Computer Engineering and Intelligent Systems
ISSN 2222-1719 (Paper) ISSN 2222-2863 (Online)
Vol.8, No.7, 2017
www.iiste.org

Intrusion Detection System: A Survey Using Data Mining and Learning Methods

Yogita Sharma*
PG Scholar, CSE, VITS, Bhopal, India

Prof. Sumit Sharma
HOD CSE, VITS, Bhopal, India

Abstract
In spite of growing information system widely, security has remained one hard-hitting area for computers as well as networks. In information protection, Intrusion Detection System (IDS) is used to safeguard the data confidentiality, integrity and system availability from various types of attacks. Data mining is an efficient artifice applied to intrusion detection to ascertain a new outline from the massive network data as well as it used to reduce the strain of the manual compilations of the normal and abnormal behavior patterns. Intrusion Detection System (IDS) is an essential method to protect network security from incoming on-line threats. Machine learning enable automates the classification of network patterns. This piece of writing reviews the present state of data mining techniques and compares various data mining techniques used to implement an intrusion detection system such as, Support Vector Machine, Genetic Algorithm, Neural network, Fuzzy Logic, Bayesian Classifier, K- Nearest Neighbor and decision tree Algorithms by highlighting a advantage and disadvantages of each of the techniques. This paper review the learning and detection methods in IDS, discuss the problems with existing intrusion detection systems and review data reduction techniques used in IDS in order to deal with huge volumes of audit data. Finally, conclusion and recommendation are included.
Keywords: Classification, Data Mining, Intrusion Detection System, Security, Anomaly Detection, Types of attacks, Machine Learning Techniques

I. INTRODUCTION

Securing computer systems from unauthorized use, such as hackers, and protecting it against any kind of abuse from any legitimate access, such as insider threats, can be defined as Intrusion Detection (ID). Such a computer system breach can cause a loss of data integrity, access denied for online resources, the leak of confidential data, and take advantage of private resources [1]. Intrusion Detection System (IDS) was first implemented by Denning (1987), [2], and since then the IDS has become a hot research topic as an important tool for computer network security. Intrusion in network is the case when there was a breakthrough into the network by external or internal bad guys who can have different goals but mainly to have access to protected data or denying legitimate users from accessing their resources. The technology grows and many techniques are put in place to control the security on network; at the same time attackers keep innovating new techniques to succeed their mission. Attackers have their technique to disguise themselves (spoofing) so that they can cross the security perimeters like firewall. An intrusion detection system (IDS) can take a form of software or hardware platform used to monitor the state of network or system for unusual activities (behavior) [3]. Its main purpose is not to prevent attack like firewall does but to detect breakthrough if any and alert the security team about it. Whatever form it takes, IDS have four components: the sensor of events as it is generated from source, the analyzer to check on characteristics or behavior of the event, the log for future references, the last part is the case when IDS is operating in active mode [3]. The source of event depends on what type of IDS deployed; if that was network-based IDS (NIDS), the source reflects any promiscuous network port of the hardware and this last one captures packet flow events. But if it was host-based IDS (HIDS) the source is log of events already collected and stored somewhere on a host. Distinction of these sources and associated events is a key to understanding functionality of IDS. So, we distinguish three categories of intrusion detection systems: network-based, host based and hybrid one.

Host-based intrusion detection system (HIDS) is always in a form of software which is an intelligent system installed on a host and can be configured to operate as standalone or in distributed mode just like other systems. HIDS mainly uses operating system audit event logs to watch pattern of activities and attempt to discover unusual activities on single computer Since HIDS relies on log, it means that the scope of attack detection is the activities initiated by local users. HIDS creates database of state of every installed program on the host where it is running and continue to watch its behavior continuously [3].

Network-based intrusion detection system (NIDS) is a type of IDS which relies solely on the network traffic as raw data in order to detect suspicious activities. It can be either software or hardware; an example of hardware IDS is CISCO ISA 5000 series. This last one is capable of logging suspected packet flow. Furthermore, the administrator can access and view it. If it is deployed as software the good example is the use of packet analyzer

systems installed on a host in LAN, so that all traffic flows must have a copy sent to that host. The technique is known as mirroring. Both using network traffic flow, they can detect flows which have deviated from expected metrics or rules on the entire network segment or subnet but they cannot interpret behavior inside any other machine.

A hybrid of the two categories is the deployment option of both HIDS and NIDS on computer running sensitive application accessible by many users on the network. In this paper we focus on this network-based IDS. The IDS is further distinguished into two types based on the mode of detection: rule-based and anomaly based. The rule-based IDS compare every packet flow against rules of well-known types of attack. The systems running this type of IDS are believed to detect presence of intrusion with fewer false alarms if any. It means that if an attack is occurring, it is alerted and normal traffic is not alerted as attack. The disadvantage of such detection mode is that if attackers change techniques the system remains blind and calm as if nothing has occurred. Anomaly-based IDS is capable of detecting novel attacks because it does not rely on prior knowledge of attack; rather, it uses statistical distribution of traffic flow with their features and machine learning techniques. This type of detection mode poses the problem of false alarms; some normal packets are alerted as attack; and not-normal packets are not alerted. The limitation of anomaly-based IDS (e.g., producing higher false alarms) became a general research problem in network intrusion detection. So many methods have been proposed for solution mostly using machine learning and artificial intelligent techniques.

Generally, there is a problem when artificial intelligence techniques are used to design IDS. Mostly, clustering techniques like k-mean, c-mean, or hierarchical clustering methods require the predefinition of the number of quality clusters from the dataset arbitrary. When the number of clusters is small, there is a risk that some data point will be forced into inappropriate clusters (called under fitting) and when the numbers of clusters are too many there is a risk of having unnecessary clusters (called over fitting). The problem becomes more serious when the dataset has increased or reduced because the previously defined number of clusters can no more fit the new size (robustness) [4].

Another problem lies on dataset pre-processing especially on features reduction. There is a risk of over reduction of features until the data loses its main characteristics. This is what was stated in the survey of [5] where some research could not be included as network-based IDS because the proposed method has excluded all basic network packet features. In our proposed method, we seek to use basic TCP features exclusively.

II.TRADITIONAL INTRUSION DETECTION

There are two types of traditional intrusion detection system

Anomaly Detection

It refers to detect abnormal behavior of host or network. It actually refers to storing features of user's usual behaviors hooked on database, then its compare user's present behavior with database. If there arises a deviation huge enough, it is said that the data tested is abnormal. The patterns detected are called anomalies. Anomalies are also referred to as outliers.

Misuse Detection

In misuse detection approach, it defines abnormal system behavior at first, and then defines any other behavior, as normal behavior. It assumes that abnormal behavior and activity has a simple to define model. It advances in the rapid of detection and low percentage of false alarm. However, it fails in discovering the non-pre-elected attacks in the feature library, so it cannot detect the abundant new attacks. IDS provide the following security functions

Data Confidentiality

It checks whether the information stored on a system is protected against unconstitutional access. Since systems are sometimes used to manage sensitive information, data confidentiality is often a gauge of the ability of the system to protect its data.

Data Availability

The network should be tough to Denial of Service attacks. Intrusion detection system based on sources of audit information it can be divided into 3 subcategories

Data Integrity

It refers to maintaining and assuring the correctness and consistency of data over its entire life-cycle. No corruption
or data loss is acknowledged either from random events or malicious activity.

Host Based IDS

It refers to intrusion detection that takes place on a single host system. It gets audit data from host audit trails and monitors activities such as integrity of system, file changes, host based network traffics, and system logs. If there is any unlawful change or movement is detected, it alerts the user by a pop-up menu and informs to the central management server. Central management server blocks the movement, or a combination of the above two. The judgment should be based on the strategy that is installed on the local system.

Network Based IDS

It is used to supervise and investigate network traffic to protect a system from network-based threats. It tries to detect malicious activities such as denial-of-service (Dos) attacks and network traffic attacks. Network based IDS includes a number of sensors to monitors packet traffic, one or more servers for network management functions, and one or more management relieves for the human interface.

Hybrid Intrusion Detection

The recent development in intrusion detection is to combine both types host-based and network-based IDS to design hybrid systems. Hybrid intrusion detection system has flexibility and it increases the security level. It combines IDS sensor locations and reports attacks are aimed at particular segments or entire network.

III.TYPES OF ATTACKS

DoS Attack

A denial-of-service attack or distributed denial-of-service attack is an effort to make a computer resource out of stock to its Intended users. In this type of attack it slow down the system or shut down the system so it disrupt the service and deny the legitimate authorized user. Due to this attack high network traffic occurs.

User to Root Attack (U2R):

In this type of attack the attacker starts with user level like taking down the password, dictionary attack and finally attacker achieves root to access the system.

Probing:

In this type of attack an attacker examines a network to gather information or discover well-known vulnerabilities. An attacker who has a record, of which machines and services are accessible on a known network, can make use of this information to look for delicate points.

Remote to User Attack (R2U):

In this type of attack an attacker have the capability to send packet to a machine over a network but does not have an account on that machine, make use of some vulnerability to achieve local access as a user of that machine.

Eavesdropping Attack

Eavesdropping is a network layer attack consisting of capturing packets from the network transmitted by others' computers and reading the sensitive information like passwords, session tokens, or any kind of confidential information.

Man-in-the-Middle Attack

In this the attacker makes independent connections with the victims and relays messages between them and making them believe that they are talking directly to each other over a private connection, but the fact is entire conversation is controlled by the attacker.

Drawbacks of IDS

Intrusion Detection Systems (IDS) have become an important component in security infrastructures as they permit networks administrators to identify policy variations. These policy violations range from outside attackers trying to gain unconstitutional access to intruders abusing their access. Current IDS have a number of considerable drawbacks

False positives: A major problem is the amount of false positives IDS will produce. Developing distinctive signatures is a complicated task. It is much trickier to pick out a legitimate intrusion attempt if a signature also alerts regularly on valid network activity.

False negatives: In these IDS does not generate an alert when an intrusion is actually taking place. It simply put if a signature has not been written for a particular exploit there is an tremendously good chance that the IDS will not detect it.

IV. DATA MINING ASSISTS FOR INTRUSION DETECTION

The central theme of intrusion detection using data mining approach is to detect the security violations in information system. Data mining can process large amount of data and it discovers hidden and ignored information. To detect the intrusion, data mining consist of following process like classification, clustering, association rule learning and regression. It monitors the information system and raises alarms when security violations are founded. Fig 1[30]: The Data Mining Process of Building ID Models

Support Vector Machine (SVM)

SVM [6][7] is a learning method for the Classification and Regression analysis of both linear and nonlinear data. It uses a hypothesis space of linear functions and maps input feature vectors into a higher dimensional space all the way through some nonlinear mapping.SVM constructs a hyper plane or set of hyper planes only the good separation is achieved by the hyper plane. [8] The hyper plane searching process in SVM is achieved by the leading margin. The related margin gives the major separation between classes. While training an SVM it creates a quadratic optimization problem.

In SVM the classifier is created by linear separating hyper plane but all the linear separation cannot be solved in the original input space. SVM uses a function called kernel to solve this problem. The Kernel transforms linear problem into nonlinear one by mapping into feature spaces. Radial basis function, polynomial, two layer sigmoid neural nets are the some of the kernel functions. At the time of training classifier, user may provide one of these functions, which selects support vectors along the surface of this function. The implementation of SVM tries to accomplish maximum separation between the classes. Intrusion detection system has two phases: training and testing. SVMs can learn a larger set of patterns and be able to provide better classification, because the categorization difficulty does not depend on the dimensionality of the feature space. SVMs also have the ability to update the training patterns dynamically whenever there is a new pattern during classification.

Genetic Algorithms

Genetic algorithms were initially introduced in the meadow of computational biology. After that they have been bloomed into various fields with promising result. Nowadays the researchers have tried to incorporate this algorithm with IDSs. The REGAL System is based on distributed genetic algorithm. REGAL is a concept learning system that learns First Order Logic multi-model concept descriptions. The learning examples are stored in relational database that are represented as relational tuples. Gonzalez and Dasgupta [9] applied a genetic algorithm, though they were examined host based IDSs, not network based. They used the algorithm only for the Meta learning step instead of running algorithm directly on the feature set. It uses the statistical classifiers for labeled vectors.2-bit binary encoding methodology is used for identifying the abnormality of a particular feature, ranging from normal to abnormal. Chittur [10] used a genetic algorithm with decision tree. Decision tree is used to represent the data. They used the high detection rate that reduces the false positive rate. The false positive occurrence was minimized by utilizing human input in a feedback loop.

Neural Networks

Neural Network was traditionally used to refer a network or biological neurons. [11] In IDSs neural network has been used for both anomaly and misuse intrusion detection. In anomaly intrusion detection the neural networks were modeled to recognize statistically significant variations from the user's recognized behavior also identify the typical characteristics of system users. In misuse intrusion detection the neural network would collect data from the network stream and analyze the data for instances of misuse. In neural network the misuse intrusion detection can be implemented in two ways. The first approach incorporates the neural network component into an existing system or customized expert system. This method uses the neural network to sort the incoming data for suspicious events and forward them to the existing and expert system. This improves the efficiency of the detection system. The second method uses the standalone misuse detection system. This system receives data from the network stream and analyzes it for misuse intrusion. It has the ability to learn the characteristics of misuse attacks and identify instances that are unlike any which have been observed before by the network. It has high degree of accuracy to recognize known suspicious events. Generally, it is used to learn complex non linear input-output relationships.

Fuzzy Logic

Fuzzy logic is derived from fuzzy set theory it uses the rule based systems for classification. Fuzzy can be thought of as the application side of fuzzy set theory dealing with sound thought out real world expert values for a complex problem. The fuzzy data mining techniques used to extract patterns that represent normal behavior for intrusion detection. The sets of fuzzy association rules are used to mine the network audit data models and to detect the anomalous behavior the set of fuzzy association rules are generated. [12] The audit data and mined normal data have been compared to identify the similarity. If the similarity values are below an upper limit, an alarm raises.

Bayesian Classifier

A Bayesian Classifier provides high accuracy and speed for handling large database. In network model Bayesian classifier encodes the probabilistic relationship among the variable of interest. In intrusion detection this classifier is combined with statistical schemes to produce higher encoding interdependencies between the variables and predicting events. Bayesian belief networks based on the joint conditional probability distributions. The graphical model of casual relationships performs learning technique. This technique is defined by two components-a directed acyclic graph and a set of conditional probability tables.DAG represents a random variable these variables may be discrete or continuous. For each variable classifier maintains one conditional probability table (CPT). It require higher computational effort

K-Nearest Neighbour

K-Nearest Neighbor (k-NN) is a type of Lazy learning, it simply stores a given training tuple and waits until it is given a test tuple. It is an instance based learner that classifies the objects based on closet training examples in the feature space. For a given unknown tuple, a k-Nearest neighbor looks the pattern space for the k-training tuples that are closest to the unknown tuple. It is the simplest algorithm among all the machine learning algorithms. Here the object is classified by a majority vote of its neighbors. The object is simply assigned to the

class of its neighbor only in the case of K=1. For a target function this algorithm uses all labeled training instances model. To obtain the optimal hypothesis function algorithm uses similarity based search. The intrusion is detected with the combination of statistical schemes. This technique is computationally expensive and requires efficient storage for implementation of parallel hardware.

Decision Tree

Decision tree is a classification technique in data mining for predictive models. Decision tree is a flowchart like tree structure where internal node represents a test on attribute, branch represents an outcome of the test and leaf node represents a class label. From the pre classified data set it inductively learns to construct the models. Here each data item is defined by the attribute values. Initially decision tree is constructed by set of pre-classified data. The important approach is to select the attributes, which can best divide the data items into their respective classes based on these attributes the data item is partitioned. This process is iteratively applied to each partitioned subset of the data items. If all the data items in current subset belongs to the same class then the process get terminate. Each node contains the number of edges, which are labeled along with a possible value of attribute in the parent node. An edge connects either a node or two nodes. Leaves are always labeled with a decision value for classification of the data. To classify an unidentified object, the process is starts at the root of the decision tree and follows the branch. Decision trees can be used for misuse intrusion detection that can learn a model based on the training data and predict the future data from the various types of attacks. It works well with large data sets. Decision tree model also be used in the rule-based techniques with minimum processing. It provides high generalization accuracy.

II. LEARNING METHODS IN IDS

In Machine Learning (ML), the instance which has a corresponding label is known as supervised learning, and the instance which has no label is known as unsupervised learning. In a situation where some of the samples are labeled and some are not, this called Semi-supervised learning.

A. Supervised learning

The process of supervised classification is to build a model that can differentiate between at least two classes based on numerical features with minimal errors for the new unseen before samples [13]. To build that model, the classifier needs a labeled training dataset contains normal and attacks samples. Since supervised learning provides classifier with more information than semi-supervised and unsupervised techniques, theoretically, it can help for a better rate of detection. However, supervised learning suffers some issues. (i) Unavailability of a given datasets in the training time that can cover all legitimate aspects. (ii) Accurate labels are not always guaranteed. (iii)High false alarm rate if the training dataset contains noise [14].

B. Semi-supervised learning

The semi-supervised method is kind of in-between supervised and unsupervised methods. In the application of real time anomaly intrusion detection, the semi-supervised method is more practical since it requires only the labeled data of the normal class, but such method is not widely used because there are available labels for every possible anomaly in the training time [14].

C. Unsupervised learning

Unsupervised learning does not use labeled data, it uses statistical models to partition data into normal and anomalies without any prior knowledge, based on two assumptions [15]. (i) It presumes that amount of data is normal and the anomalies represent a very small amount of it. (ii) Statistically, the normal and anomalies data are different from each other.

III. DETECTION METHODS IN IDS

There are two detection methods in IDS, misuse (signature) and anomaly (behavior) [16]. In misuse detection, the system stores known attacks signature and looking for such signatures in network traffic, if something matches, it will be considered as a Misuse. Misuse detection can detect attacks with less false alarm rate, but cannot detect new attacks which have no defined signatures [17]. In anomaly detection Building a model is usually established by recording normal activities traffic in the LAN, and then once a system monitors any deviation from that model, it will be considered as an anomaly or an attack. To build an anomaly model, the system needs to be trained and tested on a specific dataset applying Machine Learning (ML) methods. In the training phase, the dataset is described by features and associated with labels for each record. Labels are a set of binary values that describe each record whether it is normal or attack. [18].

IV. PROBLEMS WITH EXISTING INTRUSION DETECTION SYSTEMS

In ideal IDS should have an important characteristic, such as high accuracy detection rate, low false positive and negative rate, and low computational cost [19] Present IDS should cope with a new challenge such as.

- Fast expansion in network systems and digital devices, i.e. smart mobile phones, tablets, laptops.
- Increasing numbers of sophisticated network intrusions, such as systematic multistep detailed process

taken by a hacker before conducting the attack.

- Occurrence of a new intrusions and software vulnerabilities.
- Low-level detection with independent intrusion reporting. No matter what detection principle has been applied and what audit data has been examined in intrusion detection, the deployed intrusion detection mechanism or IDS normally detects the intrusion entirely from low-level raw information (for example, raw network traffic, host log file record). Though low-level information contains the potential implications for intrusion possibilities, it is difficult for the intrusion detector to directly report the high-level alarm abstraction (that is, the system has been compromised). An issue in independent intrusion reporting is when hacker implements the same attack repeatedly, the IDS will be busy generating repetitive alarms in one specific time, meanwhile, the system administrator is overwhelmed by the flood of redundant alarms, and the hacker can gather additional information and create other intrusions.
- Self-existed weaknesses in applied IDS can allow the hacker to disable it remotely without notification for the system administrator.

V. DATA REDUCTION TECHNIQUES USED IN IDS

Data reduction is a method that offers a tool for analysis which makes it possible to derive useful information from a huge dataset so it can be used in more exploration. Most of the data mining and machine learning methods could not operate well in IDS due to the huge size of network data, and that can cause a long computational time. One reason for that is the nature of collected network data which contain a big number of extracted features that IDS should process [20]. The number of features in data is essential for a good classification as if the features are too high, that will make a loss of generalization, and if it is too small, that also could degrade classification quality. Feature reduction is a good method that helps people to understand the quality and importance of features and how they are related to each other [21]. Besides, empirical results in IDS show that when using a feature reduction technique, the classifier performance in terms of accuracy and computational cost is improved [22]. Methods of feature and dimensionality reduction, such as clustering, feature extraction, and selection, has been employed recently by many researchers in IDS as a preprocessing step to improve the accuracy and reduce the computational complexity.

A. Feature selection

The purpose of feature selection (FS) is to find the best feature subset that could enhance classification process and produces less generalization error. Further objectives of FS are [23]:

- Increase classifier's speed and reduces storage capacity
- Save resources for the next process of data collecting.
- Obtain more information and understanding how data generated. In the area of IDS, feature selection has been used to improve the classification process. Principal component analysis (PCA), information gain (IG), and genetic algorithm (GA) are three popular feature selection approaches. Features selection can be based on the wrapper, filter, and embedded methods [23],
- Wrapper method uses a classifier for evaluating the optimal features, this classifier is utilized as a black box for such a task. Such a method can achieve good generalization, but may suffer high-dimensionality due to the computational cost of training the classifier in a combinatorial number of times.
- Filter methods are relatively robust against over fitting, it doesn't use any classifiers for feature assessments, but it uses independent measurement methods, such as consistency measures, correlation measures, and distance measure.
- Embedded method is a mix between wrapper and filter. Although it has less computational cost than the wrapper, it is much slower than filter method.

There are three principal approaches to feature selection. The shades show the components used by the three approaches: filter, wrappers, and embedded methods. The principal are [24]:

1. (Search strategy) also called feature subset generation. Wrapper and filter methods share the same search strategies to look for all potential features in the domain space.
2. Evaluation criterion definition. The most likely difference between filters and wrappers are by the evaluation criterion. Filters do not use machine learning, while wrapper does.
3. Assessment method also called evaluation criterion estimation.

Where the machine learning or performance of relevance index is defined by estimating the limited amount of data used for training. There are two possible strategies for that:

I. (out-of-sample strategy, where machine learning applied by dividing the training data into training and validation subsets. The training used for parameters estimation of the model, while validation is used to evaluate the classifier performance. k-fold cross validation known as multiple splitting is a common method for decrease the classifier variance.

II. Ensemble strategy, referring to classical statistics, where all data is used to calculate the empirical

estimate, then find the importance features based on a statistical test.

B. Feature extraction

Feature extraction is an approach applied to find the best set of features. In data, columns representing features, and rows representing samples, where features are an outcome of qualitative and quantitative observations. Linear and non-linear feature extraction methods are well known in the domain. i.e. self-organizing feature map (SOM), is a non-linear, while Principal component analysis (PCA) is a linear feature extraction method.

C. Clustering

Clustering is unsupervised learning method that is essential in data mining task. It divides the data into similar groups for the simplicity sake, but that can be on the cost of information detail loss.ped on the foundation of classic techniques.

Conclusion

As the intrusions and attacks to the networks are getting stronger day by day, the smart intrusion detection systems should be used to keep the networks safe and secure. In this research various Intrusion Detection Systems were analyzed and surveyed. The current techniques have their own advantages and disadvantages. Building an effective intrusion detection system using machine learning methods has received much attention for network security. Data set always contain a huge number of features where most of it are redundant or irrelevant. Employing feature reduction method is an essential to reduce the computational cost and increase the classifier performance. Feature selection and feature extraction are having advantages and disadvantages, which make it hard to choose a single method to implement. It's recommended to use feature extraction followed by feature selection as a hybrid approach to increase the accuracy of intrusion detection.

References

[1.] Aburomman and M. B. I. Reaz, "A novel svm-knn-pso ensemble method for intrusion detection system," *Applied Soft Computing*, vol. 38, pp. 360–372, 2016.

[2.] D. E. Denning, "An intrusion-detection model," IEEE Trans. Softw. Eng., vol. 13, no. 2, pp. 222–232, Feb. 1987. [Online]. Available: http://dx.doi.org/10.1109/TSE.1987.232894

[3.] K. Scarfone and P. Mell, "Guide to Intrusion Detection and Prevention Systems (IDPS)," Nist Spec. Publ., vol. 800–94, p. 127,2007.

[4.] K. Muchammad and T. Ahmad, "Detecting Intrusion Using Recursive Clustering and Sum of Log Distance to Sub-centroid," Procedia - Procedia Comput. Sci., vol. 72, pp. 446–452, 2015.

[5.] J. J. Davis and A. J. Clark, "Data preprocessing for anomaly based network intrusion detection: A review," Comput. {&} Secur., vol. 30, no. 6–7, pp. 353–375, 2011.

[6.] W. Feng, Q. Zhng, G. Hu, J Xiangji Huang," Mining network data for intrusion detection through combining SVMs with ant colony networks"Future Generation Computer Systems,2013.

[7.] Y. Li u, X. Yu, J.X. Huang, A." An, Combining integrated sampling with SVM ensembles for learning from imbalanced datasets", Information Processing &Management 47 (4) (2011) 617–631.

[8.] L. Khan, M. Awad, B. Thuraisingham, "A new intrusion detection system using support vector machines and hierarchical clustering", The VLDB Journal 16(2007) 507–521.

[9.] Dasgupta, D. and F. A. Gonzalez,"An intelligent decision support system for intrusion detection and response", In Proc. of International Workshop on Mathematical Methods, Models and Architectures for Computer Networks Security (MMM-ACNS),St.Petersburg. Springer-Verlag, 21-23 May, 2001

[10.] Chittur, A., "Model generation for an intrusion detection system using genetic algorithms", High School Honors Thesis, Ossining High School. In cooperation with Columbia Univ, 2001

[11.] J. Ryan, M.-J. Lin, R. Miikkulainen,"Intrusion detection with neural networks", in: Proceedings of AAAI-97 Workshop on AI Approaches to Fraud Detection and Task Management, 1997, pp. 92–97.

[12.] Luo, J.,"Integrating fuzzy logic with data mining methods for intrusion detection", Master's thesis, Mississippi State Univ., 1999.

[13.] Oberth¨ur and P. Warnat, "Supervised classification," Encyclopedia of Cancer, pp. 3563–3565, 2012.

[14.] Daneshpazhouh and A. Sami, "Entropy-based outlier detection using semi-supervised approach with few positive examples," Pattern Recognition Letters, vol. 49, pp. 77–84, 2014.

[15.] S. Jiang, X. Song, H. Wang, J.-J. Han, and Q.-H. Li, "A clustering-based method for unsupervised intrusion detections," Pattern Recognition Letters, vol. 27, no. 7, pp. 802–810, 2006.

[16.] J. J. Davis and A. J. Clark, "Data preprocessing for anomaly based network intrusion detection: A review," Computers & Security, vol. 30, no. 6, pp. 353–375, 2011.

[17.] Kruegel, F. Valeur, and G. Vigna, Intrusion detection and correlation: challenges and solutions. Springer Science & Business Media, 2005, vol. 14.

[18.] Kruegel, G. Vigna, and W. Robertson, "A multi-model approach to the detection of web-based attacks," Computer Networks, vol. 48, no. 5,pp. 717–738, 2005.

[19.] Z. Zhang, P.-H. Ho, and L. He, "Measuring ids-estimated attack impacts for rational incident response: A decision theoretic approach," Computers & Security, vol. 28, no. 7, pp. 605–614, 2009.
[20.] Guo, Y.-J. Zhou, Y. Ping, S.-S. Luo, Y.-P. Lai, and Z.-K. Zhang, "Efficient intrusion detection using representative instances," Computers & Security, vol. 39, pp. 255–267, 2013.
[21.] S. Saha, A. S. Sairam, A. Yadav, and A. Ekbal, "Genetic algorithm combined with support vector machine for building an intrusion detection system," in Proceedings of the International Conference on Advances in Computing, Communications and Informatics. ACM, 2012, pp. 566–572.
[22.] Amiri, M. R. Yousefi, C. Lucas, A. Shakery, and N. Yazdani, "Mutual information-based feature selection for intrusion detection systems," Journal of Network and Computer Applications, vol. 34, no. 4, pp. 1184–1199, 2011.
[23.] J. R. Vergara and P. A. Est´evez, "A review of feature selection methods based on mutual information," Neural Computing and Applications, vol. 24, no. 1, pp. 175–186, 2014.
[24.] Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, Feature extraction:foundations and applications. Springer, 2008, vol. 207.