International Journal of Computer Science
ISSN:2348-6600
Oddity...Probe...Reviste...
Published by SK Research Group of Companies(SKRGC)

http://www.ijcsjournal.com Volume 3, Issue 2, No 1, 2015. ISSN: 2348-6600
Reference ID: IJCS-096 PAGE NO: 549-555.

A Review of Anomaly-Based IDS's and Techniques

Chandrima Dutta Prof. Amit Saxena
Department of Computer Science , Truba Institute of Science & IT, Bhopal,India
Cdutta03@gmail.com,amitsaxena@trubainstitute.ac.in.

Dr. Manish Manoria
Director, Truba Institute of Science & IT,Bhopal, India.
manishmanoria@trubainstitute.ac.in

***Abstract*— Due to rapid growth and deployment of network technologies and global internet services has made better administration and protection of unauthorized networks activity a difficult research problem. This development is go along with by an exponential expansion in the number of network attacks over insecure channel, which have become more difficult, more categorized, more active, and more rigorous than ever. Modern network protection techniques are static, time-consuming in responding to attacks, and inefficient due to the large number of false alarms.**

***Index Terms*—**IDS, Anomalies, Machine Learning, Support vector machine, Signature based Detection.

I. INTRODUCTION

The field of intrusion detection has received increasing awareness in modern time. One explanation for this is the unstable expansion of the Internet and the large amount of networked systems that exist in all types of organizations. The increase in the number of networked machines has lead to an increase in unauthorized activity, not only from outside attackers, but also from inside attackers, for instance disgruntled employee and people abusing their privileges for personal gain.

One of the main reasons behind the under-deployment of anomaly-based IDSs is that they cannot be without problems organized. A common of anomaly-based IDSs exploit data mining and machine learning algorithms, which take feature vectors containing some complicated attributes as their inputs. By monitoring several concert metrics, and all the way through training on what is measured normal activity, an anomaly-based system is able to determine when the system is operating in an uncharacteristic situation. Due to its less-precise environment an anomaly-based system can detect abnormal conditions on a system without detailed knowledge of any particular threat [1]. While this can avoid many of the issues that plague signature-based systems with large rule sets, an anomaly-based system cannot openly identify any given threat. Rather, an anomaly-based system can only report that something is out of the ordinary. Additionally, it is extremely difficult to train an anomaly-based system to recognize normal machine activity, as user activity changes over time [1]. Minimizing the false-positive rate has proven challenging.

Despite its growing importance, current IDS solutions available have limited response mechanisms. While the researches focus is on better intrusion detection techniques, reply and efficient response to threats are still mostly manual and rely on human agents to take effect [2].

II. THEORETICAL BACKGROUND

The increase in the number and diversity of computing devices in use at the moment, in addition to the complication of enterprise computing environments, pose a maintenance and management challenge to a highly skilled workforce [4]. The high demand for skilled IT personnel is already outstripping supply and labor cost is exceeding equipment costs. As a result of the aforementioned factors, and motivated by the idea that a system should be able to manage itself, there is an overwhelming economic and practical need to automate, as

International Journal of Computer Science
ISSN:2348-6600
Oddity...Probe...Reviste...
Published by SK Research Group of Companies(SKRGC)
http://www.ijcsjournal.com Volume 3, Issue 2, No 1, 2015. ISSN: 2348-6600
Reference ID: IJCS-096 PAGE NO: 549-555.

much as possible, today's system maintenance and administration.

Machine Learning frequently forms the basis for anomaly method and it can detect novel attacks by comparing suspicious ones with normal traffics but has high false alarm rate due to difficulty of modeling normal behaviors for protected system [5]. With misuse method, pattern matching on known signatures leads to high accuracy for detecting threats but it cannot detect novel attacks as novel attack signatures are not available for pattern matching.

III. Intrusion Detection Systems (IDSs) & Its Methodologies

The widespread use of corporate networks with sophisticated technologies, e.g. Web services, distributed databases and remote access, has raised concerns in terms of security issues. Network Intrusion Detection Systems (NIDSs) are one of the major techniques used to protect such networks against well-designed dispersion. Traditionally, to make safe computer systems, network services and running applications, alternative was made to the formation of defensive "shields". Security mechanisms such as firewalls [6], authentication mechanisms and Virtual Private Networks (VPN) have been developed in order to protect the systems of organizations. However, these security mechanisms have almost inevitable vulnerabilities and are usually insufficient in ensuring the complete security of the infrastructure. Attacks are continually being adapted to exploit the system's limitations, often reasoned by casual design and implementation faults. This accounts for the need for protection technology that can monitor systems and identify security policy violations. This is called *intrusion detection*, and complements conventional security mechanisms [7].

Understandably, intrusion is popularly defined as a malicious and externally or internally induced operational fault. Nowadays, computer intrusions and attacks are often regarded as synonymous. But more theoretically, an attack is an challenge to intrude into what is supposedly a secure network, while an intrusion is actually the result of an attack that has been partially or completely successful [8]. "Intrusions in the computer systems are usually caused by attackers accessing the systems from the Internet, or by allowed users of the schemes who attempt to misuse the privileges given to them and/or to gain additional privileges for which they are not authorized" [8]. For this reason, the discrepancy that intrusion is a consequence of attack, however, unsuccessful attack is not necessary to result in an intrusion. Therefore, throughout this thesis, both terms are used from the viewpoint of the defender, and thus checking an attack is comprehensive of stopping an intrusion.

An IDS is a system for detecting and preventing such intrusions. A methodological definition make available by the National Institute of Standards and Technology [9] is that it is "the process of monitoring the events occurring in a computer system or network and analyzing them for signs of intrusions, defined as efforts to cooperate the privacy, reliability, accessibility, or to bypass the security mechanisms of a computer network". An IDS assure its explanation for being by observing the network traffic or looking at OS events [10]. An IDS can be defined as "a combination of software and/or hardware components that monitors computer systems and raises an alarm when an intrusion happens" [10].

Axelsson [11] proposed a generalization model of IDSs as an alternative taxonomy, as shown in Figure 1. The classification is mainly based on detection principles and operational aspects. Even though several methodologies have arisen to classify IDSs from the time when 1980, these fall into three common approaches: 1) anomaly- (behavior) based, 2) signature- (information) based, and 3) mix systems (anomaly and signature).

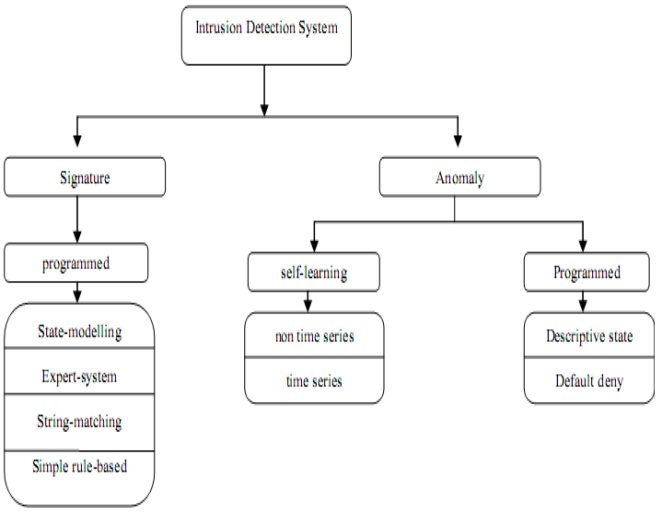

Figure 1: Axelsson's classification of Intrusion Detection Systems (IDSs) [11].

International Journal of Computer Science
ISSN:2348-6600
Oddity...Probe...Reviste...
Published by SK Research Group of Companies(SKRGC)
http://www.ijcsjournal.com Volume 3, Issue 2, No 1, 2015. ISSN: 2348-6600
Reference ID: IJCS-096 PAGE NO: 549-555.

A. Anomaly-based detection

Anomaly-based detection methods are based on a deviation of abnormal activities from the normal or expected behavior of the system. A set of characteristics of the system are observed and analyzed to create a model of normal behavior using collections of information about the system over a particular time period. IDSs can detect anomalies when they evaluate current behavior to the normal system model with the purpose of recognize, description and block any destruction. Furthermore, anomaly-based methodologies are based on the assumption that any anomaly is an indication of a potential attack.

Normal behavior is learned by the system during an online/offline training phase (heuristic systems). Collected data from the learning stage is analyzed, pre-processed and processed; then the normal model is built according to these observations. Therefore, audit data is inspected for any abnormal patterns deviating from the normal model baseline, and these are think about malicious. The usefulness of these methodologies depends on the selected variables and parameters to build the model of the system outline. These parameters contrast from uncomplicated statistical data to comprehensive measures. Therefore, robustness of these systems is proportional to the amount and accuracy of measured data. In addition, these sorts of systems should be adaptable due to the complex and changing nature of protected situations, for instance communication networks. Its reviews [12] the anomaly-based IDS process into three stages: 1) a *parameterization* stage to collect the observed instances of normal system behavior; 2) a *training stage* to characterize the normal and abnormal models, which can be achieved either manually or automatically; and 3) a *detection stage* to detect any deviation exceeding a pre-defined threshold. These methods are hypothetically able to detect novel and 0-day attacks [13]. On the other hand, their effectiveness is robustly dependent on model construction and threshold choice. Numerous methods are used to put together anomaly-based systems.

- **Statistical techniques**

The objective of statistical techniques is to observe the system's activities in order to determine its behavior, and then to generate system outlines. Selected variables are illustration over a particular period of time to measure the normal behavior of the system. The observed activities can be system logs, spatial and temporal distinctiveness of network traffic, or system calls. Two models are built: a model stored or programmed and a current model; and detection is based on the degree of abnormality in the comparison of the two models considering a threshold metric. The benefit of these come within reach of is that they do not require prior knowledge of the observed systems. Haystack's prototype [14] was developed as one of the earliest statistical anomaly-based IDSs. The detection system considers a combination of two models: user behavior and generic group behavior. It takes into account a range of normal behavior events between two limits and each occasion has a score, with a high score demonstrating an anomaly. On the other hand, normal method features are takeout offline only.

- **Expert systems**

Expert systems **[**15**]** are knowledge-based and used to build the profile of a system or its users based on rules obtained from statistical measures of normal behavior over a period of time. Primarily, these approaches are intended for data classification according to the extracted rules. In the first stage, training data is used to define certain variables and classes, and then classification rules are inferred and applied to audit data.

- **Machine learning**

Learning is a process to learn the dependency between two sets of information to generate an unknown input-output model based on a limited number of inspections [16]. A perfect surveillance that explains the constructed model requires an accurate problem definition. Machine learning techniques have been used widely in computer systems to provide intelligence in the automatic process. The tasks of machine learning include: classification, acting and planning, and understanding. IDSs can be acknowledged as a classification problem (with two classes: normal and abnormal) [16].

- **Data-mining techniques**

Data-mining techniques have also been employed in anomaly detection systems in many researches to extract a knowledge model from a large number of prototypes. Association rules from the structure patterns are utilized to create features that construct the detection system. Two types of methods applied in data mining are 1) predictive methods involving certain variables to predict unknown variables; and 2) descriptive methods where human interpretation are used to detect unknown patterns.

International Journal of Computer Science
ISSN:2348-6600
Oddity...Probe...Reviste...
Published by SK Research Group of Companies(SKRGC)
http://www.ijcsjournal.com Volume 3, Issue 2, No 1, 2015. ISSN: 2348-6600
Reference ID: IJCS-096 PAGE NO: 549-555.

B. Signature-based detection

Signature-based detection methods are knowledge-based techniques where well-defined attack patterns are used to detect malicious security breaches. A new attack has to be considered and examined to identify its features and then generate its accurate signatures. The detection system observes and analyses activities amongst audit data, and the detection mechanism is based on the comparison between attack signatures and observed prototypes. Signatures can be described as a set of circumstances characterizing the direct manifestation of intrusion activities in terms of system calls and network data [17], which is to say that when these conditions are met, a type of intrusion event is indicated. In networks, unauthorized behavior is detected by sniffing packets and using the sniffed packets for analysis.

IV. Anomaly Based Detection Techniques

With the enormous and fast growth of network attacks, constructs to utilize of signature based or classification based detection techniques ineffective, costly, slow, and erratic. This involve that the only recovery for realistic and effective detection technique is the anomaly based detection. The following is a classification of those techniques anomaly based methods, with examples on each:

A. Statistical based methods

This method measures the behavior of certain variables over time. If the average or the variance of any of these variables deviates from a definite doorstep, then an anomaly is distinguished. This kind of anomaly detection technique is the simplest and usually it has a lot of false positive due to its permittivity. One example of this technique is Snort Spade which is a plug-in for the Snort IDS. What it does is that it monitors the TCP-SYN flag and finds the average of those in a period of time. If the average of those flags is beyond a predefined threshold then an anomaly alert is triggered and given a score. To reduce the number of false alarms, they prepared a rule for scoring. That rule is the fewer times a particular packet has occurred in the past, the higher the anomaly score will be occurred.

B. Distance based methods

In this method, the whole monitored space is divided into regions. Each of these regions is tagged with normal or abnormal flags. And a point inside that region is chosen to characterize the area. With the intention of numeral out if a assured point is belonging to certain region, the distances from that point to all normal regions' representing points are computed. The point is then considered to be belonging to a normal region if all the distances are below a certain predefined threshold. Another method is to discover all the distance from that point to all points that are known to be usual, if any of the distances move away from a confident threshold then the points are noticeable as uncharacteristic. The dimensions of space in this system are the supervised features. Consequently, the features should be stabilized to overcome the problem of being biased.

One example is MINDS (Minnesota Intrusion Detection System), in this intrusion detection system, netflow is monitored and the desired features are selected. The feature forms the space dimension. Then each monitored point (record) is compared to all the known normal points, and given a score based on the distance to those points. Then analysts start looking at the connections with the highest scores and determine if they are abnormal.

One problem with this technique is that it requires a huge processing power, which makes the system hard to implement in real time environment. Also it is hard to distinguish small region of behaviors that fits into a region of different type of behavior.

The following diagram explains that. Where point **a** represent region **A** and point **b** represent behavior region **B**. When trying to find where point **c** belong, we find it has almost the same distance to both points **a** and **b**.

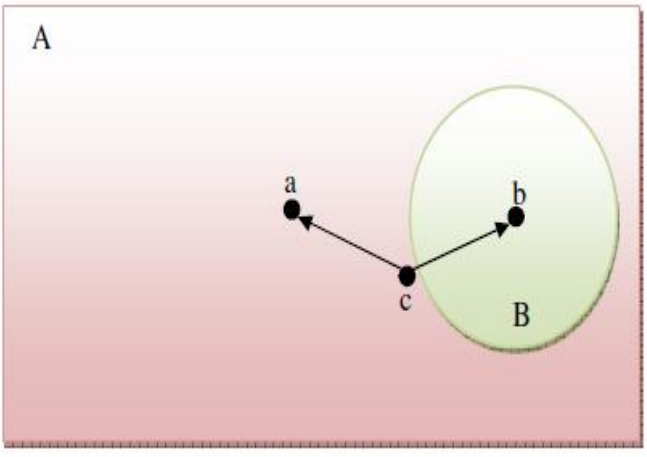

Figure 2.3: Distance Based decision problem

International Journal of Computer Science
ISSN:2348-6600
Oddity...Probe...Reviste...
Published by SK Research Group of Companies(SKRGC)
http://www.ijcsjournal.com Volume 3, Issue 2, No 1, 2015. ISSN: 2348-6600
Reference ID: IJCS-096 PAGE NO: 549-555.

C. Rule based methods

In this method, the normal behavior profile is captured in a set of rules. Every one of the rules is measured as a region of performance. Contrasting the distance based technique, this process doesn't require a enormous processing power in detection, also the process of making decision is easy even when there are two different behavior regions one inside the other one.

Wisdom and Sense (W&S), a computer security anomaly detection system developed at Los Alamos National Laboratory (LANL) is one example on the rule based behavior analysis system. Automatically generates rules (Wisdom) from past data. W&S recognize computer operations that are at variance with historically established usage patterns.

In general a rule based system generates its set of rules from a pre-tagged normal and abnormal. The set of rules generated are all normal rules. While responsibility the detection nobody of the standard rules are triggered then the record is marked as abnormal.

D. Profile based methods

In these methods, the normal behaviors are captured by records the past warehouse. And then data mining methods are used to find if the current records have formerly been distinguished. It necessitates an enormous amount of data to evaluate. And the evaluation procedure takes a long time and enforces a lot of overhead on the system.

V. LITERATURE SURVEY

Zhang et al. [18] introduced a Network Intrusion Detection System (NIDS) which used SVM method to identify the scan based on RST to reduce the number of features from 41 to 29. They skilled SVM using assorted frequencies of normal and attack packets which were collect from the network in every 4 seconds. Thus, SVM was proficient to categorize whether the incoming unknown packet was usual or attack. With this move toward, they maintained to appropriately detect 95% of the attack packets. Their proposed system used two packages, WINPCAP and JPCAP; to capture all the packets in the network for investigation. 14 features were removed for pre-processing in their implementation, and they have used packet sniffer to store network packet in sequence. Feature Selection was done using RST technique which helped in reducing the dimensionality of the patterns. Their representation recognized the intrusion by profiling normal network behavior with various attack behavior by using PCA detection.

Kholidy approach describes how to extend the current technology and IDS systems. His proposal is based on hierarchical IDS [19]) to experimentally detect DDoS, host-based, network based and masquerade attacks. It provides capabilities for self-resilience preventing illegal security event updates on data storage and avoiding single point of failure across multiple instances of intrusion detection components.

His proposal consists on a hierarchical arrangement, autonomic and cloud-based, make longer his earlier work [19] with features such as autonomic response and forecast. Especially, it evaluates vulnerabilities and risks in the system through a mechanism that builds a security model based on risk assessment and security event strategies seriously. It also makes available the opportunity of automatic response to actions based on a set of policies defined by the system administrator. On the other hand, a black box format does not elucidate possible answers or makes clear how to choose the best answer leaving that decision to a system administrator.

Anomaly-based intrusion detection systems organize network traffic occurrences by evaluates them with a representation of the standard network behavior. To be efficient, such methods are anticipated to accurately detect intrusions i.e. high true positive rate while containing the number of false alarms i.e. low false positive rate. Alternatively, there subsist a typical trade-off between detecting all anomalies, and missing anomalies. In this paper [20], author has presented a new approach an autonomic approach for tuning the parameters of anomaly-based intrusion detection systems in case of SSH traffic. In exacting, author concern is twofold. First, they stress the consequence of tuning the constraints of an anomaly detection system to accomplish an optimal concert period. We anticipate that our analysis can bring new understanding while generating or organizing a new anomaly-based system. In the paper, here they propose to new move toward the parameter-tuning crisis by formalizing it in expressions of an optimization process. The optimization process permits to explicitly relating the system constraints and the performance determines in the form of an optimization difficulty. Proposed process that aims to repeatedly alter the arrangement restrictions and, by doing so to optimize the structure presentation. A key distinctive of the proposed way out is that it considers optimality according to high-level policies.

International Journal of Computer Science
ISSN:2348-6600
Oddity...Probe...Reviste...
Published by SK Research Group of Companies(SKRGC)
http://www.ijcsjournal.com Volume 3, Issue 2, No 1, 2015. ISSN: 2348-6600
Reference ID: IJCS-096 PAGE NO: 549-555.

Second, in this paper authors [20] try to show how to appropriate the optimization procedure to HMM-based IDS. In this part of procedure, the optimization problem consists in maximizing the number of properly classified inspection series. As per analysis, the rule states the comparative consequence of detecting all the attacks against maintaining the false alarm rate small. They authenticate our approach by search it on a flow-based probabilistic detection method for the detection of SSH attacks. Our authentication exhibited that, by unpredictable the comparative consequence, we are capable to fine-tune the method to favor either the detection of all the anomalies or the detection of attacks only when they are promised. Our discovering also show that the relative importance has impact on the detection rate and on how appropriate the system is able to become aware of an attack or make progress from it. Accordingly think about that, when communicating a process guiding standard in terms of the relative consequence it should be taken into account that such a policy concern the system performance in multiple performances. The variety of which would be an suitable policy is left to the system manager.

In this paper [21] author has examines real-time intrusion response systems with the purpose of mitigate attacks that cooperation reliability, privacy and accessibility in cloud computing policies. This paper recommends the exploit of autonomic computing to provide reply to attacks on cloud computing environments. Thus, it is probable to make available self-awareness, self configuration and self-curative in the cloud. A structural design that uses the anticipated effectiveness function for deciding a suitable answer is a statistical representation to alter the answers given with the intention of provides more consequences.

To accomplish this objective, author has proposed [21] IRAS, an Intrusion Response Autonomic System, using Big Data methods for data analytics and anticipated utility function for decision satisfying. Additionally, the effort proposes the use of Big Data communications using Hadoop to manage the large quantity of data and take out information using the Map-Reduce structure. Thus, we might make available intrusion detection, respond and self-curing in cloud environment.

VI. Conclusion

Anomaly detection based on externally has always been a difficult job for real-time detection. When it concern with protection challenges of network technologies and global internet services our assumption was that anomaly based intrusion detection could be relevant as a second line of defense in both the effect of alteration the constraints of an anomaly detection system to complete a most favorable concert era. On the other hand, the use of a learning based anomaly detection method, which permits the characterization of all the discrepancies of the system regularity, has proved to be efficient. During the literature review, it was examined that in recent times, many researchers were and is still achieving their research to enhance the use of anomaly based intrusion detection.

.

References

[1] Matt Bishop. Computer Security Art and Science. 2003.

[2] K. Lumpur, "An investigation and survey of response options for Intrusion Response Systems (IRSs)," 2010.

[3] N. Stakhanova, S. Basu, and J. Wong, "A taxonomy of intrusion response systems," International Journal of Information and Computer Security, vol. 1, no. 1, pp. 169–184, 2007.

[4] P. Horn, "Autonomic computing: IBM's perspective on the state of information technology," Computing Systems, vol. 15, no. Jan, p. 140, 2001.

[5] E. Bertino, A. Kamra, E. Terzi, and A. Vakali, "Intrusion detection in rbac-administered databases," in Proc. 21st Annual Computer Security Applications Conference. Washington, DC, USA: IEEE Computer Society, 2005, pp. 170-182.

[6] S.W. Lodin and C.L. Schuba, "Firewalls fend off invasions from the Net," *IEEE Spectrum*, vol. 35, no. 2, 1998, pp. 26-34.

[7] A. Lazarevic, L. Ertoz, V. Kumar, A. Ozgur and J. Srivastava, "A comparative study of anomaly detection schemes in network intrusion detection," *Proc. the Third SIAM International Conference on Data Mining* 2003, pp. 25–36.

[8] V. Kumar, J. Srivastava and A. Lazarevic, "Intrusion Detection: A Survey," *Managing Cyber Threats*, Massive Computing 5, Springer US, 2005, pp. 19-78.

[9] R. Bace and P. Mell, *Intrusion detection systems*, US Dept. of Commerce, Technology Administration, National Institute of Standards and Technology, 2001.

[10] I. Ristic, *Apache security*, O'Reilly Media, Inc., 2005.

[11] S. Axelsson, *Intrusion detection systems: A survey and taxonomy*, Technical Report, Chalmers University of Technology, Dept. of Computer Engineering, 2000.

[12] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Macia-Fernandez and E. Vazquez, "Anomaly-based network intrusion detection: Techniques, systems and challenges," *Computers & Security*, vol. 28, no. 1-2, 2009, pp. 18-28.

International Journal of Computer Science
ISSN:2348-6600
Oddity...Probe...Reviste...
Published by SK Research Group of Companies(SKRGC)
http://www.ijcsjournal.com Volume 3, Issue 2, No 1, 2015. ISSN: 2348-6600
Reference ID: IJCS-096 PAGE NO: 549-555.

[13] S. Zanero, ―Detecting 0-day attacks with learning intrusion detection system,‖ *Blackhat Briefings, USA*, 2004.

[14] S.E. Smaha, T.A.S. Inc and T.X. Austin, "Haystack: An intrusion detection system," *Proc. the IEEE fourth Aerospace Computer Security Applications Conference*, IEEE Computer Society Press, 1988, pp. 37-44.

[15] T.F. Lunt, A. Tamaru, F. Gilham, R. Jagannathan, P.G. Neumann and C. Jalali, "IDES: a progress report [Intrusion-Detection Expert System]," *Proc. the Sixth Annual Computer Security Applications Conference*, IEEE Computer Society Press, 1990, pp. 273-285.

[16] M. Bishop, *Introduction to computer security*, Addison-Wesley Professional, 2004.

[17] D.F. Gong, "White Paper: Deciphering Detection Techniques: Part II Anomaly-based Intrusion Detection," *Network Associates (McAfee Security)*, 2003.

[18] X. Zhang, L. Jia, H. Shi, Z. Tang, and X. Wang, "The Application of Machine Learning Methods to Intrusion Detection," in 2012 Spring Congress on Engineering and Technology (S-CET), 2012, pp. 1-4.

[19] H. A. Kholidy, A. Erradi, S. Abdelwahed, and F. Baiardi, "Ha-cids: A hierarchical and autonomous ids for cloud systems," in Computational Intelligence, Communication Systems and Networks (CICSyN), 2013 Fifth International Conference on, pp. 179–184, IEEE, 2013.

[20] Sperotto, Michel Mandjes, Ramin Sadre, Pieter-Tjerk de Boer, and Aiko Pras, "Autonomic Parameter Tuning of Anomaly-Based IDSs: an SSH Case Study" IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT, VOL. 9, NO. 2, JUNE 2012.

[21] Kleber M.M. Vieira, Fernando Schubert, Guilherme A. Geronimo, Rafael de Souza Mendes, Carlos B. Westphall, "Autonomic Intrusion Detection System in Cloud Computing with Big Data" 2014.