Exploiting Efficient Data Mining Techniques to Enhance Intrusion Detection Systems

Chang-Tien Lu, Arnold P. Boedihardjo, Prajwal Manalwar
Department of Computer Science
Virginia Polytechnic Institute and State University
7054 Haycock Road, Falls Church, VA 22043
{ctlu,aboediha,pmanalwa}@vt.edu

Abstract *- Security is becoming a critical part of organizational information systems. Intrusion Detection System (IDS) is an important detection that is used as a countermeasure to preserve data integrity and system availability from attacks. Data mining is being used to clean, classify, and examine large amount of network data to correlate common infringement for intrusion detection. The main reason for using Data Mining Techniques for Intrusion Detection Systems is due to the enormous volume of existing and newly appearing network data that require processing. The amount of data accumulated each day by a network is huge. Several Data Mining techniques such as clustering, classification, and association rules are proving to be useful for gathering different knowledge for Intrusion Detection. This paper presents the idea of applying data mining techniques to intrusion detection systems to maximize the effectiveness in identifying attacks, thereby helping the users to construct more secure information systems.*

Keywords: Data mining, Information Security, Intrusion Detection, Anomaly Detection, Intrusion Detection System

1 Introduction

Computer based Information Systems are becoming an integral part of our organizations. An Information System is a computerized system which contains organization information which serves the organization in its various activities and functions. Computer Security is the ability to protect a computer system and its resources with respect to confidentiality, integrity, and availability. Various protocols, firewalls are in existence to protect these systems from computer threats. Intrusion is a type of cyber attack that attempts to bypass the security mechanism of a computer system. Such an attacker can be an outsider who attempts to access the system, or an insider who attempts to gain and misuse non-authorized privileges.

Intrusion Detection System (IDS) is an important detection used as a countermeasure to preserve data integrity and system availability from attacks. Intrusion Detection Systems (IDS) is a combination of software and hardware that attempts to perform intrusion detection. Intrusion detection is a process of gathering intrusion related knowledge occurring in the process of monitoring the events and analyzing them for sign or intrusion. It raises the alarm when a possible intrusion occurs in the system. The network data source of intrusion detection consists of large amount of textual information, which is difficult to comprehend and analyze. Many IDS can be described with three fundamental functional components – Information Source, Analysis, and Response. Different sources of information and events based on information are gathered to decide whether intrusion has taken place. This information is gathered at various levels like system, host, application, etc. Based on analysis of this data, we can detect the intrusion based on two common practices - Misuse detection and Anomaly detection. Misuse detection is based on extensive knowledge of patterns associated with known attacks provided by human experts. Pattern matching, data mining, and state transition analysis are some of the approaches for Misuse detection. Anomaly detection is based on profiles that represent normal behavior of users, hosts, networks, and detecting attacks of significant deviation from these profiles. Statistical methods, expert system are some of the methods for intrusion detection based on Anomaly detection.

The main motivation behind using intrusion detection in data mining is automation. Pattern of the normal behavior and pattern of the intrusion can be computed using data mining. To apply data mining techniques in intrusion detection, first, the collected monitoring data needs to be preprocessed and converted to the format suitable for mining processing. Next, the reformatted data will be used to develop a clustering or classification model. The classification model can be rule-based, decision-tree based, association-rule based, Bayesian-network based, or neural network based. Intrusion Detection mechanism based on IDS are not only automated but also provides for a significantly elevated accuracy and efficiency. Unlike manual techniques, Data Mining ensures that no intrusion will be missed while checking real time records on the network. Credibility is important in every system. IDS are now becoming important part of our security system, and its credibility also adds value to the whole system. Data mining techniques can be applied to gain insightful knowledge of intrusion prevention mechanisms. They can help detect new vulnerabilities and intrusions, discover previous unknown patterns of attacker behaviors, and provide decision support for intrusion management.

In this paper, we investigate several data mining techniques as applied to Intrusion Detection Systems. For this purpose, we study different Data Mining techniques such as classification, clustering, and association rules to

evaluate their usefulness for IDS in serving their purpose. We compare and contrast current available Intrusion Detection System. Detailed comparison based on different criteria is performed and evaluated to punctuate the significance of Data Mining in IDS. We also identify several research issues such as performance, delay and many others that are involved in incorporating Data Mining in Intrusion Detection Systems.

The paper is outlined as follows. Section 2 gives details for different data mining techniques and discusses how these techniques are used for intrusion detection. Section 3 discusses current available intrusion detection system which makes use of data mining techniques and compares them in various aspects. Section 4 discusses some of the research issues of data mining as applied to IDS. Finally, section 5 presents our conclusion.

2 Data Mining and Intrusion Detection

Data Mining is assisting various applications for required data analysis. Recently, data mining is becoming an important component in intrusion detection system. Different data mining approaches like classification, clustering, association rule, and outlier detection are frequently used to analyze network data to gain intrusion related knowledge. This section will elaborate on several of these data mining techniques and will describe how they are used in the context of intrusion detection.

2.1 Clustering

The amount of available network audit data instances is large, human labeling is time-consuming, and expensive. Clustering is the process of labeling data and assigning it into groups. Clustering algorithms can group new data instances into similar groups. These groups can be used to increase the performance of existing classifiers. High quality clusters can also assist human expert with labeling. A cluster is 100% pure if it contains only data instances from one category. Clustering techniques can be categorized into the following classes: pairwise clustering and central clustering. Pairwise clustering (i.e., similarity-based clustering) unifies similar data instances based on a data-pairwise distance measure. On the other hand, Central clustering, also called centroid-based or model-based clustering, models each cluster by its “centroid”. In terms of runtime complexity, centroid-based clustering algorithms are more efficient than similarity-based clustering algorithms.

Clustering discovers complex intrusions occurred over extended periods of time and different spaces, correlating independent network events. The sets of data belonging to the cluster are modeled according to pre-defined metrics and their common features. It is used to detect hybrids of attack in the cluster. Clustering is an unsupervised machine learning mechanism for finding patterns in unlabeled data with many dimensions. K-means clustering is used to find natural groupings of similar alarm records. Records that are far from any of these clusters indicate unusual activity that may be part of a new attack. The network data available for intrusion detection is primarily categorical with attributes having a small number of unordered values. Recently, Stolfo [11] presented a clustering method for detecting intrusions from unlabeled data.

Most of the clustering techniques discussed above are the basic steps involved in identifying intrusion. These steps are as follows - (1) Find the largest cluster, i.e., the one with the most number of instances, and label it normal; (2) Sort the remaining clusters in an ascending order of their distances to the largest cluster; (3) Select the first K1 clusters so that the number of data instances in these clusters sum up to ¼ ´N, and label them as normal, where ´ is the percentage of normal instances ;(4) Label all the other clusters as attacks. Unlike traditional anomaly detection methods, they cluster data instances that contain both normal behaviors and attacks, using a modified incremental k-means algorithm. After clustering, heuristics are used to automatically label each cluster as either normal or attacks. The self-labeled clusters are then used to detect attacks in a separate test dataset.

2.2 Classification

Classification is similar to clustering in that it also partitions customer records into distinct segments called classes. But unlike clustering, classification analysis requires that the end-user/analyst know ahead of time how classes are defined. It is necessary that each record in the dataset used to build the classifier already have a value for the attribute used to define classes. As each record has a value for the attribute used to define the classes, and because the end-user decides on the attribute to use, classification is much less exploratory than clustering. The objective of a classifier is not to explore the data to discover interesting segments, but to decide how new records should be classified. Classification is used to assign examples to pre-defined categories. Machine learning software performs this task by extracting or learning discrimination rules from examples of correctly classified data. Classification models can be built using a wide variety of algorithms. Classification categorizes the data records in a predetermined set of classes used as attribute to label each record; distinguishing elements belonging to the normal or abnormal class. This technique has been popular to detect individual attacks but has to be applied with complementary fine-tuning techniques to reduce its demonstrated high false positives rate. Classifications algorithms can be classified into three types [8]: extensions to linear discrimination (e.g., multilayer perceptron, logistic discrimination), decision tree and rule-based methods (e.g., C4.5, AQ, CART) [8], and density estimators (Naïve Bayes, k-nearest neighbor, LVQ).

Data classification for intrusion detection can be achieved by the following basic steps. (1) In order for a machine learning program to learn the classification models of the *normal* and *abnormal* system call sequences, we need to supply it with a set of training data containing pre-labeled *normal* and *abnormal* sequences. Different mechanisms based on either linear discrimination, decision tree or rule based methods can be used to scan the normal network traces and create a list of unique sequences of system calls. This list is generally named as *normal* list. (2) Next step is to scan each of the intrusion traces. For each sequence of system calls, first look it up in the normal list. If an exact match can be found then the sequence is labeled as *normal*. Otherwise it is labeled as *abnormal*. (3) Then ensure that the normal traces include nearly all possible *normal* short sequences of system calls. An Intrusion trace contains many normal sequences in addition to the abnormal sequences since the illegal activities only occur in some places within a trace.

As compared to the clustering technique, classification technique is less popular in the domain of intrusion detection. The main reason for this phenomenon is the large amount of data needed to be collected to apply classification. To build the traces and form the normal and abnormal groups, significant amount of data need to be analyzed to ensure its proximity. Using the collected data as empirical models, false alarm rate in such case is significantly lower when compared to clustering.

Classification approach can be useful for both misuse detection and anomaly detection, but it is more commonly used for misuse detection. In intrusion detection, data mining classification can be applied to a standard set of malicious virus and benign executable using derived features. Secondly, RIPPER, Naive Bayes and multi-Bayes classifiers can be used to detect malicious virus code. A decision Tree can be exploited to formulate genetic algorithm to create rules that match a set of anomalous connection. There are alternative classifier approaches which we can use for intrusion detection. With intrusion it is observed that over the time user establishes profile based on the number and types of commands he/she executes. In such case, the attributes will be the number and types of commands invoked by the user. We can reduce dimensionality of such collected data by applying data mining classifier approaches like SOM (Self Organizing Maps) and LQM (Learning Vector Quantization). Nearest neighbor classifier approaches based on SOM and LVQ can be used to refine the collected network data in intrusion detection. Thus the various classification approaches can be employed on network data for obtaining specific information and detecting intrusion.

2.3 Outlier Detection

An outlier is an infrequent observation that immensely deviates from the characteristic distribution of other observations. Outlier detection has many applications, such as data cleaning, fraud detection and network intrusion. The existence of outliers indicates that individuals or groups that have very different behavior from most of the individuals of the dataset. Many times, outliers are removed to improve accuracy of the estimators.

Most anomaly detection algorithms require a set of purely normal data to train the model. They assume that anomalies can be treated as previously unobserved patterns. Since an outlier may be defined as a data point which is very different from the rest of the data, we can employ several outlier detection schemes for intrusion detection which are based on statistical measures, clustering methods and data mining methods. Commonly used outlier techniques in intrusion detection are Mahalanobis distance, detection of outliers using Partitioning around medias (PAM), and Bay's algorithm for distance-based outliers. Outlier approaches for categorical data, such as in Guha [6] are not generally available commercially. Unsupervised approaches for detecting outliers in large data sets for the purposes of fraud or intrusion detection are now appearing in the many research, but these approaches are primarily based on ordered data. Knorr and Ng [9] recently developed a distance-based clustering approach for outlier detection in large data sets. Ramaswarny [12] defines a new outlier criterion based on the distance of a point to its kth nearest neighbor. Breunig [3] define a new local outlier factor, which is the degree to which a data point is an outlier.

Outlier detection is very useful in anomaly based intrusion detection. With outlier detection approach, we can detect novel attack/intrusion by identifying them as deviation from normal behavior. The basic steps in detecting intrusion based on outlier detection are as follows. (1) As outlier detection technique is used in anomaly detection, we first need to identify normal behavior. This behavior can be data set or pattern of some events on the network. (2) Then useful set of feature need to be constructed and (3) similarity function need to be defined between them. We will need to run specific outlier detection algorithm on the set of feature. The algorithm can be based on a statistical based approach, a distance based approach, or a model based schema. All these approaches are based on finding the deviation between collected and scanned data sets. In case of intrusion detection, the collected set of data set will be the set of events and their relation to intrusion. Such relation can be calculated based on normal behavior and any other behavior which significantly deviates from normal behavior. As with such deviation we can preempt attacks based on their behavioral deviation. Outlier detection approaches can useful for detecting any unknown attacks. This is the primary reason that makes outlier detection a popular approach for intrusion detection systems.

Statistical based outlier detection scheme uses a probabilistic model as representation of underlying mechanism of data generation. Such probabilistic model can be useful in intrusion detection environment to decide the probability before alarming the system for intrusion. Finite mixture and BACON are major statistical based outlier detection approaches. Distance based outlier detection approaches such as Nearest Neighbor and Mahalanobis approach are engaged in finding outlier that do not have enough neighbors per define density of local neighborhood. Such outlier detection is very useful in anomaly based intrusion detection systems that are involved in detecting abnormal behavior or deviating patterns. Such density based and distance based approaches can help us to identify abnormal behavior from the set of normal behavior and enable us to detect any unknown intrusions.

2.4 Association Rule

The Association rule is specifically designed for use in data analyses. The association rule considers each attribute/value pair as an item. An item set is a combination of items in a single network request. The algorithm scans through the dataset trying to find item sets that tend to appear in many network data. The objective behind using association rule based data mining is to derive multi-feature (attribute) correlations from a database table. Association rule mining finds associations and/or correlation relationships among large set of data items. Association rules show attributes value conditions that occur frequently together in a given dataset. A typical and widely-used example of association rule mining is Market Basket Analysis. Association rules provide information of this type in the form of "if-then" statements. These rules are computed from the data. Association rules are probabilistic in nature. In addition to the antecedent (the "if" part) and the consequent (the "then" part), an association rule has two numbers that express the degree of uncertainty about the rule. In association analysis the antecedent and consequent are sets of items (called itemsets) that are disjoint. The first number is called the support for the rule. The support is simply the number of transactions that include all items in the antecedent and consequent parts of the rule. The other number is known as the confidence of the rule. Confidence is the ratio of the number of transactions that include all items in the consequent as well as the antecedent to the number of transactions that include all items in the antecedent.

Many association rule algorithms have been developed in the last decades, which can be classified into two categories: (1) candidate-generation-and-test approach such as Apriori [1] and (2) pattern-growth approach. The challenging issues of association rule algorithms are multiple scans of transaction databases and a large number of candidates. Apriori was the first scalable algorithm designed for association-rule mining algorithm. The Apriori algorithm searches for large itemsets during its initial database pass and uses its result as the basis for discovering other large datasets during subsequent searches. There are variations of the Apriori algorithm such as AprioriTID and AprioriHybrid. AprioriTID works as Apriori but uses the generated itemsets to search the support instead of rescanning the database. AprioriHybrid is a hybrid of Apriori and AprioriTID. It uses Apriori for its initial passes and switches to AprioriTID when it expects the sets generated would be able to fit into memory.

Use of association rule in analyzing network data in intrusion detection is useful in many ways. Basic steps for incorporating association rule for intrusion detection as follows. (1) First network data need to be formatted into a database table where each row is an audit record and each column is a field of the audit records. (2) There is evidence that intrusions and user activities shows frequent correlations among network data. For example, one of the reasons that “program policies”, which codify the access rights of privileged programs, are concise and capable to detect known attacks is in that the intended behavior of a program, e.g., read and write files from certain directories with specific permissions is very consistent. These consistent behaviors can be captured in association rules. (3) Also rules based on network data can continuously merge the rules from a new run to the aggregate rule set of all previous runs. Thus with the association rule, we get the capability to capture behavior in association rule for correctly detecting intrusion and hence lowering the false alarm rate.

3 Data Mining Based IDS

Besides expert systems, state transition analysis, and statistical analysis, data mining is becoming one of the popular techniques for detecting intrusion. Recently, many IDS vendors are adopting different data mining techniques for detecting intrusions. We explore several such available IDS which use data mining technique for intrusion detection. This section will provide information about these systems and how they are making use of data mining in their overall framework. IDS can be classified on the basis of their strategy of detection. There are two categories under this classification: misuse detection and anomaly detection.

3.1 Misuse Detection Based IDS

Misuse detection searches for known patterns of attack. This strategy is employed by the current generation of commercial intrusion detection systems. One disadvantage of this strategy is that it can only detect intrusions which are based on known patterns. Example misuse detection systems that use data mining include JAM (Java Agents for Metalearning), MADAM ID (Mining Audit Data for Automated Models for Intrusion Detection), and Automated Discovery of Concise Predictive Rules for Intrusion Detection.

JAM

JAM (developed at Columbia University) uses data mining techniques to discover patterns of intrusions. It then applies a meta-learning classifier to learn the signature of attacks. The association rules algorithm determines relationships between fields in the audit trail records, and the frequent episodes algorithm models sequential patterns of audit events. Features are then extracted from both algorithms and used to compute models of intrusion behavior. The classifiers build the signature of attacks. So thus, data mining in JAM builds misuse detection model.

Classifiers in the JAM are generated by using rule learning program on training data of system usage. After training, resulting classification rules is used to recognize anomalies and detect known intrusions. The system has been tested with data from Sendmail-based attacks, and with network attacks using TCP dump data.

Automated Discovery of Concise Predictive Rules for ID (IOWA-IADCPRID)

Researchers at Iowa State University working on Automated Discovery of Concise Predictive Rules for Intrusion Detection [7]. This system performs data mining to provide global and temporal views of intrusions on a distributed system. The rules detect intrusions against programs (such as Sendmail) using feature vectors to describe the system calls executed by each process. A genetic algorithm selects feature subsets to reduce the number of observed features while maintaining or improving learning accuracy. This is another example of data mining being used to develop rules for misuse detection.

3.2 Anomaly Detection Based IDS

Misuse detection can not detect the attack which signatures have not been defined. Anomaly detection addresses this shortcoming. In anomaly detection, the system defines the expected network behavior in advance. Any significant deviations from the problem are then reported as possible attacks. Such deviations are not necessarily actual attacks. The Data mining technique are used to extract implicit, unknown, and useful information from data. Applications of data mining to anomaly detection include ADAM (Audit Data Analysis and Mining), IDDM (Intrusion Detection using Data Mining), eBayes.

MINDS

The MINDS [5] system is being developed and used by the University of Minnesota. As the first step in MINDS, the net flow tools are used to collect the network traffic data from CISCO routers. This data is filtered to remove network connections not interesting for analysis and preprocessed to collect basic features. Such created data is fed into the MINDS system. The known attack detection module detects network connections that correspond to attacks for which the models are known. The remaining connections are fed to the anomaly detection modules, which assigns a score that reflects how anomalous the connection is compared to the normal network traffic. Connections that are highly anomalous are analyzed by the UM network security analysts to determine if they are truly intrusions are false alarms.

EBays

The eBayes system is a newly developed component for the statistical anomaly detector of EMERALD. It applies Bayesian inference on observed and derived variables of the session on hypotheses. Hypotheses can be either normal events or attacks. Given a Bayes model table is built for the hypotheses and variables. Table is adjusted for the current observations. The eBayes can dynamically generate the new hypothesis that helps it detect new attacks. The eBayes may be computationally expensive as the number of hypothesis states increases. Kohavi [10] study different approaches for handling unknowns and zero counts when estimating probabilities for naïve Bayes classifiers, and propose a new variant of the estimator that shows better performance.

3.3 IDS Using both Misuse and Anomaly Detection

Following are the IDSs that use both misuse and anomaly intrusion detection techniques. Thus they are capable for detecting both known and unknown intrusions. Most of the IDS are used and developed at an international scope.

IIDS (Intelligent Intrusion Detection System Architecture)

The IIDS [2] is an active intrusion detection research effort being performed at Center for Computer Security Research (CCSR) at Mississippi State University. It is distributed and network-based modular architecture to monitor activities across the whole network. It is based on both anomaly and misuse detection. In IIDS multiple sensors, both anomaly and misuse detection sensors serving as experts, monitor activities on individual workstations, activities of users, and network traffic by detecting intrusion from and within network traffic and at individual client levels. These components use different methods to detect intrusion information and then, pass anomalous/malicious system behavior indications to the Decision Engine which assess the overall network health. Currently, the IIDS architecture runs in a high speed cluster environment. In this environment, the Decision Engine resides in the head node and monitors intrusion activities across our experimental cluster.

RIDS-100

RIDS(Rising Intrusion Detection System) is provided by Rising Tech. Rising Tech. is a leader in antivirus and content security software and services in China. The company is a leading provider of client, gateway and server security solutions for virus protection, firewall and

intrusion detection technologies and security services to enterprises and service providers around China. RIDS make the use of both intrusion detection technique, misuse and anomaly detection. Distance based outlier detection algorithm is used for detection deviational behavior among collected network data. For misuse detection, it has very vast set of collected data pattern which can be matched with scanned network data for misuse detection. This large amount of data pattern is scanned using data mining classification Decision Tree algorithm.

3.4 Summary

IDSs employ various mining techniques as per system requirement. We observe that the association rule is a common approach to Misuse detection. As the association rule facilitates the correlation of different data set, associated set of events or patterns can be easily correlated with the intrusion. Anomaly detection studies normal behavior and raises the alarm when any behavior seems abnormal. Outlier detection technique provides the clustering outlier detection mechanism for detecting any set of data which falls outside cluster or group. Thus outlier detection mechanism is widely used for anomaly detection.

4 Discussion

Intrusion detection systems have been an area of active research for over fifteen years. Current commercial intrusion detection systems employ misuse detection. As such, they completely lack the ability to detect new attacks. The absence of this capability is a recognized gap in current systems. With the shortcomings of current commercial systems, an important research focus is anomaly detection using data mining. A critical issue for anomaly detection is the need to reduce false alarms, since any activity outside a known problem raises an alarm. Research combining data mining and classification has shown great promise in this area. Data mining in intrusion detection is a relatively new concept. Thus there will likely be obstacles in developing an effective solution. As stated previously, it is possible for a company to collect millions of records per day which need to be analyzed for malicious activity. With this amount of data to analyze, data mining will become quite computationally expensive. Processing power or memory are costly. Though we only need samples of the data in order to generate profiles, analyzing network traffic, without all the data could lead to false conclusions.

5 Conclusion

In this paper, we describe different data mining technique and their usefulness in the context of an intrusion detection system. This paper also provides the description of the current Intrusion Detection Systems that make use of data mining for detecting intrusion. Misuse detection techniques are not sufficient for identifying unknown attacks. For detecting unknown intrusion, we need to study normal behavior inside the data. Data mining provide effective mechanism for understanding normal behavior inside the data and use this knowledge for detecting unseen intrusions. Data mining is becoming an integral part of current IDS. Different data mining techniques like clustering, classification, association rule, and outlier detection techniques are helping the various aspects of intrusion data analyses. More research will help us to overcome the limitations in existing data mining technology and will give us effective mechanisms through which we can identify intrusion with low false alarm rate.

References

[1] Agrawal R., Mannila H., Srikant R., Toivonen H., and Verkamo A., "Fast Discovery of Association Rules," *Advances in Knowledge Discovery and Data Mining, MIT*, 1996.

[2] Ambareen Siraj, Rayford B. Vaughn, and S. M. Bridges, "Intrusion Sensor Data Fusion in an Intelligent Intrusion Detection System Architecture," *Proceedings of the 37th Hawaii International Conference on System Sciences*, 2004.

[3] Breunig, Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying Density-Based Local Outliers", Proceedings of the ACM Sigmod 2000 Intl. Conference On Management of Data, Dallas, TX.," 2001.

[4] Eric Bloedorn, Alan D. Christiansen, William Hill, Clement Skorupka, Lisa M. Talbot, and Jonathan Tivel, "Data Mining for Network Intrusion Detection: How to Get Started."

[5] Ertoz L. et Al, "MINDS - Minnesota Intrusion Detection System,", Next Generation Data Mining Chapter 3, 2004

[6] S. Guha, Rastogi R., and Shim K., "ROCK: A Robust Clustering Algorithm for Categorical Attributes," *Proceedings of the 15th Int. Conference On Data Eng., Sydney, Australia.*, 1999.

[7] Helmer, Wong, Honavar, and Miller, "Automated discovery of concise predictive rules for intrusion detection.," *Technical Report TR 99-01, Department of Computer Science, Iowa State University, Ames, IA.*, 2001.

[8] Henery R. J., "Classification," *Machine Learning Neural and Statistical Classification*, 1994.

[9] Knorr and R. T. Ng, "Algorithms for Mining Distance-Based Outliers in Large Datasets," *Very Large DataBases Proceedings of the 24th Int. Conference on Very Large Databases, Aug 24-27, 1998, New York City, NY, pp. 392-403.*, 1998.

[10] Kohavi, Beckeer, and Sommer, "Improving simple bayes.," *In European Conference on Machine Learning, Prague, Czech Republic.*, 1997.

[11] Portnoy L., E. Eskin, and Stolfo S., "Intrusion detection with unlabeled data using clustering," *In ACM Workshop on Data Mining Applied to Security.*, 2000.

[12] Ramaswarny, R. R. S., and K. Shim, "Efficient Algorithms for Mining Outliers from Large Data Sets," *Proceedings of the ACM Sigmod 2000 Int. Conference on Management of Data, Dallas, TX.*, 2000.