IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 9, September 2024

DOI: 10.17148/IJARCCE.2024.13926

THREAT INTELLIGENCE INTEGRATION BEHAVIOUR ANALYSIS OF NETWORK INTRUSION DETECTION CONTROL SYSTEMS IN INDUSTRIAL SECTOR

Mr. Bijjam Venkateswara Reddy[1], Dr. N.Pughazendi[2]

MCA, CMR University, Bengaluru[1]

Professor & Head, School of Science and Computer Studies, CMR University,Bengaluru[2]

Abstract: The tremendous growth of the usage of computers over the network and development in applications running on various platforms captures the attention toward network security. This paradigm exploits security vulnerabilities on all computer systems that are technically difficult and expensive to solve. Hence intrusion is used as a key to compromise the integrity, availability, and confidentiality of a computer resource. The Intrusion Detection System (IDS) plays a vital role in detecting anomalies and attacks in the network. In this work, the data mining concept is integrated with IDS to identify the relevant, hidden data ofinterest for the user effectively and with less execution time. Four issues such as Classification of Data, High Level of Human Interaction, Lack of Labeled Data, and Effectiveness of Distributed Denial of Service Attack are being solved using the proposed algorithms like the EDADT algorithm, Hybrid IDS model, Semi-Supervised Approach, and Varying HOPERAA Algorithm respectively. Our proposed algorithm has been tested using the KDD Cup dataset. The entire proposed algorithm shows better accuracy and reduced false alarm rate when compared with existing algorithms. Threat intelligence integration is a critical component of modern cyber security strategies, helping organizations stay one step ahead of cyber threats and minimize security risks. It empowers security teams to make data-driven decisions and respond effectively to evolving threats in today's complex threat landscape.

Keywords: Data Mining, Intrusion Detection, Network Intrusion, EDADT, HIDS, Hoperaa, Predictive Data Mining, Network Data Systems, Denial of Service (DNS), Distributed Data Mining, Host-based Intrusion Detection Systems, Threat Intelligence Integration Systems (TIIS).

INTRODUCTION

In this modern world, intrusion occurs in a fraction of a seconds. Intruders cleverly use the modified version of the command and thereby erasing their footprints in audit and log files. Successful IDS intellectually differentiate both intrusive and nonintrusive records. IDS was first introduced by James Anderson in the year 1980 [1]. Most of the existing systems have security breaches that make them easily vulnerable and cannot be solved.. It has also become a priority and challenging task for network administrators and securityexperts. So it cannot be replaced by more secure systems. A network intrusion detection system (IDS) is a device or software application that monitors network or system activities for malicious activities or policy violations and produces reports to a management station. IDS come in a variety of “flavors” and approach the goal of detecting suspicious traffic in different ways [3]. There are network based (NIDS) and host- based (HIDS) intrusion detection systems are primarily focused on identifying possible incidents, logging information about them, and reporting attempts., documenting existing threats, and deterring individuals from violating security policies.

DOI: 10.17148/IJARCCE.2024.13926

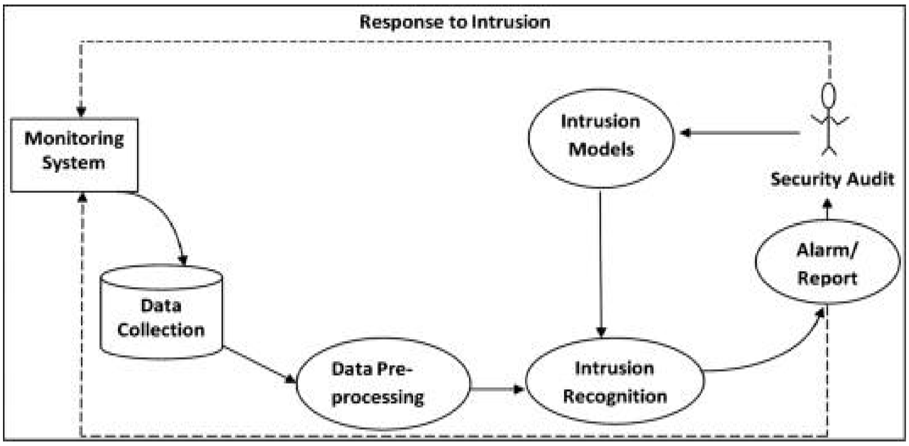

Figure 1: Processing of Intrusion Detection in a Network

Network behavior analysis is the ability to identify traffic patterns that are not considered normal in the day-to-day traffic of the network. Simply put, this is the industry's attempt to identify irregularities in the network beyond simple threshold settings for excessive traffic [6]. One of the most watched-for network security breaches is an abnormal traffic pattern known as a Distributed Denial of Service attack (DDoS). Itis a significant security threat to internet service providers and large network infrastructures.

Problem Statement & Related Work

Data mining approaches have been implemented by many authors to solve the detection problem. This implies that we are close to the solution. Since the pattern signature approach is currently utilized only by network administrators. The fact is that the existing works deal with the subset of problems that are needed for achieving intrusion detection and not others. To solve the problem of Classification of Data, an enhanced data-adapted decision tree algorithm [2] is proposed. This algorithm works differently normal decision tree algorithm. It efficiently classifies the data into normal and attack without any misclassification.

The problem of implementing supervised and unsupervised methods can be solved by using the Semi- Supervised Approach where with a small amount of labeled data, a large amount of unlabeled data can be labeled. Distributed Denial of Service Attacks can be greatly reduced using varying clock drift, with the help of varying clock drift in network-based applications [3], the adversary finds it difficult to access the port that has been used by the legitimate client. At the same time, any client can communicate with the server for longer time intervals without any interruption. This activity takes place on the client side to initiate a request to the server for connection and further communication.

To overcome this issue, the pseudorandom function and index value can be issued by the server to intimate the client to choose the next set of ports through this the bandwidth would be maintained. Once the server receives the contact initiation message it sends the varying clock value of the client and the server's clock value is $t_1 \dots t_n$ which denotes the arriving time of the same message at different rates. Then, equation (1)

Vhc(t1)={low rate(L),medium rate(M),high rate(H)}

It would be stored by the client to estimate the variable clock drift. The client waits for the acknowledgmentfrom the server for a specific period of time say $2\mu + L$. After that, it chooses another interval and starts sending messages. It may take any number of trials to get access to the server. In our proposed algorithm, the number of trials made by the client in the contact initiation part has been minimized so as to improve thereliability of the application [5]. Once the message is received, the server waits for the port to open. As soon as the port is open it sends the acknowledgment with pseudorandom function, index value, and varying time and t_1, t_2 to the client.

IDS Models

Classification of IDS essentially falls under two models: the *misuse or signature-based model* and the *anomaly model.*

DOI: 10.17148/IJARCCE.2024.13926

The misuse or signature-based is the most-used IDS model. Signatures are patterns that identify attacks by checking various options in the packet, like source address, destination address, source and destination ports, flags, payload, and other options. The collection of these signatures composes a knowledge base that is used by the IDS to compare all packet options that pass by and check if they match a known pattern. The anomaly model tries to identify new attacks [7] by analyzing strange behaviors in the network. To make this possible, it first has to ``learn" how the traffic in the network works and later try to identify different patterns to then send some kind of alert to the sensor or console. IDS made using this model have a higher tendency for raising false alarms, as they were often suspicious about all network behavior irrespective of whether malicious or legitimate.

- Data mining encompasses a wide range of techniques and algorithms for discovering patterns, relationships, and insights in large datasets. Some common data mining algorithms include:
- Association Rule Mining: Algorithms like Apriori and FP-growth are used to discover associations or patterns in data, often applied in market basket analysis.
- Clustering: Algorithms like K-means, hierarchical clustering, and DBSCAN group similar data points together based on certain criteria.
- Classification: Algorithms like Decision Trees, Random Forest, Support Vector Machines, and Naive Bayes are used for classifying data into predefined categories.
- Regression: Algorithms like Linear Regression and Logistic Regression are used to predict numerical or categorical values based on input features.
- Anomaly Detection: Techniques like Isolation Forest and One-Class SVM are used to identify outliers or anomalies in data.
- Dimensionality Reduction: Methods like Principal Component Analysis (PCA) and t-distributed Stochastic Neighbor Embedding (t-SNE) help reduce the dimensionality of data while preserving important information.
- Neural Networks: Deep learning techniques, such as Convolutional Neural Networks (CNNs) for image data and Recurrent Neural Networks (RNNs) for sequence data, are used for various data mining tasks.
- Text Mining: Techniques like Natural Language Processing (NLP) are used to extract insights from text data, including sentiment analysis, topic modeling, and named entity recognition.
- Time Series Analysis: Algorithms like ARIMA and Exponential Smoothing are used to analyze and forecast time-series data.
- Recommendation Systems: Collaborative filtering and content-based recommendation systems are used for personalized recommendations.

If "hoperra" is a more recent or specialized algorithm developed after my last knowledge update, I recommend consulting recent research papers, online resources, or academic literature to learn more about it. Data mining is a rapidly evolving field, and new algorithms and techniques are constantly being developed to address specific data analysis challenges.

IDSs can respond to a suspicious event in one of several ways, which includes displaying an alert, logging the event, or even paging an administrator. In some cases, the IDS may be prompted to reconfigure the network to reduce the effects of the suspicious intrusion. An IDS specifically looks for suspicious activity and events that might be the result of a virus, worm or hacker. This is done by looking for known intrusion signatures or attack signatures that characterize different worms or viruses and by tracking general variances that differ from regular system activity [6]. The IDS is able to provide notification of only known attacks

DOI: 10.17148/IJARCCE.2024.13926

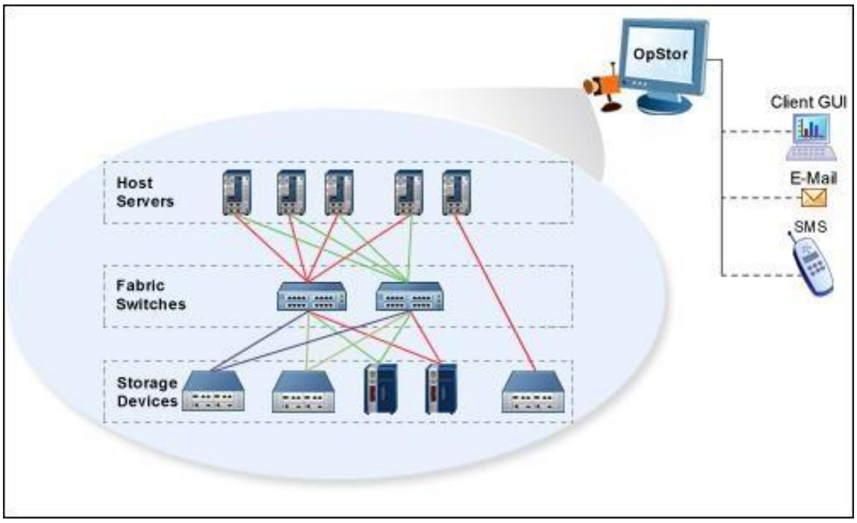

Figure 2: Host Functional Model

The term IDS actually covers a large variety of products, for which all produce the end result of detecting intrusions. An IDS solution can come in the form of cheaper shareware or freely distributed open programs, to a much more expensive and secure vendor software solution. Additionally, some IDSs consist of both software applications and hardware appliances and sensor devices which are installed at different points along your network [2].

Misuse Detection vs. Anomaly Detection

In misuse detection, the IDS analyzes the information it gathers and compares it to large databases of attack signatures [5]. Essentially, the IDS looks for a specific attack that has already been documented. Like a virus detection system, detection software is only as good as the database of intrusion signatures that it uses to compare packets against. In anomaly detection, the system administrator defines the baseline, or normal, state of the network's traffic load, breakdown, protocol, and typical packet size. The anomaly detector monitors network segments to compare their state to the normal baseline and look for anomalies.

Network-based vs. Host-based IDS

Intrusion detection systems are network or host based solutions. Network-based IDS systems (NIDS) are often standalone hardware appliances that include network intrusion detection capabilities. It will usually consist of hardware sensors located at various points along the network or software that is installed to system computers connected to your network, which analyzes data packets entering and leaving the network. Host-based IDS systems (HIDS) [4] do not offer true real-time detection, but if configured correctly are close to true real-time. Host-based IDS systems consist of software agents installed on individual computers within the system. HIDS analyze the traffic to and from the specific computer on which the intrusion detection software is installed on. HIDS systems often provide features you can't get with network-based IDS.

IPS — An Active Security Solution

IPS or intrusion prevention system, is definitely the next level of security technology with its capability to provide security at all system levels from the operating system kernel to network data packets. It provides policies and rules for network traffic along with IDS for alerting system or network administrators to suspicious traffic but allows the administrator to provide the action upon being alerted. Where IDS informs of a potential attack, an IPS makes attempts to stop it. Another huge leap over IDS is that IPS has the capability of being able to prevent known intrusion signatures, but also some unknown attacks due to its database of generic attack behaviors

Network-based vs. Host-based IPS

Host-based intrusion prevention systems are used to protect the servers and workstations through software that runs between your system's applications and OS kernel. The software is preconfigured to determine the protection rules based on intrusion and attack signatures. The HIPS will catch suspicious activity on the system and then, depending on the predefined rules, it will either block or allow the event to happen. HIPS monitors activities such as application or data requests, network connection attempts, and read or write attempts to name a few networks. Network-based intrusion prevention systems (often called inline prevention systems) are a solution for network-based security.
Problems associated with implementing NIPS exist as well. We already mentioned the possibility of blocking

legitimate traffic, and you also have to take network performance into consideration. Since all data moving through the network will pass through the IPS it could cause your network performance to drop. To combat this problem, network-based IPSs that consist of appliance or hardware and software packages are available today (at a larger cost), but it will take most of the load from running a software- based NIPS off your network.

Threat Intelligence Integration is the process of collecting, aggregating, analyzing, and applying threat intelligence data from various sources to enhance an organization's cyber security posture. Threat intelligence provides valuable insights into potential security threats, vulnerabilities, attack tactics, and malicious actors. By integrating threat intelligence effectively, organizations can proactively identify and respond to security risks and incidents. Implementing a Threat Intelligence Integration System is crucial for modern cyber security strategies. It enables organizations to proactively defend against emerging threats and vulnerabilities by leveraging timely and relevant information.

HOPERAA Algorithm

The Varying HOPERAA Algorithm has been illustrated using Fig. 6 in which L_c does not drift apart from the server. Since the varying clock drift is maintained depending on the length of the message. In this section, client *c* has a variable clock drift Vh_c related to the server. The server maintains threshold value to estimate the nature of the message. The server keeps the part of a HOPERAA algorithm [6] estimate the clock drift and also evaluates the state of message consequently. At the maximum the client runs Varying HOPERAA one time to estimate the clock drift but in case of fixed clock drift the client runs HOPERAA over three times to know the clock drift is slower or faster than the server. Through the proposed Varying HOPERAA Algorithm growth the growth of intervals is extremely reduced.

Hoperaa Algorithm

*****This pseudo code is called in Initiation Stage and should be plugin the subroutine Receive <ReMsg. 2. c. timestamp. V h < else (t1). t: > in Initiation Stage*

```
Varying HOPERAA Algorithm treply □the
arrival time of Re Msg
ρv □ (treply – Ts)/ (timestamp – TA )
*timestamp is included in the contact- else initiation reply message ReMsg*ρL □ (Treply –
Ts)/ (timestamp – TA +2A )
If 1 ≤ ρL ≤ ρv || ρL ≤ ρv ≤ 1 then
Interval varrying Hoperra □( ρv ρL Δ) / (ρv – ρL)If 1 ≤ ρL ≤
ρv then
If V hc(t) = = L then goto step 3else if V hc(t) == M
then goto step 2 else if V hc(t) == H then goto step 3
Step 1: Lc □ L L ρLElse
Lc □ L L ρv

End if
Step 2: Lc □ L M ρLElse
Lc □ L M ρvEnd if
Step 3: Lc □ L H ρLElse
Lc □ L H ρvEnd if

Else End if
Interval Varying HoperAA □ min { (ρL Δ/ 1- ρl), (ρU Δ/ ρL -1)}
```

Data mining based IDS can efficiently identify these data of user interest and also predicts the results that can be utilized in the future. Data mining or knowledge discovery in databases has gained a great deal of attention in IT industry as well as in the society. Much like choosing between standard securities devices like routers and firewalls, it is important to remember that no single security device will stop all attacks all the time. IPS and IDS work best when integrated with additional and existing security solutions. Network security has recently received an enormous attention due to the mounting security concerns in today's networks.

Network Behavioral Analysis

Behavioral monitoring for your network & systems is essential for spotting unknown threats [10]. It's also useful in investigating suspicious behavior and policy violations. When it comes to identifying threats in your environment, the best approach is a multi-layered one. Intrusion detection systems (network, host- based, and wireless IDS) identify known threats, and network behavior analysis can help you identify anomalies and other patterns that signal new, and unknown threats.

```
_p_ pattern to be matched m 0

Π[1] 0 k 2 to m for q do while k > 0 and p[k+1] :

p[q] Π[k] :

do k If p[k+1] = p[q] k +1 then kk Π[q]
                    return Π [21]:
```

There are two primary approaches to NIDS implementation: signature based, and anomaly detection based. The first approach has become a commercial success. A signature based NIDS maintains a collection of signatures, each of which characterizes the profile of a known security threat (e.g. a virus, or a DoS attack). These signatures are used to parse the data streams of various flows traversing through the network link; when a flow matches a signature, appropriate action is taken (e.g. block the flow or rate .limit it).

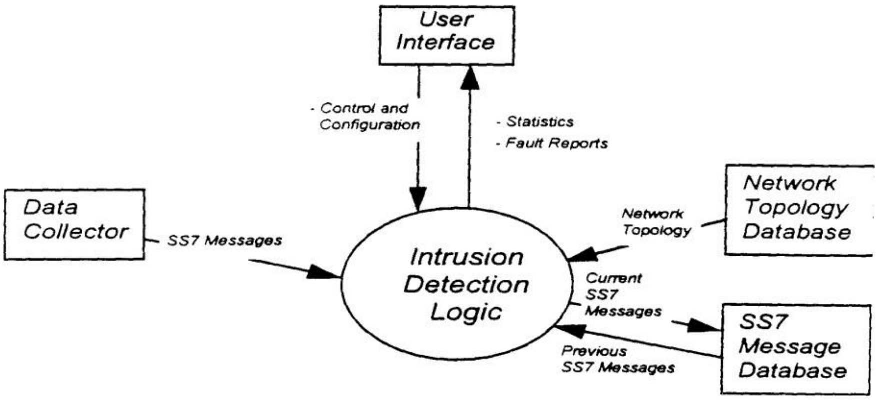

Figure 3. NIDS Traffic baseline block

String signatures are a string of ASCII symbols that characterizes a known attack. For example, such a string signature in UNIX can be "cat "+ +" > /.hosts" , which if executed, can cause the system to become extremely vulnerable to network attack. Simple strings may lead to high false positives, therefore it is important to refine the string signature; for this purpose one may use a compound string signature.

Implementation of Network Behavioral Model:

Anomaly based NIDS monitors network traffic and compares it against an established baseline of normal traffic profile. The baseline characterizes what is "normal" for the network - such as the normal bandwidth usage, the common protocols used, correct combinations of ports numbers and devices - and alerts the administrator or user anomalous traffic is detected which is significantly different from the baseline [9]. It is highly subjective to decide what can be considered normal and what an anomaly, but a widely accepted rule of thumb is that, any incident which occurs on a frequency greater than two standard deviations from the statistical norm should be considered suspicious.

An example of such behavior would be if a normal user logs on and off of a machine 20 times a day instead of the normal course of 1 or 2 times. Intrusion Detection Systems (IDS) play a crucial role in the manufacturing industry to protect critical infrastructure, sensitive data, and maintain operational continuity.

These systems help identify and respond to unauthorized access, cyber threats, and anomalies within the manufacturing environment.

Clearly, such anomaly based intrusion detection may lead to a high rate of false detection, which we call false positives. It is generally considered difficult to keep low false positives in any system that sets aggressive policies to detect anomalies. For example, it may be difficult to distinguish flash crowd from a Distributed Denial of Service attack (DDoS), thus a system may raise false alarm during a flash crowd event assuming that it is a DDoS attack [6]. Similarly network reconfigurations and transient failures may abruptly change the traffic profile falsely raising the alarm. The second challenge concerns with the assumption made by these systems that attacks are always anomalous, which may not necessarily be true.

Network Intrusion Detection		**Predicted label**	
		Normal	Intrusion
Actual Class	Normal	True Negative (42138)	False Positive (18455)
	Intrusion	False Negative (12528)	True Positive (237908)

Table 1 Standard metrics for system evaluation

Data Mining Techniques for Network Intrusion Detection Many researchers have investigated the deployment of data mining algorithms and techniques for intrusion detection Examples of these techniques include the feature selection data analysis:. The main idea in feature selection is to remove features with little or no predictive information from the original set of features of the audit data to form a subset of appropriate features [7]. Feature selection significantly reduces computational complexity resulting from using the full original feature set. Other benefits of feature selection are: improving the prediction of ID models, providing faster and cost-effective ID models and providing better understanding and virtualization of the generated intrusions.

Subset selection algorithms use heuristic search such as genetic algorithms, simulated annealing and greedy hill climbing to generate and evaluate a subset of features as a group for suitability. On the other hand, feature ranking uses a metric to rank the features based on their scores on that metric and removes all features that do not achieve an adequate score. The goal of classification is to assign objects (intrusions) toclasses based on the values of the object’s features. Classification algorithms can be used for both misuse and anomaly detection. [6]

Threat Intelligence Integration Systems (TIIS) or Threat Intelligence Platforms (TIPs) are softwaresolutions that help organizations collect, aggregate, correlate, analyze, and act upon threat intelligence data from various sources. These platforms are essential components of modern cyber security operations, as they enable organizations to proactively defend against cyber threats by leveraging up-to-date information about potential risks and vulnerabilities

Conclusion and Future Recommendations:

Network Intrusion Detection System has a major role to play in safeguarding the network resources against various kinds of attacks. With the advent of new vulnerabilities and sophistications in the nature of attacks, new techniques for intrusion detection have evolved. The main objectives of the research being increasing the detection accuracy while keeping the false positive rate low.. So the viable alternative would be to analyses the behavior of the network as a whole and trying to build the model based on the observations. The major concerns in this method are identifying the appropriate network features to characterize the network and build a behavioral model and also the rate of false positives may increase sharply if the IDS is not trained sufficiently in the target network

The IDS is designed to provide the basic detection techniques so as to secure the systems presentin the networks that are directly or indirectly connected to the Internet. Performing such a duty always goes in hand on hand diving success as well as failure in fulfilling the objective. This software does not completely shield network from Intruders, but IDS helps the Network Administrator to track down bad guys on the Internet whose very purpose is to bring your network to a breach point and make it vulnerable to attacks [7].

It can be employed and tested on various other machines which run on different Operating systemsand which satisfy the requirements and prerequisites for the IDS system. The present IDS system employs a log that is valid only for the current session and doesn’t store information about past sessions. The system may be enhanced by incorporating techniques corresponding to the future works listed below: The present system just displays the log information but doesn’t employ any techniques to analyze the information present in the log records and extract knowledge. The system can be extended again by incorporating Data Mining techniques using semantics and simulation techniques to analyze the information in the log records which may help in efficientdecision making.

DOI: 10.17148/IJARCCE.2024.13926

REFERENCES

[1]. Intrusion Detection: Challenges and myths by Marcus J. Ranum “Protect your network with an Intrusion Detection system”, Gartner Research http://www.techrepublic.com/article.jhtml.

[2]. F Esponda, S Forrest and P Helman (2004) A formal framework for positive and negative detection schemes’, IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, 34(1), pp 357-373. V Fuller and T Li and J Yu and K Varadhan (1993)

[3]. Classless Inter-Domain Routing (CIDR): an Address Assignment and Aggregation Strategy’, RFC 1519. Free Software Foundation Inc (2006) ’GNU’,

[4]. F Gomez and F Gonzalez and D Dasgupta (2003) ’An immuno-fuzzy approach to anomaly detection’, Proc. of the IEEE International Conference on Fuzzy Systems.J Hoagland and S Staniford (2003)

[5]. Ashok Kumar. (2014). Multidimensional Clustering Methods of Data Mining for Industrial Applications. *International Journal of Engineering and Science Innovation* , 7-8.

[6]. J, Tedesco G, Twycross J (2007): ’Immune System Approaches to Intrusion Detection - A Review ’, Natural Computing, Springer, forthcoming. G Lawton (2002)

[7]. Open Source Security: Opportunity or Oxymoron Institute of Electrical and Electronics Engineers Inc, http://www.computer.org/computer/ Lincoln Lab ’MIT data’ (2009)

[8]. T Mitchell (1997) ’Machine Learning’, McGraw Hill. P Ning and D Xu, ’Hypothesizing and Reasoning about Attacks Missed by Intrusion Detection Systems’, ACM Transactions on Information and System Security (TISSEC).

[9]. Petra Hunziker, Andreas Maier, Alex Nippe, Markus Tresch, Douglas Weers, and Peter Zemp, Data Mining at a major bank: Lessons from a large marketing application http://homepage.sunrise ch/homepage/pzemp/info/pkdd98.pdf.

[10]. Zubair Md. Fadlullah, Hiroki Nishiyama, “An Intrusion Detection System (IDS) for Combating Attacks Against Cognitive Radio Networks” 2018 IEEE

[11]. Ashok Kumar. (2014). Techniques to Enhance Productivity in Manufacturing Environments UsingData Mining. *International Journal of Engineering and Science Invention* , 7-8.

[12]. Mueen Uddin , Azizah Abdul Rehmanl etl “Signature-based Multi-Layer Distributed Intrusion Detection System using Mobile Agents” International Journal of Network Security, Vol.15, No.1, PP.79-87, Jan. 2019.