive overview of research on anomaly detection techniques. The different aspects and approaches for anomaly detection are described. We hope that this survey will provide a better understanding of the different directions in which anomaly detection research has been done.

Keywords— Intrusion, Anomaly, Analysis, Security, Machine Learning

Introduction

With the tremendous growth of network-based services and sensitive information on networks, network security is getting more and more importance than ever. Intrusion poses a serious security risk in a network environment. An intrusion is a sequence of events that deliberately try to cause harm such as accessing unauthorized information or manipulating such information. Detecting either failed or successful attempts to compromise the system is called an Intrusion Detection. Detecting the intrusions and preventing the possible attacks

is a critical aspect of computer based system security. Intrusion detection factors related to anomaly detection. Intrusion detection refers to a broad range of approaches that detect malicious attacks on computers and networks. These approaches can be categorized into misuse detection and anomaly detection.

Misuse detection or rule based detection uses pattern matching. To detect attacks it compares network traffic to known attack patterns called signatures. Misuse detectors can be successful in known attacks but fail to detect unknown patterns. When a new attack is found, signature is required to be constructed and misuse detector is required to be reconfigured [9].

An anomaly is something that is different from the normal or that cannot be classified. Anomaly detection or profile based detection creates a profile system that flags any events that differs from a normal pattern and passes this information to output routines. An anomaly detector looks for deviation from normal behaviour. When deviation exceeds a threshold an alarm is raised. Anomaly detectors are able to detect previously unseen attack but they suffer from high false alarm rate. False alarm rate is high because some behaviour may be rare but legitimate. [9]

This paper presents the literature review completed for the research work on anomaly detection. Section II presents the various aspects and terminologies related to anomaly detection. Section III describes different approaches for anomaly detection. In the next section

Aspects of anomaly detection

An important aspect of an anomaly detection technique is the nature of the desired anomaly. Anomalies can be classified into following three categories: anomalies also known as outliers, exceptions or peculiarities are patterns in data that do not conform to a well defined notion of normal behaviour of a system

Point Anomalies: If an individual data instance can be considered as anomalous with respect to the rest of data, then the instance is termed a point anomaly. This is the simplest type of anomaly and is the focus of majority of research on anomaly detection.

Contextual Anomalies: If a data instance is anomalous in a specific context, but not otherwise, then it is termed a contextual anomaly

Collective Anomalies: If a collection of related data instances is anomalous with respect to the entire data set, it is termed a collective anomaly. Anomaly based schemes fall into three main categories: behavioural, traffic pattern and protocol. Behavioural anomalies look for anomalies in the types of behaviour that have been statistically base lined. Traffic-pattern analysis looks for specific patterns in network traffic. Protocol analysis looks for network protocol violations. Protocol analysis has the benefit of identifying possible attacks that are not yet identified.

A key aspect of any anomaly detection technique is the nature of the input data. Input is generally a collection of data instances. Each data instance can be described using a set of attributes. The attributes can be of different types

such as binary, categorical, or continuous. Each data instance might consist of only one attribute (univariate) or multiple attributes (multivariate). In the case of multivariate data instances, all attributes might be of same type or might be a mixture of different data types[6]. The nature of attributes determines the applicability of anomaly detection techniques. For example, for statistical techniques different statistical models have to be used for continuous and categorical data. Similarly, for nearest-neighbor-based techniques, the nature of attributes would determine the distance measure to be used. Often, instead of the actual data, the pair wise distance between instances might be provided in the form of a distance or similarity matrix. In such cases, techniques that require original data instances are not applicable, for example, many statistical and classification-based techniques. Input data can also be categorized based on the relationship present among data instances. Most of the existing anomaly detection techniques deal with record data or point data, in which no relationship is assumed among the data instances. An important aspect for any anomaly detection technique is the manner in which the anomalies are reported. Typically, the outputs produced by anomaly detection techniques are one of the following two types:

- *Scores.* Scoring techniques assign an anomaly score to each instance in the test data depending on the degree to which that instance is considered an anomaly. Thus the output of such techniques is a ranked list of anomalies. An analyst may choose to either analyze the top few anomalies or use a cutoff threshold to select the anomalies.
- *Labels.* Techniques in this category assign a label (*normal* or *anomalous*) to each test instance.

Approaches for anomaly detection

Several anomaly detection techniques based on machine-learning or statistics based approaches have been developed for the purpose of network security. In this section we will overview statistical approach and machine learning approach for anomaly detection. Statistical approaches attempt to define normal or expected behaviour, whereas rule base approaches attempt to define proper behaviour. Statistical methods monitor the user or system behaviour by measuring certain variables over time (e.g. login and logout time of each session in intrusion detection domain). The basic models keep averages of these variables and detect whether thresholds are exceeded based on the standard deviation of the variable. More advanced statistical models also compare profiles of long-term and short-term user activities. Statistical anomaly detection is effective against masquerades that are unlikely to mimic the behaviour patterns of the accounts they appropriate. On the other hand, such techniques may be unable to deal with illegitimate users.

Statistical Anomaly Detection falls into two categories: threshold detection and profile based systems. Profile based anomaly detection focuses on characterising the past behaviour of individual users or related groups of users and then detecting significant deviations. A profile may consist of a set of

parameters, so that deviation on just a single parameter may not be sufficient in itself to signal an alert. Audit records serve to define typical behaviour. The intrusion detection model analyses incoming audit records to determine deviation from average behaviour. Using these general metrics the statistical test like mean and standard deviation of a parameter can be conducted. This gives a reflection of the average behaviour and its variability. A multivariate model is based on correlations between two or more variables. Intruder behaviour may be characterized with greater confidence by considering such correlations. A markov process model is used to establish transition probabilities among various states. A time series model focuses on time intervals, looking for sequences of events that happen too rapidly or too slowly. An operational model is based on judgment of what is considered abnormal rather than an automated analysis of past audit records. Intrusion is suspected for an observation outside limits.

In contrast to statistical techniques, machine learning techniques are well suited to learning patterns with no prior knowledge of what those patterns may be. Machine learning is the study of computer algorithms that improve automatically through experience. Applications range from data mining programs that discover general rules in large datasets to information filtering systems that automatically learn user's interest.Clustering and classification are the two most popular machine learning techniques.

Anomaly detection techniques can operate in one of the following modes

Supervised Anomaly Detection

Supervised anomaly detection techniques require a data set that has been labeled as "normal" and "abnormal" and involves training a classifier (the key difference to many other statistical classification problems is the inherent unbalanced nature of outlier detection). A typical approach in such cases is to build a predictive model for normal vs. anomaly classes. Any unseen data instance is compared against the model to determine which class it belongs to. There are two major issues that arise in supervised anomaly detection. First, the anomalous instances are far fewer compared to the normal instances in the training data. Second, obtaining accurate and representative labels, especially for the anomaly class is usually challenging. The other main issue with the supervised anomaly detection is building predictive models. The classifier has to be trained with labeled patterns to be able to classify new unlabeled patterns. The given labeled training patterns are use to learn the description of classes. Some supervised methods include support vector machines, neural network and genetic algorithms among others.[2]

Unsupervised Anomaly Detection

Unsupervised anomaly detection techniques detect anomalies in an unlabeled test data set under the assumption that the majority of the instances in the data set are normal by looking for instances that seem to fit least to the remainder of the data set. Techniques that operate in unsupervised mode do

not require training data, and thus are most widely applicable. The techniques in this category make the implicit assumption that normal instances are far more frequent than anomalies in the test data. If this assumption is not true then such techniques suffer from high false alarm rate. Such adaptation assumes that the test data contains very few anomalies and the model learned during training is robust to these few anomalies. Data clustering is very useful when little priori information about the data is available. Clustering methods can be classified into two categories: hierarchical clustering algorithms and partitional clustering algorithms.

Machine Learning for Anomaly detection

Several anomaly detection techniques are proposed in literature some of the popular techniques are:

- Distance based techniques (k-nearest neighbor, Local Outlier Factor).
- One Class Support Vector Machines.
- Replicator Neural Networks.
- Cluster analysis based outlier detection.
- Pointing at records that deviate from association rules

We will see One class Support Vector Machines and Cluster Analysis based outlier detection in detail.

Support vector machines for anomaly detection

In supervised learning, learning is based on labelled training data while in unsupervised learning, the training data is unlabeled. Unsupervised learning is beneficial for intrusion detection domain as unlabeled data can be obtained easily from audit records and log files. SVM is basically two class based and supervised learning while it is adapted to one class SVM. In one class SVM noises in positive data also called as outliers is used as negative examples. The basic idea of the one class SVM is to map the input data into a high dimensional feature space using an appropriate kernel function and constructs a decision function to best separate one class data from the second class data with the maximum margin.The one class SVM can be formulated as follows:

f(x) = +1 if x € S and

f(x) = -1 if x € S

In our context, let $x1, x2, \ldots, xl$ be training examples belonging to one class X, where X is a compact subset of R^N. Let $\Phi : X \longrightarrow H$ be a kernel map which transforms the training examples to another space. The kernel functions can be used at the time of training of the classifiers which selects support vectors along the surface of this function.

An SVM maps linear algorithms into non-linear space. It uses a feature called, kernel function, for this mapping. Kernel functions like polynomial, radial basis function are used to divide the feature space by constructing a

hyperplane. The kernel functions can be used at the time of training of the classifiers which selects support vectors along the surface of this function. SVM classify data by using these support vectors that outline the hyperplane in the feature space. This process will involve a quadratic programming problem, and this will get a global optimal solution. Suppose we have N training data points {(x1, y1), (x2,y2), (x3, y3), ..., (xN , yN)}, where xi € Rd and yi €{+1,−1}. Consider a hyper-plane defined by (w, b), where w is a weight vector and b is a bias. The classification of a new object x is done with

$$f(x) = \text{sign}(w \cdot x + b) = \text{sign}\left(\sum_{i}^{N} \alpha_i y_i (x_i \cdot x) + b\right)$$

The training vectors xi occur only in the form of a dot product. For each training point, there is a Lagrangian multiplier αi. The Lagrangian multiplier values αi reflect the importance of each data point [2]. When the maximal margin hyper-plane is found, only points that lie closest to the hyper-plane will have αi > 0 and these points are called support vectors. All other points will have αi = 0. That means only those points that lie closest to the hyperplane, give the representation of the hypothesis/classifier. These data points serve as support vectors. Their values can be used to give an independent boundary with regard to the reliability of the hypothesis/classifier.

Cluster analysis based outlier detection

The K-means clustering is a classical clustering algorithm. After an initial random assignment of example to K clusters, the centres of clusters are computed and the examples are assigned to the clusters with the closest centres. The process is repeated until the cluster centres do not significantly change. Once the cluster assignment is fixed, the mean distance of an example to cluster centres is used as the score. Using the K-means clustering algorithm, different clusters were specified and generated for each output class. K-means clustering is a well known Data Mining algorithm that has been used in an attempt to detect anomalous user behaviour, as well as unusual behaviour in network traffic. There are two problems that are inherent to K-means clustering algorithms. The first is determining the initial partition and the second is determining the optimal number of clusters. The figure7 as shown below depicted the K-means algorithm. As the algorithm iterates through the training data, each cluster's architecture is updated. In updating clusters, elements are removed from one cluster to another. The updating of clusters cause the values of the centroids to change[3]. This change is a reflection of the current cluster elements. Once there are no changes to any cluster, the training of the K-Means algorithm is complete.

Algorithm:

1. Initialize K clusters (randomly select k elements from the data)
2. While cluster structure changes, repeat from 2.
3. Determine the cluster to which source data belongs Use Euclidean distance formula. Add element to cluster with min (Distance (xi, yj)).
4. Calculate the means of the clusters.
5. Change cluster centroids to means obtained using Step 3.

At the end of the K-Means training, the K cluster centroids are created and the algorithm is ready for classifying traffic. For each element to be clustered, the cluster centroids with the minimal Euclidean distance from the element will be the cluster for which the element will be a member.

Analysis Model of anomaly detection

There are many possible data-analysis schemes for an analysis engine but we can basically break down intrusion analysis in four steps.

1. **Preprocessing** The first step is to collect data about intrusions from sensors. The data are put into a numeric form and is then formatted. The information is then classified into statistical profile that is based on different algorithms in the knowledge base.
2. **Analysis** The event data are typically reduced to a profile vector, which is then compared to historical records for that particular user. Any data that fall outside the baseline of normal activity is labelled a deviation.
3. **Response** if the event matches any pattern of attack, the analysis engine sends an alert. In anomaly detector response is triggered automatically or generated manually.
4. **Refinement** The data records must be kept updated. The profile vector history will typically be deleted after a specific number of days. [9]

The basic model for anomaly detection is shown as figure 1. The model explains the basic steps required for anomaly detection.

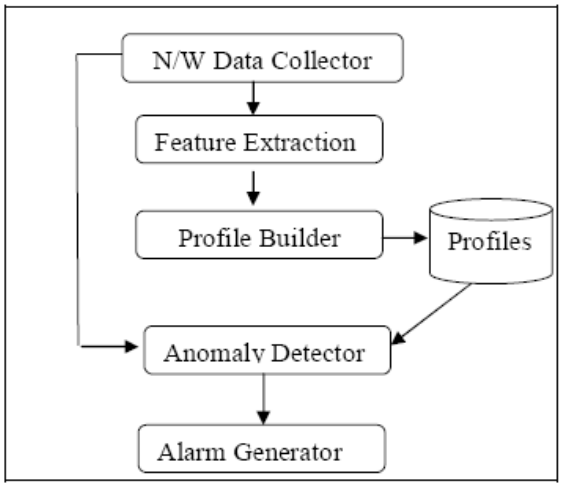

Figure 1. Anomaly Detection Model

Data collector is a network sensor which generates the unlabeled input data for processing. Feature extraction is an step in the processing of raw network traffic or audit data in order to apply data mining techniques to it. Most machine learning algorithms are designed to process data in a feature vector format, where each record consists of a set of feature values and possibly a label for training and/or evaluation purposes. Careful feature selection is critical for improving the performance of machine learning algorithms. Anomaly detectors are based on profiles. Creating profiles of normal behaviour is essential step of anomaly detection. the profiles are updated when new knowledge about system behavior is available. This part of the system handles all output from the intrusion detection system. The output may be either an automated response to an intrusion or a suspicious activity alert for a system security officer.

Conclusion

In the last many years, intrusion detection systems have slowly evolved from host and operating system specific applications to distributed systems that involve a wide array of operating systems. The challenges that lie ahead for the next generation of intrusion detection systems and, more specifically, for anomaly detection systems are many. In this paper, we have provided a comprehensive survey of anomaly detection systems and the different approaches. We have also discussed different technological trends in anomaly detection and identify open problems and challenges in this area. While anomaly detection systems are attractive conceptually, a host of technological problems need to be overcome before they can be widely adopted. These problems include: high false alarm rate, failure to scale to gigabit speeds, etc[10]. We hope that this comprehensive survey on anomaly detection should allow a reader to not only understand the motivation behind using a particular anomaly detection technique, but will also provide a comparative analysis of various techniques.

References

[1] Hans-Peter Kriegel, Peer Kröger, Arthur Zimek (2009). "Outlier Detection Techniques (Tutorial)". *13th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2009)* (Bangkok, Thailand). http://www.dbs.ifi.lmu.de/Publikationen/Papers/tutorial_slides.pdf. Retrieved 2010-06-05.

[2] Varun Chandola, Arindam Banerjee, and Vipin Kumar, Anomaly Detection: A Survey, ACM Computing Surveys, Vol. 41(3), Article 15, July 2009

[3] Ivan Tomek (1976). "An Experiment with the Edited Nearest-Neighbor Rule". *IEEE Transactions on Systems, Man and Cybernetics*. pp. 448-452. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4309523&tag=1.

[4] Michael R Smith and Tony Martinez (2011). "Improving Classification Accuracy by Identifying and Removing Instances that Should Be Misclassified". *Proceedings of International Joint Conference on Neural Networks (IJCNN 2011)*. pp. 2690-2697. http://axon.cs.byu.edu/papers/smith.ijcnn2011.pdf.

[5] Breunig, M. M.; Kriegel, H. -P.; Ng, R. T.; Sander, J. (2000). "LOF: Identifying Density-based Local Outliers". *ACM SIGMOD Record* 29: 93.doi:10.1145/335191.335388. http://www.dbs.ifi.lmu.de/Publikationen/Papers/LOF.pdf.edit

[6] Denning, Dorothy, "An Intrusion Detection Model," Proceedings of the Seventh IEEE Symposium on Security and Privacy, May 1986, pages 119-131.

[7] Teng, Henry S., Chen, Kaihu, and Lu, Stephen C-Y, "Adaptive Real-time Anomaly Detection Using Inductively Generated Sequential Patterns," 1990 IEEE Symposium on Security and Privacy

[8] Jones, Anita K., and Sielken, Robert S., "Computer System Intrusion Detection: A Survey," Technical Report, Department of Computer Science, University of Virginia, Charlottesville, VA, 1999 Retrieved from "http://en.wikipedia.org/w/index.php?title=Anomaly_detection&oldid=454689533"

[9] Carl Endorf, Eugene Schultz, Jim Mellander, “Intrusion Detection and Prevention”, Tata Mcgraw Hill.

[10] Animesh Patcha and Jung-Min Park, “An Overview of Anomaly Detection Techniques:Existing Solutions and Latest Technological Trends”Computer Networks (2007),doi: 10.1016/j.comnet.2007.02.001