Use of Big Data for Improving Real-time Anomaly-based Intrusion Detection

Contributors: Chenfeng Nie, Eric Schlieber, Erwen Shan

1. Abstract

This paper focuses on the use of big data to improve the overall performance of anomaly based intrusion detection system (A-NIDS). Key metrics for improvement include detection accuracy, efficiency, and usability of the modern IDS. We have reviewed several papers to obtain an overview of intrusion detection systems, big data's applicability to intrusion detection, the challenges of big data, and critiques for using pre-built datasets for developing IDS. While big data presents opportunities for growth in the field of IDS, there are still many challenges that must be overcome. One recommended solution is the use of Rathore and Paul's four-layer real-time anomaly intrusion detection system prototype. To evaluate our solution, we trained, tested and evaluated our model by using KDD Cup 99 and DARPA dataset and Euclidean distance and Mahalanobis distance techniques.

KEYWORDS: Machine learning, Real-time Intrusion Detection System, Big Data

2. Introduction

This report begins with an overview of intrusion detection systems(IDS) then narrows its focus to the use of big data for anomaly based intrusion detection techniques. This paper addresses the issues and possible solutions to the challenge of big data usage for anomaly based intrusion detection. Although techniques using big data have already proven to be effective, they still face significant challenges that limit their current effectiveness.

Intrusion detection systems can be classified in a variety of ways based on their detection method and efficiency. The foundation for the modern field of intrusion detection began in 1987 with Denning's intrusion detection model. This model presented a rule-based pattern matching system that contained six main components: subjects, objects, audit records, profiles, anomaly records, and activity rules.[1] The general concept behind Deming's model is that the system monitors normal operations and looks for any deviations in usage.

Another important concept from Deming's foundational work includes the description of activity profile, which creates a standard of behavior for the subject(s) and object(s).[2] This concept has continued in the use of current IDS. The behavior that constructs this profile is

[1] , Dorothy E. Denning, "An Intrusion Detection Model," IEEE Transactions on Software Engineering," Volume SE-13, No. 2, February 1987, 223.
[2] Denning, 224.

characterized by a statistical model and metrics, and it uses information from audit records to detect abnormal behavior.[3]

The "Revised Taxonomy for Intrusion-Detection Systems" is from 2000 and also provides an overview into the classification of intrusion detection systems. It first proposes that there had previously been three critical metrics that can be used to measure the efficiency of intrusion detection systems. The first metric is accuracy, which is the ability to properly detect actual attacks from false alarms. The next is performance, which measures the rate that the system can process events, and the third is completeness which measures the ability to detect all attacks.[4] This paper introduced two additional metrics, the first being fault tolerance which measures the system's ability to ward off against attacks. The second is timeliness, which measures not only the speed of the IDS but the time required to publish the information and react to it.[5] These five metrics can be used to measure the system.

In addition, this paper also provides a taxonomy of five concepts to classify a system: detection method, behavior on detection, audit source location, detection paradigm, and usage frequency.[6] Detection methods are classified as either behavior based, which uses normal system behavior and is now more commonly known as anomaly detection. Knowledge based which uses attack information and is now more commonly known as misuse detection. This paper focuses on anomaly based intrusion detection.

Anomaly based intrusion detection Systems (A-NIDS)

Anomaly based intrusion detection focuses on establishing a baseline of normal behavior, then comparing it against the desired traffic. The main challenge to this process lies within creation of accurate training data within the detection analysis phase. Appropriate feature selection and extraction are critical to creating a representative dataset during this phase.

Although specific A-NIDS have nuanced differences, all adhere to the same functional model for execution. The basic stages are as follows[7]:

1. Parameterization: gathering data in its raw form.
2. Training - data is organized and translated into a model.
3. Detection - the system model is compared against the parameterized (observed) traffic.

A-NIDS can also be classified into three main categories:

1. Statistical-based: system behavior is seen from a random perspective
2. Knowledge-based: captures "claimed behaviour from available system data (protocol specifications, network traffic instances, etc.)"
3. Machine-learning based: establishes models that allow for pattern recognition[8].

---

[3] Denning, 222.
[4] Debar, H., Dacier, M. & Wespi, A. Ann. Télécommun. (2000) 55: 361.
[5] Debar, 363.
[6] Debar, 363.
[7] P. Garcia-Teodoro, Diaz-Verdejo, Macia-Fernandez, et. al, "Anomaly-Based Network Intrusion Detection: Techniques, Systems, and Challenges," Computers and Security (28), 2009, 19.
[8] Garcia-Teodoro, 20.

In comparison to misuse detection, anomaly based intrusion detection systems are better at detecting new, or zero-day, attacks. In addition, they can help to detect “abuse of privilege” or “escalation of privilege” types of attacks.[9] These attacks may not exploit any specific vulnerability and may be more difficult for misuse systems to detect.

Anomaly based IDS is usually based on statistical measurements. Network data is captured, and then a stochastic, statistical, or probabilistic profile of behavior is created.[10] Traffic that differs from the calculated baseline will be flagged as an intrusion, and will either be a true positive or a false positive. Machine learning has greatly contributed to the performance of IDS. These systems can use Bayesian networks, Markov models, or other techniques.[11]

Expansive networks face an overwhelming amount of data and must use traffic classification as a critical component to building an effective IDS. The TCP/IP model provides a framework to separate the traffic into data flows, which can then be associated with applications on a host. These data flows are a sequence of IP packets that share the same basic characteristics, that can be associated with an application running on a host.[12] As organizations face an increasing amount of data, big data can be leveraged to improve IDS.

3. Literature Review

General Challenges to Anomaly Detection

Anomaly detection has many general challenges, including the concept that what is anomalous is not necessarily bad.[13] Just because a behavior is rare or unexpected does not mean that it is malicious. Detecting too many of these rare behaviors lead to the over identification of false positives. Another challenge is defining normal traffic is highly complex systems can be both variable and highly complex. It can change depending on the time period, with certain times being much busier than others due to numerous variables. This means that the baseline for normal behavior is constantly changing, which adds to the complexity.[14]

Another challenging aspect of anomaly detection is the difficulty in continuously updating the system to maintain the standard of normal system behaviour. Periodic updates can create additional overhead in the system, and may even lead to security vulnerabilities.[15] In contrast, misuse detection systems add new threats to their database, which is a less intensive task from a workload perspective. Advances in the field of data analytics can assist with this challenge.

---

[9] Dacier, 365.
[10] Awais Ahmad, Anand Paul, Mazhar Rathore, “Real Time Intrusion Detection System for Ultra-High speed Data Environments,” Springer: 23 Feb. 2016, 3493.
[11] Ahmad, 3493.
[12] Mark R. Heckman, “The Promise and Limits of Big Data for Improving Intrusion Detection,” ISSA Journal, June 2017.
[13] Heckman, 29.
[14] Heckman, 28.
[15] Ahmad, 3493.

Big Data and IDS

When describing big data, it is important to make the distinction between data and information. Data consists of objective records, statistics, or other empirical observations, while information is the useful interpretation of data. Big data techniques, therefore, permit the collection, storage, and processing of data to produce useful information. For example, authentication events in a log may be seen as data, but the processing of this to draw a conclusion would be the information. All of these steps contain challenges that are continually being improved upon to make the use of big data more efficient.

Until relatively recently, the efficient use of big data may not have been feasible due to a variety of factors; however, that is no longer the case. Decreasing cost of storage, increase in CPU power, cloud computing, and tools such as Hadoop have exponentially increased an organization's ability to feasibly employ the use of big data.[16] The ability to more easily store and process big data has expanded its feasibility for use in intrusion detection. From an economic standpoint, a greater number of organizations can now afford to both store and analyze big data. The data can soon become overwhelming, however. Large organizations may have trouble both storing and processing the information. For example, in 2013 Hewlett Packard (HP) estimated that it generated one trillion events per day, or twelve million events per second.As the numbers grow larger, existing techniques become less effective and lead to a greater number of false positives.[17]

Big Data's applicability to Intrusion Detection

Big data has presented new opportunities within the field of intrusion detection. In "The Promise and Limits of Big Data for Improving Intrusion Detection," Heckman asserts that although big data has solved many problems for intrusion detection systems, there are still significant challenges to overcome. Techniques such as data mining and machine learning are associated with the use of big data, and have already proven to be effective in improving IDS.[18] One of the central improvements is in improving the false negative and false positive reports.

Big Data techniques assist with IDS through assisting with the collection, storage, and analysis of large data sets. Anomaly based detection has always used this process to determine threats, so better improvements in big data have been able to assist with the process. The movement of data throughout a system can be classified in terms of volume, velocity, and variety. Techniques in this field involve the use of Hadoop and Hive applications, which allow for the search and organization within this vast quantity of data.[19]

Even within the more recent research involving big data and machine learning, much of Denning's foundational work is still used. Her metrics and statistical models for anomaly

---

[16] Cloud Security Alliance, "Big Data Analytics for Security Intelligence," www.cloudsecurityalliance.org/research/big-data, 2013.
[17] Cloud Security Alliance.
[18] Mark R. Heckman, "The Promise and Limits of Big Data for Improving Intrusion Detection," ISSA Journal, June 2017, 22.
[19] Heckman, 29.

detection still provide the basis for these methods. The statistical models used include the operational, mean and standard deviation, multivariate, Markov process, and time series.[20]

Big Data - Challenges

Drawing meaningful relationships from within such large amounts of data is a very difficult and time consuming task. Machine learning techniques have been able to assist with this process by the use of algorithms to detect patterns and the creation of analytical models.[21] One classification of these systems is “supervised,” which requires human input to label whether or not data is malicious. In contrast, unsupervised systems allow the machine to identify data clusters on its own, which can be used to label it as “normal” or “anomalous”.[22]

Controversy in Using Pre-built Dataset for Developing IDS

Training datasets are extremely vital and necessary because the quality of dataset affects several crucial parameters such as attack detection rate, false positive rate and overall accuracy rate directly during machine learning process. In this case, there is controversy focusing on building your own dataset and using a pre-built dataset such as KDD Cup dataset to train, verify, and test machine learning models.

KDD Cup 1999 dataset[23] (The 1999 Knowledge Discovery and Data Mining Tools Competition), original constructed from 1998 DARPA Intrusion Detection Evaluation, is one of the most popular pre-built datasets which are widely used among the network security community in last decade. The dataset has two parts, the training dataset and testing dataset. For the training dataset, it contains about 4.9 million connections extracted from seven weeks of traffic, and for the test data, it contains around 300,000 connections of traffic in two weeks. A lot of researchers and companies have used the KDD Cup 99 dataset to evaluate and develop machine learning techniques for their intrusion detection systems since this data set provides a common platform for NIDS evaluation.

To develop an effective IDS, we must understand various potential types of attacks aimed at the valuable system that we need to protect. In the report “A Scheme for Building A Dataset for Intrusion Detection Systems “, it describes four main categories of attacks, including Denial of Service (DOS), unauthorized access from a remote machine (R2L), unauthorized access to local superuser privileges (U2R) and surveillance and other probing (Probe).

Cao, Hoang and Nguyen[24] introduce a feasible scheme for building a dataset for Intrusion Detection Systems. Briefly, there are three main steps for the implementation to collect and build the dataset. First, they used a packet sniffer (e.g. WinPcap in their study) to capture

---

[20] Heckman, 27.
[21] Heckman, 30.
[22] Heckman, 44.
[23] KDD Cup 1999 data (Computer network intrusion detection). http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html.
[24] V. L. Cao, V. T. Hoang and Q. U. Nguyen, "A scheme for building a dataset for intrusion detection systems," 2013 Third World Congress on Information and Communication Technologies (WICT 2013),

all traffic flow packets generated by background traffic and simulated attack sessions. After that, they extracted both protocol and payload features(e.g. 16-feature vectors used in their research) from traffic flow dumps. Since the collected dataset is too enormous to analyze and evaluate in real time or short time, it is extremely vital to modify, reduce, and extract critical information and features to build up our dataset. Finally, they labeled each feature vector as either a normal or intrusive connection in order to make the dataset be ready for machine learning algorithm.

However, as the report “A Scheme for Building A Dataset for Intrusion Detection Systems” [25] mentioned, more and more researchers criticize that KDD Cup 99 dataset has a few limitations, especially in recent years. For instance, Sabhani and Serpen[26] indicate that there is a problem when applying an machine learning technique for two rare classes in KDD Cup datasets (e.g., R2L and U2R) in misuse detection context due to the difference between target hypotheses in training and testing dataset. Moreover, as Cao, Hoang and Nguyen[27] pointed out, KDD Cup 99 dataset may become gradually out-dated because of the rapid development of network technologies and computing power. Thus, it would be more and more unreliable and irresponsible to continue using the KDD Cup 99 dataset to develop, test and evaluate state-of-the-art intrusion detection systems.

4. Problem Definition

By now, most intrusion detection systems are hand-crafted, aka signature IDS. This type of IDS has worked well for the past decades, yet as the capability to process data volume, various data features, and diverse new attacks without known signatures increases, signature based IDS has a more difficult time surviving in the era of big data. Thus, a contemporary and proper anomaly intrusion detection system model must be chosen to improve the overall performance metrics of IDS.[28] However, as Lee, Stolfo, Chan, Eskin, Fan, Mille, Hershkop, and Zhang[29] mention, despite anomaly IDS’s ability to detect unknown attacks in comparison to signature-based IDS, it has several inherent drawbacks as mentioned before, briefly in three key areas: 1. Accuracy may lower than signature-based IDS, for instance, anomaly-based IDS usually have higher false positive rates; 2. Efficiency may leave much to be desired due to high computationally expense during collecting, training and evaluating procedures; 3. Usability needs to be considered due to the large amount and complexity of the dataset. In our paper, we will find a feasible solution to overcome these addressed problems.

---

[25]Van Loi Cao, Van Thuy Hoang, Quang Uy Nguyen, “A Scheme for Building A Dataset for Intrusion Detection Systems,” 2013 Third World Congress on Information and Communication Technologies (WICT).

[26]Maheshkumar Sabhnani and G'ursel Serpen, "Why machine learning algorithms fail in misuse detection on KDD intrusion detection data set". Intelligent Data Analysis, 8(4):403-415, 2004

[27]V. L. Cao, V. T. Hoang and Q. U. Nguyen, "A scheme for building a dataset for intrusion detection systems," 2013 Third World Congress on Information and Communication Technologies (WICT 2013), Hanoi, 2013, pp. 280-284. doi: 10.1109/WICT.2013.7113149

[28] Heckman, 45.

[29] Lee, Stolfo, Chan, Eskin, Fan, Mille, Hershkop, and Zhang, “Real Time Data Mining-based Intrusion Detection,” IEEE, 2001.

5. Solution and Discussion

Based on our research, implementing a real-time data mining-based IDS is a viable solution that could be applied to improve anomaly-based intrusion detection by using Big Data methodology. There are plenty of research paper and work depicted the concepts related to real-time anomaly IDS topic. And after researching a couple of paper, we noticed that the high-level concepts for a real-time anomaly IDS are quite similar. In this case, we pick Rathore and Paul's[30] real-time intrusion detection prototype as an example to demonstrate how it could solve addressed problems(figure 1 shown below).

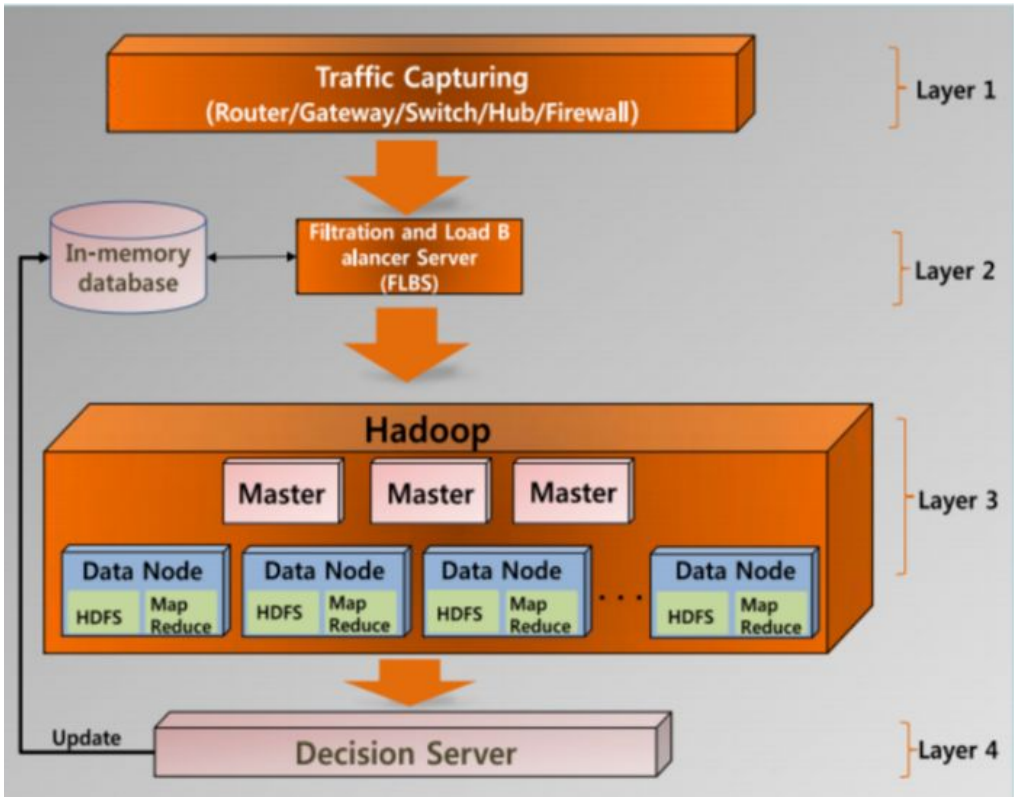

Figure 1. Proposed Architecture for Real-time IDS

The proposed real-time anomaly IDS architecture is an ecosystem with four layers, namely data capturing layer, data filtering and sorting layer, data feature extraction layer, and decision maker layer. These are designed to improve detection accuracy, efficiency and usability of modern anomaly IDS. We will describe them in detail below.

I. Data capturing layer

[30] Lee, Stolfo, Chan, Eskin, Fan, Mille, Hershkop, and Zhang, "Real Time Data Mining-based Intrusion Detection," IEEE, 2001.

In the data capturing layer, there are multiple high-speed network traffic capturing devices, sniffers and sensors deployed for capturing and collecting huge amount of raw data transmitted through network terminals such as routers, gateways, switches, hubs and firewalls. Later on, all captured raw data will be delivered to data filtering layer.

II. Data filtering and sorting layer

In data filtering layer, there is a server or sensor used for filtering and sorting raw data from previous layer. For filtering functionality, server or sensor observes all raw data and only keeps unascertainable traffic connections and packets, in other words, the traffic flows that cannot be distinguished as normal activity or malicious intrusion by analysing its meta and header information based on its knowledge database. The knowledge database is worth to be mentioned, it has some preset rules and standards for determining normal and malicious traffic flow and the knowledge would update automatically after machine learning process. The knowledge database also facilitates the integration of data from multiple sniffers and sensors in layer 1. After that, the filtered unidentified data flows/packets will be transmitted to data feature extraction layer.

III. Data feature extraction layer

Hadoop is one of the most popular and powerful technologies used widely in big data processing area. Most researchers and companies use Hadoop to preprocess their raw data for building their anomaly IDSs. Briefly, Hadoop framework provides two main features, namely, Hadoop Distributed File System(HDFS) for storing large amount of raw data files and the MapReduce functionality for reducing frequently occurring large-scale data.

In order to increase overall efficiency and usability and also achieve the real-time IDS objective, Rathore and Paul's model[31] has a notable difference in comparison to other similar models. They use hadoop with multiple so called master nodes and data nodes to preprocess different types of data flow and packets filtered and sorted by their ip address from Filtration and Load Balance Server in data filtering and sorting layer. Master nodes are designed to extract the necessary information from each filtered data flow and packet by using certain network packet format such as Pcap-Input Format, then write them into sequence file and send them to data nodes. Data nodes will receive large-scale sequence file and store them into HDFS, after that, running customized MapReduce function to calculate the traffic features by reading sequence files line by line in multiple parallel processes at real time. Once data flows and packets are processed by data nodes, feature values are transmitted to Decision maker layer for machine learning approaches.

IV. Decision maker layer

The main purpose of the decision maker layer is to facilitate the rapid development and real-time updating anomaly IDS. By now, we have acquired enough dataset prepared for machine learning. In decision maker layer, there is a machine learning system or server

[31] Lee, Stolfo, Chan, Eskin, Fan, Mille, Hershkop, and Zhang, "Real Time Data Mining-based Intrusion Detection," IEEE, 2001.

deployed for classifying normal and malicious data flows and packets based on their feature values by applying various machine learning algorithms and classifiers such as Random Forest and SVM(support vector machines). Generated results will be sent to knowledge database and update in real-time so that increase overall efficiency and detection accuracy for servers or sensors in data filtering and sorting layer.

After describing overview of this proposed architecture, we start to simulate, demonstrate, and explain why this model can be used to solve the problems addressed in Problem Definition section, in other words, how this model improve overall performance included accuracy, efficiency and usability of a modern anomaly IDS. In this theoretical prototype, there are two scenarios can be simulated hypothetically.

Scenario 1: when a suspicious intrusion or malicious flow or packet was detected and identified in data filtering and sorting layer, in other words, the exemplary features of certain data flow or packet was matched and indicated with knowledge base, the IDS will send alarms and reports to IT department and related people. After that all data features of this certain flow or packet such as meta information and header information will be processed by decision maker server or system in decision maker layer and update to knowledge base in order to increase detection accuracy if data filter server or system detects intrusion with similar exemplary feature in future detection

Scenario 2: when an intrusion or malicious flow or packet was not able to be detected or identified in data filtering and sorting layer. It is also called false positive. This flow or packet would be sorted and sent to decision maker layer, after machine learning approach processed by the decision maker server or system, when it predicts this flow or packet has high possibility as an intrusion based on huge data volume trained before, it automatically modify the model that can detect the new unknown malicious intrusion and update its knowledge database at real-time. Thus, the sensors and filtering server will detect this kind of intrusion at next time. Moreover, the real-time anomaly IDS will improve its accuracy continually and automatically with more and more data trained and tested because big data will reveal and classify the exemplary features of malicious intrusion from obscure to clear state, which improve the IDS's overall accuracy.

To sum up, more and more researchers and companies interested in data mining-based mechanism to build and optimize their real-time intrusion detection model in recent years. And the real-time data mining-based IDS would become one of the most effective model which could solve the addressed problems in Problem Definition section theoretical ideally.

6. Evaluation and results

Now we put forward some possible ways to evaluate our model. First we can perform our results on a Naive Bayesian Classifier to evaluate our model. Elementary, we can self-test our result through evaluating the extracted data with raw training data sets. And then we can use KDD 99 and DARPA dataset for testing purpose. While testing our results, we can compare our model's processing time, efficiency and performance to the existing best results of the KDD

99 pre-built dataset. And we can also evaluate the model in the exhibition of the detection tested with the DARPA 1999 Evaluation data.

From another point of view, we can evaluate our lab performance using two kinds of distance metrics:Euclidean distance and Mahalanobis distance.[32] We can use Euclidean distance to calculate continuous data and use the Mahalanobis distance to take the dependency among variables into account. We can evaluate our data collection and extraction result through calculate and compare these two kinds of distance metrics of the data.

What's more, Due to the lack of academic research and experiment results, we may can performance the evaluation of results based on comparing to other similar attack evaluations such as SQL injection attacks.

7. Summary

The main purpose of our research is to present a novel method for collecting real ground-truth training data and doing efficient data extraction for intrusion detection system.

Efficient data collection and extraction is undoubtedly significant. However, from early research, a lot of academic researches were not able to evaluate their proposals due to the lack of ”ground-truth” training data. Most of them used an old data set originated from the 1998 DARPA Intrusion Detection Program, which is adopted in the Data-Mining and Knowledge Discovery (KDD) competition, and investigated the impact of features selection process on the classification accuracy artifacts[33].

We manage an optimized method to do the data collection and extraction through the four layers of Data capturing layer, Data filtering and sorting layer, Data feature extraction layer and Decision maker layer. We are trying to solve the problem to make the results more accurate, more usable and cost little. Throughout the whole study process, we become more familiar with the overall concept of modern Intrusion Detection Systems and the from-front-to-end procedure. We are using big data technique for improving the Intrusion Detection System, in order to find a solution to maximize overall accuracy and efficiency by detecting attacks and protecting our targeted information systems.

For future work, we would like to further improve our system to enhance its ability to handle an increased amount of data, conduct more feature selections and hence be more powerful and more efficient. Furthermore, we are interested in combining our work in studying the impact of real-work traffic parameters in the future.

---

[32] Dina Said, Lisa Stirling, Peter Federolf, and Ken Barker, “Data Preprocessing for Distance-based Unsupervised Intrusion Detection,” University of Calgary, 2011.

[33] Abdelhamid Makiou, Ahmed Serhrouchni, “Efficient Training Data Extraction Framework for Intrusion Detection Systems,” IEEE, 2015.

References

1. Abdelhamid MAKIOU & Ahmed SERHROUCHNI “Efficient Training Data Extraction Framework for Intrusion Detection Systems” *2015 International Conference on the Network of the Future, NOF 2015*, November 20, 2015

2. Ahmad, A., Paul, A., & Rathore, M.M. (2015). “Real time intrusion detection system for ultra-high-speed big data environments.” The Journal of Supercomputing, 72, 3489-3510.

3. Cloud Security Alliance, “Big Data Analytics for Security Intelligence,” www.cloudsecurityalliance.org/research/big-data, 2013.

4. Debar, H., Dacier, M. & Wespi, A. Ann. Télécommun. (2000) 55..

5. Denning, Dorothy E., “An Intrusion Detection Model,” IEEE Transactions on Software Engineering,” Volume SE-13, No. 2, February 1987, 223.

6. Dina Said∗, Lisa Stirling†, Peter Federolf†, and Ken Barker∗ “Data Preprocessing for Distance-based Unsupervised Intrusion Detection” *2011 9th Annual International Conference on Privacy, Security and Trust, PST 2011*, p 181-188,

7. Garcia-Teodoro, P., Diaz-Verdejo, Macia-Fernandez, et. al, “Anomaly-Based Network Intrusion Detection: Techniques, Systems, and Challenges,” Computers and Security (28), 2009.

8. Heckman, Mark R. “The Promise and Limits of Big Data for Improving Intrusion Detection,” ISSA Journal, June 2017.

9. Lee, Stolfo, Chan, Eskin, Fan, Mille, Hershkop, and Zhang, “Real Time Data Mining-based Intrusion Detection,” IEEE, 2001.

10. Makiou, Abdelhamid , Serhrouchni, Ahmed, “Efficient Training Data Extraction Framework for Intrusion Detection Systems,” IEEE, 2015.

11. Sabhnani, Maheshkumar and Serpen, G'ursel, "Why machine learning algorithms fail in misuse detection on KDD intrusion detection data set". Intelligent Data Analysis, 8(4):403-415, 2004.

12. Said,Dina, Stirling, Lisa, et al, “Data Preprocessing for Distance-based Unsupervised Intrusion Detection,” University of Calgary, 2011.

13. V. L. Cao, V. T. Hoang and Q. U. Nguyen, "A scheme for building a dataset for intrusion detection systems," *2013 Third World Congress on Information and Communication Technologies (WICT 2013)*, Hanoi, 2013, pp. 280-284. doi: 10.1109/WICT.2013.7113149

URL:http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7113149&isnumber=7113092

14. Wenke Lee *et al*., "Real time data mining-based intrusion detection," *DARPA Information Survivability Conference & Exposition II, 2001. DISCEX '01. Proceedings*, Anaheim, CA, 2001, pp. 89-100 vol.1. doi: 10.1109/DISCEX.2001.932195 URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=932195&isnumber=20170