IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 19, Issue 1, Ver. I (Jan.-Feb. 2017), PP 65-72
www.iosrjournals.org

A Survey on Various Data Mining Technique in Intrusion Detection System

Snehil Dahima[1], Dr. Jitendra Shitlani[2]
[1](MCA, SIES College of Management Studies, Mumbai, India)
[2](CSE, SSUTMS Bhopal, India)

***Abstract**: The intrusion detection plays an essential role in computer security. Data Mining refers to the process of extracting hidden, previously unknown and useful information from large databases. Thus data mining techniques help to detect patterns in the data set and use these patterns to detect future intrusions. Data Mining based Intrusion Detection System is combined with Multi-Agent System to improve the performance of the IDS. This paper concerned with the brief review of comparative study on applied data mining based intrusion detection techniques with their merit and demerits. This paper relay more number of applications of the data mining and also focuses extent of the data mining which will useful in the further research.*
***Keywords**: Data Mining, Datamining based IDS architecture, Intrusion Detection, Multi Agent System, ,Multi-Agent based IDS.*

I. Introduction

The Knowledge Discovery in Databases (KDD) Process is used to represent the process of derive useful knowledge from larger data sets. Data mining is the process of discovering riveting patterns (or knowledge) from large amounts of data. The source of can include databases, data warehouses, the Web, any other information stores, or data that are flowed into the system (dynamically). Data can be related with classes or concepts that can be described in summarized, compact, and yet precise, terms, such descriptions of a concept or class are called class/concept descriptions. These descriptions can be obtained via Data Characterization and Data Discrimination. Data describes the real state of the world and the Knowledge describes the structure of the world and consists of directives principles, and rules. The KDD process involves a number of steps and is often interactive, iterative and user-driven decision making rules. Data mining is the most important step in the KDD process, and it applies data mining techniques to extract patterns from the data.

1.1 Know the application domain: To understand the back ground of the knowledge and to specify the goal.

1.2 Data Collection: Includes creating a target dataset which is relevant to the analysis.

1.2.1 ***Data Mining:*** Applying an appropriate algorithm to extract useful information using techniques.

1.2.2 **Data Interpretation**: to understand the discovered patterns and to confirm the goal is achieved.

1.2.3 **Knowledge Representation*:*** The final stage of representing the discovered knowledge. Data mining functionalities are used to specify the kind of patterns to be found in data mining tasks and it can be classified into two categories:

- ***Descriptive:*** To characterize the general properties of data in the database
- ***Predictive:*** to perform inference on data and to make predictions [1].

II. Goals of data mining

Widely Speaking, the purpose of Data Mining falls into the following groups: prediction, identification, classification and optimization.

2.1 Prediction:

Prediction discovers relationship between dependent and independent variables and the relationship between independent variables . Data mining showing how particular attributes within the data will behave in future. In some application, business logic is used coupled with data mining.

2.2 Identification:

Data patterns are using to identify the existence of an item, an event, or an activity or some new patterns of customer behavior. The area known as authentication is a layout of identification.

2.3 Classification:

Data Mining can separate the data so that different classes or categories can be recognized based on combination of parameters to find a clever say to display the data.

2.4 Optimization:

Data Mining can be optimized the use of limited resources such as time, space, money or materials and to maximize output variables under a given set of constraints.

III. Advantages of data mining

Data mining applications are continuously developing in various industries to provide more hidden knowledge that enable to increase business efficiency and grow businesses. Data mining approaches plays an important role in various domains. For the categorization of security problems, a large amount of data has to be study containing historical data. It is difficult for human beings to find a pattern in such an huge amount of data. Data mining, however, seems well-suited to overcome this problem and can therefore be used to discover those patterns [2].

IV. Intrusion detection system

An IDS is a combination of hardware and software which are used for detecting intrusion. It gathers and analyzes the network traffic & detects the malicious patterns and finally alert to the proper authority. The main function of IDS includes:

1.3 Monitoring and analyzing the information gathered from both user and system activities.
1.4 Analyzing configurations of system and evaluating the file integrity and system integrity.
1.5 For static records, it finds out the abnormal pattern.
1.6 To recognize abnormal pattern, it use static records and alert to system administrator.

According to techniques used for intrusion detection based on whether attack's patterns are known or unknown, IDS classified into two categories

1. Misuse detection
2. Anomaly detection

Misuse detection: It is Signature based IDS where detection of intrusion is based on the behaviors of known attacks like antivirus software. Antivirus software compares the data with known code of virus. In Misuse detection, pattern of known malicious activity is stored in the dataset and identify suspicious data by comparing new instances with the stored pattern of attacks.

Anomaly detection*:* It is different from Misuse detection. Here baseline of normal data in network data in network load on network traffic, protocol and packet size etc is defined by system administrator and according to this baseline, Anomaly detector monitors new instances. The new instances are compared with the baseline, if there is any deviation from baseline, data is notified as intrusion. For this reason, it is also called behavior based Intrusion detection system [3].

V. Types of ids

There are various types of IDS; they are characterized on the basis of different monitoring and analysis approach. Another way of classifying IDS is to group them by information source. There are some IDS which analyze information sources generated by the application software or Operating system for signs of intrusion. Other analyzes the network packet captured from network link to find attackers. Protected systems of IDS are Network based system and Host based system. Host based system monitors an individual host machine. Network based system monitors the traversing of packet on network link. People need to use the IDS in order to identify attacks in host based system and network based system.

5.1 Network Based System

Network Based IDS observe the packet that traverses through LAN segment and analyzes the network activity to identify attacks. Listening on a LAN segment, network based Intrusion detection system can observe the network traffic affecting multiple host that are connected to the network segment, so that it can protect those hosts. Network-based IDS often consist of hosts or a set of single-purpose sensors placed at several points in a LAN. Most of these Sensors are design to run in ―stealth mode, for the purpose of making it more difficult for an attacker/intruder to determine their presence and location. It is most commonly deployed at a boundary between networks, such as in virtual private network servers, wireless networks and remote access servers.

The following are the advantages of using network based IDS:

1) Network-based IDSs can be made invisible to number of attackers to give security against attack.
2) A few network based IDSs can monitor a large network.
3) Network-based IDSs are normal passive devices that listen on a network wire without interfering with the usual operation of a network. Thus, it is usually easy to fit in a current network to include network-based IDSs with least effort.

Disadvantages of using network based IDS are:

5 Network-based IDSs is not able to analyze encrypted information because various organizations use virtual private networks.
6 Most of the advantages of network based IDS don't apply to small segment of network i.e. switch based network. When it monitors range of switches, they are not universal, this limits the network based IDS monitoring range to single host.

7 Some network based IDS have also problem in dealing with network based attacks. The network based IDS involve the packet fragmentation. This anomalously formed packets cause the IDS to become unstable and crash.

5.2 ***Host based System***

A host-based IDS monitors activities which are associated with a particular host and focus at collecting information about activity on a host system or within an individual computer system. In host based IDS separate sensors would be required for an individual computer system. Sensor monitors the event which takes place on the system. Sensors collect the data from system logs, application activity, file access and modification, logs generated by operating system processes,. These log file can be simple text file or operation on a system.

The following are the advantages of using Host based IDS:

a) Host based IDS can identify attacks which cannot be seen by network based IDS because they monitor local events of a host.
b) Host based IDS works on operating system audit trails, that can help to detect attacks involve in software integrity breaches.
c) Host-based does not affect by switched networks.

Disadvantages of using Host based IDS are:

a) Host based IDS can be disabled by some DOS attacks.
b) Host based IDS are not good for detecting attacks, those who targets an entire network.
c) Host based IDS are difficult to manage, as for every individual system; information is configured and managed [4].

VI. Data mining based intrusion detection system architechture

The complete system architecture is designed to support a data mining-based IDS with the properties described. The architecture is made up of sensors, detectors, a data warehouse, and a model generation component. This architecture is having capabilities of supporting not only data gathering, sharing, and analysis, but also data archiving and model generation and distribution. The system is designed in such a way that is independent of the sensor data format and model representation. A piece of sensor data can contain an arbitrary number of features. Each feature can be continuous or discrete, numerical or symbolic.

6.1 Sensors

Sensors observe raw data on a monitored system and it computes features for use in model evaluation. Sensors insulate the rest of the IDS from the specific low level properties of the target system being monitored. This is ready by having the entire sensors implement a Basic Auditing Module (BAM) framework. In a BAM, features are computed from the raw data and encoded in XML.

6.2 Detectors

The processed data takes by Detector from sensors and use a detection model to evaluate the data and discover if it is an attack. The data warehouse get the result for further analysis and report which is send by the detectors. There can be several (or multiple layers of) detectors monitoring the same system. There can also be a "back-end" detector, which employs very sophisticated models for correlation or trend analysis, and various "front-end" detectors that perform simple and quick intrusion detection.

6.3 Data Warehouse

The data warehouse works as a centralized storage for data and models. One important advantage of a centralized repository for the data is that different components can alter the same piece of data asynchronously with the existence of a database, such as manually labeling and off-line training . The data warehouse also enables the integration of data from multiple sensors. By relating data/results from different IDSs or data collected over a longer period of time, the detection of complicated and large scale attacks becomes possible.

6.4 Model Generator

The main use of the model generator is to accelerate the rapid development and distribution of new (or updated) intrusion detection models. In this architecture, an attack detected initially as an anomaly can have its exemplary data processed by the model generator, which in turn, using the archived intrusion and normal data sets from the data warehouse, it automatically generates a model which can detect the new intrusion and distributes it to the detectors. Especially useful are unsupervised anomaly detection algorithms because they can operate on unlabeled data which can be directly collected by the sensors [5].

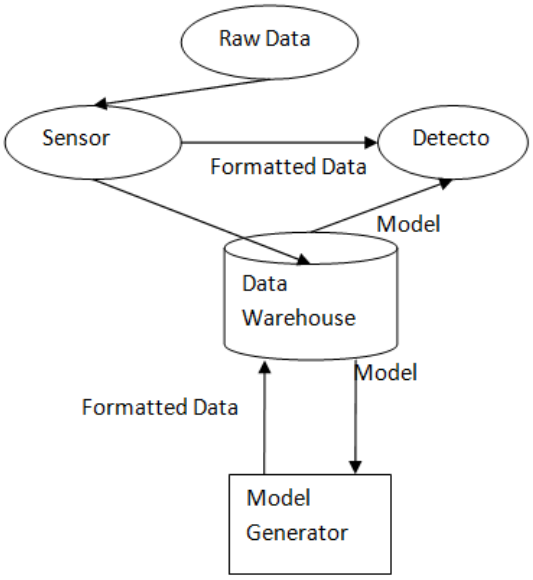

Fig.1. The architecture of data mining based IDS

VII. Need of data mining in intrusion detection

Data Mining generally refers to the process of previously unknown, extracting hidden and useful information from large databases. It is a convenient way of extracting patterns and focuses on issues relating to their utility, feasibility, scalability and efficiency. Thus data mining techniques help to detect patterns in the data set and use these patterns to detect future intrusions in similar data. The following are a some specific things that make the use of data mining important in an intrusion detection system:

- Manage firewall rules for anomaly detection.
- Analyze large volumes of network data.
- Same data mining tool can be applied to different data sources.
- Performs data summarization and visualization.
- Differentiates data that can be used for deviation analysis.
- Clusters the data into groups such that it possesses high intra-class similarity and low inter-class similarity [6].

VIII. Applied data mining based intrusion detection techniques

Data mining refers to identify for hidden patterns and trends in data warehouse which is not immediately apparent from summarizing the data, and there is no query involved but use the concept interestingness criteria i.e. specification of data such as Rarity ,Frequency, Length of occurrence ,Correlation, Repeating/ periodicity, abnormal behavior, Consistency and other patters of interestingness. The algorithms that are used for intrusion detection based on data mining techniques are listed as follows:

1.7 ***Association rule:*** Association rules mining identifies association among database attributes and its values. Association Rule is a pattern-discovery technique which does not serve to solve classification problems nor predict problems. Association rule mining requires two thresholds i.e. Minimum Confidence and Minimum support. Example: Apriori for mining Association rules Algorithm.

1.8 ***Classification*****:** Classification is the process of learning a function that maps data objects to a subset of a given class set. The two goals of classification are, first finding a good general mapping that can predict the class of so far unknown data objects with high accuracy. Second to find a understandable and compact class model for each other classes

8.3 ***Clustering techniques*****:** Clustering group's data elements into different groups based on the similarity between equivalence classes or within a single group Cluster partitions the data set into clusters. Cluster methods divided into two categories based on the cluster structure namely Hierarchical –connection oriented and Non Hierarchical.

8.4 ***Decision Tree*****:** Decision tree initially builds a tree with classification. Each node represents a binary predicate on one branch, one attribute represents the positive instances of the predicate and the other branch represents the negative instances. Construction of Decision Tree does not require any domain knowledge and can handle high dimensional data.

8.5 ***Genetic Algorithms Method***: learning examples are stored in relational database that are represented as relational tuples. It solves the problems with multiple solutions and easily transferred to models.

8.6 ***K Nearest Neighbor:*** An object classification process is achieved by the majority vote of its neighbors. The object is being assigned to the class most common between its k nearest neighbors. If k=1, then the object is simply assigned to the class of its nearby neighbor. Its Implementation task is simple and Easy for parallel implementations.

8.7 ***Support vector Machine Method:*** A support vector machine is a classification and regression technique it constructs a hyper plane or set of hyper planes in a high or infinite dimensional space. It is able to model complex and nonlinear decision boundaries.

8.9 ***Neural Network Method:*** A Neural Network is an adaptive system that changes its structure based on external or internal information which flows through the network during the learning phase. Implicitly detect the complex nonlinear relationships between dependent and independent variables. Highly tolerant the noisy data. Availability of multiple training algorithms.

8.9 ***Bayesian Method***: Bayesian classifier based on the rules. It uses the joint probabilities of sample classes and observations. The algorithms seek to estimate the conditional probabilities of classes given an observation. Naïve Bayesian Classifier simplifies the computations It exhibit high accuracy and speed when applied to large databases.

8.10 ***Fuzzy Logic*****:** The Fuzzy logic has been used for both anomaly and misuse intrusion detection. It uses linguistic variables and allows imprecise inputs, permits fuzzy thresholds. Rule base or Fuzzy sets easily modified[7].

IX. multiagent

9.1 Agents

a) What is an Agent?

The Agent, this word has many definitions in artificial intelligence circles some of the more common definitions are presented in this section. One of the most common definitions for agents is that the agent itself is entirely an umbrella term for a group of more specific types of agents. They can be classified by attributes. Some commonly used attributes are reactivity, autonomy, learning, cooperation, reasoning, communication, and mobility. The American Heritage Dictionary definition of an agent is a system that acts or has the ability to act or represent another. This essentially is using the term agent in the same way as it is used with travel agents and real estate agents. Another common definition is that an agent is anything that can autonomously communicate with its environment and an intelligent agent perceives its environment and makes informed decisions based on its perceptions and acts accordingly. For the purposes of this research we use the definition that an intelligent agent is defined, as a system which perceives its environment and acts upon the information it perceives.

b) Multi-Agent Systems

Unlike agent definitions, the definition of Multi Agent Systems (MAS) is accepted and well known as a loosely coupled network of agents that work together to find answers to problems that are beyond the individual knowledge and capabilities or of each agent and there is no global control system. The problems being beyond the individual capabilities of an agent could mean that the domain requires multiple different agent types each focused in a different area, meaning they can only solve part of the problem, or it could mean that each agent is only capable of solving a subset of the domains problems. There is a need for mechanisms for advertising, finding, fusing, using, presenting, managing, and updating agent services and information. Most MAS use a agent that is known as facilitator agents to help find agents, agents to which other agents surrender their autonomy in exchange for the facilitator's services. Facilitators can coordinate agents' activities and can satisfy requests on behalf of their subordinated agents. Other methods also exist including Mediators, Brokers, Matchmakers and yellow pages and Blackboards. Another method also exists that is called scenes, where the tasks are predetermined and every agent is told where the rooms are that they need to be. There are essentially two kinds of MAS. Closed MAS contain well-behaved agents which are designed to cooperate together easily toward a global goal. Open MAS can contain agents which are not designed to cooperate and coordinate. Most open MAS of coordination and cooperation mechanism are designed to assist the agents to working together. The most common kinds of these mechanisms are for auctions and negotiations. One example of negotiations is where an agent barters money, services, etc in exchange for assistance on a particular task or subtask by other agents. One example of an auction would be where you have a group of tasks and agents that you would like distributed as efficiently as possible. You would let the agents bid on the tasks they want to do. Assuming the agents are configured correctly they will only bid on tasks that they can complete for less than the other agents.

c) What are Agents used for?

There are some of the activities that agents have been used for include as Internet shopping assistants, game playing agents like for example soccer agents, non player (NPC) characters or even at a strategic level, personal assistants, text-learning and also for assisting decision support systems both outside and inside of the

medical domain, as detailed later in this paper. Agents are normally used in similar circumstances, in that they are used to monitor the current situation and knowledge base, and then make a decision on an action consistent with the domain they are in, and finally perform that action on the environment [8].

X. Multi agent based ids

The performance of IDS can be enhanced by using an agent. Agent Based IDS has following advantages which are following:

1.1 ***Decrease Network Flow:*** the process functions of central node to network nodes are distributed by systems and calculated by agents in network nodes. Malicious data package can be identified by system and send computing result to other nodes in network if there is abnormal information in data flow.

1.2 ***Improvement Autonomous Computing and Adaptation Capacity:*** Agent is autonomous independent unit. Other agents remain effective even though a few agents do not work for some reasons.

1.3 ***Platform Irrelevance***: agent based on IDS can work in diverse environment and implement interoperation on the application layer for agents are independent of the computer and transformation layer and work in nodes with agent.

1.4 ***Better Maintainability:*** Agent can response network topology dynamic changing as system can independent start and stop agent so IDS is configured dynamically [9].

Comparative study on applied data mining based intrusion detection techniques

YEAR	PAPER NAME	TECHNIQUE	MERITS	DEMERITS
Dec 2012	A Survey on Intrusion Detection using Data Mining Techniques	1. Association rule or Dependency Mining 2. Classification & clustering	Used in transaction data analysis Applied for KDD task Unsupervised technique	…..
Feb 2013	A Review Of Data Mining Based Intrusion Detection Techniques	1. Novel IDS 2. K-Means Clustering Algorithm 3. Data Dependency Weighted Sequence Mining 4. Hybrid IDS KDD Anomaly Detection	Used To Detect Dos Attack Detect Black Hole Attack Used To Filter Out Extra Rules Generated By This Approach Combines The Filter And Wrapper Models For Selecting Relevant Features Investigate More Efficient Methods Against Intrusions	Not Reliable Doesn't Provide Sufficient Mining Method Architecture Needs To Be Enhanced For Improvement Needs To Be Enhanced For Cryptographic Mechanism. Required To Survey More Recent Techniques
April 2013	Survey Paper On Data Mining Techniques Of Intrusion Detecion	1. Feature Selection 2. Machine Learning 3. Hybrid Approach	Used On Finite Data Set Improve Automatically Through Experience	
June 2013	A Survey On Intrusion Detection System In Data Mining	1. Data Mining, Feature Selection Multiboosting 2. K-Means Clustering Distributed IDS	Find High Detection Rates For U2R And R2L And Also To Detect Attacks False Alarm Rate Has Been Decreased Also Clustering Helps In To Identify The Attacked Data.	
Oct 2013	A Survey: Network Intrusion Detection System Based On Data Mining Techniques [7]	1. Support Vector Machine 2. Genetic Algorithm 3. K Nearest Neighbor 4. Neural Network 5. Bayesian Method	High Accuracy Solves Optimization Problem Simple And Highly Adaptive Behavior Implicitly Detect The Complex Nonlinear Relationships Between Dependent And Independent Variables Simplifies The Computations Exhibit High Accuracy And Speed When Applied To Large Database	More Time Space Complexity No Global Optimum High Storage Requirement It Requires Long Training Time The Assumptions Made In Class Conditional Independence Lack Of Available Probability Data

XI. Literature Survey

Saumya Raj et al. [2016] in this paper, multi agent based IDS presents the status of coordination issues, false alarm rates and detection rates on application of multiple agents.Finally, a hash table mechanism (Distributed Hash Table (DHT) & Internet Protocol (IP) based hash table) into the network to improve the matching efficiencies and computational speed. This survey conveys the difficulties in the traditional methods, namely, storage overhead, less matching efficiency, and adaptive nature (dynamically updating of hash tables) and false positive rates. The prediction of attackers or mis-behaving requests and the construction of adaptive reputation constitute the main problems in IDS that lead to less efficiency. The observation from the survey lead to the stone of extension of Distributed Hash Table (DHT) with fuzzy based rules in order to overcome the difficulties in traditional research works [10].

Yanjie Zhao et al. [2016] the paper's object is to develop a network intrusion detection model based on data mining technology, which can detect known intrusion effectively and has a good capacity to recognize unknown data schema which can't be detected effectively in traditional IDS. The paper mainly does the following work: by analyzing the intrusion deeply, extract the properties which can reflect intrusion characteristics effectively; combine misuse detection, anomaly detection and human intervention, establish rule library based on C.45 decision tree algorithm and use the optimal pattern matching so as to improve detection rate; the hosts are clustered to be IP group based on visit number by k means clustering algorithm, the audit data are divided into parts under the IP group's direction, and the classifiers are built up by divided audit data

respectively, then the detected Data apply different rules according to their own IP group, thereby reduce false positives [11]. Patrie Nader Paul et al. [2016] in this paper, our propose to use machine learning techniques, in particular one-class classification, in order to bring the necessary and complementary help to IDS in detecting cyber attacks and intrusions. One-class classification algorithms have been used in many data mining applications, where the available samples in the training dataset refer to a unique/single class. We propose a simple one-class classification approach based on a new novelty measure, namely the truncated Mahalanobis distance in the feature space. The tests are conducted on a real dataset from the primary water distribution system in France, and the proposed approach is compared with other well-known one-class approaches [12].
Hamid Reza Ghorbani et al. [2015] in this paper and by using data mining methods, an efficient policy driven detection strategy for intrusion detection has been proposed for the cloud environment. The proposed approach classifies different security needs, based on CIA triad model, into groups of users with the same security requirements and then selects the appropriate policy. By grouping similar users/security requirements and tuning each IDS accordingly, the proposed approach has been able to improve IDS efficiency. Results of our simulations show that the proposed approach decreases the total detection time by 21% in average while preserving adequate detection coverage. Improving IDS efficiency implies that it also processes a bigger volume of data due to reduction in time, better use of resources and also loads balancing between groups [13].

Hachmi Fatma et al [2015] in this paper, a two-stage process based on data mining and optimization is proposed having as input the outcome of multiple IDSs. In the first stage, for each IDS the set of elementary alerts is clustered to create a set of meta-alerts. Then, we remove false positives from the sets of meta-alerts using a binary optimization problem. In the second stage, our discard the meta-alerts generated by all IDSs and only those missed by one, two or most of them are left. This set is called the set of potential false negatives. In fact, at this level a meta alerts fusion is performed to avoid the redundancy between meta-alerts collected from multiple IDSs. Finally, a binary classification algorithm is proposed to classify the potential false negatives either as real attacks or not. Experimental results show that our proposed process outperforms concurrent methods by significantly reducing the rate of false positives and false negatives [14]. Fang-Yie Leu et al. [2015] in this paper, a security system, named the Internal Intrusion Detection and Protection System (IIDPS), is proposed to detect insider attacks at SC level by using data mining and forensic techniques. The IIDPS creates users' personal profiles to keep track of users' usage habits as their forensic features and determines whether a valid login user is the account holder or not by comparing his/her current computer usage behaviors with the patterns collected in the account holder's personal profile. The experimental results demonstrate that the IIDPS's user identification accuracy is 94.29%, whereas the response time is less than 0.45 s, implying that it can prevent a protected system from insider attacks effectively and efficiently [15].

Kailas Elekar et al. [2015] In this paper our have evaluated five rule base classification algorithms namely Decision Table, JRip, OneR, PART, and ZeroR. The comparison of these rule based classification algorithms is presented in this paper based upon their performance metrics using WEKA tools and KDD- CUP dataset to find out the best suitable algorithm available. The classification performance is evaluated using cross validation and test dataset. Considering overall higher correct and lower false attack detection PART classifier performs better than other classifiers.Many researchers working on number of data mining techniques for developing an intrusion detection system. For detecting the intrusion, the network traffic can be classified into normal and anomalous [16]. Ibéria Medeiros et al. [2015] this approach brings together two approaches that are apparently orthogonal: humans coding the knowledge about vulnerabilities (for taint analysis), joined with the seemingly orthogonal approach of automatically obtaining that knowledge (with machine learning, for data mining). Given this enhanced form of detection, we propose doing automatic code correction by inserting fixes in the source code. Our approach was implemented in the WAP tool, and an experimental evaluation was performed with a large set of PHP applications. Our tool found 388 vulnerabilities in 1.4 million lines of code. Its accuracy and precision were approximately 5% better than Php MinerII's and 45% better than Pixy's [17].
Irina Ioniță et al. [2013] in this paper, a multi agent based approach is used for network intrusion detection using data mining concept. In a network environment, intrusion detection is the act of detecting actions that attempt to compromise the confidentiality, integrity or availability of a resource. Due to increasing incidents of cyber-attacks, building intrusion detection systems (IDSs) remains a priority for protecting information systems security. Intrusion detection does not include prevention of intrusions. IDS should be fast enough to catch different types of intruders (external or internal intruders) before harm is done. Developing and implementing IDS is a complex task of knowledge engineering that requires an elaborate infrastructure. Modern technology such intelligent agents and data mining are appropriate to be used in network security [18].

XII. Conclusion

In this paper we briefly reviewed the several data mining applications which are used to detect the Intrusion in the network. This review would be helpful to researchers to focus on the various issues of data mining. In future course, a multi agent based approach is used for network intrusion detection using data mining

concept. The different techniques of data mining are used to extract the patterns and thus the knowledge from these different databases. Selection of data and methods for data mining is an important task in this process and needs the knowledge of the domain. Several attempts have been made to design and develop the generic data mining system but no system found completely generic. The intelligent agents up to some extent make the application generic but have limitations. Therefore it is conclude that multi agent system is used in combination with data mining technique to detect misuse and anomaly, combine IDS with network management system and develop cost sensitive Intrusion detection system.

References

This heading is not assigned a number.

[1] R.Venkatesan "A Survey on Wireless Intrusion Detection using Data Mining Techniques" International Journal of Innovative Research in Advanced Engineering (IJIRAE) Volume 1 Issue 1 (March2014).

[2] V. Jaiganesh , S. Mangayarkarasi , Dr. P. Sumathi "Intrusion Detection Systems: A Survey and Analysis of Classification Techniques" International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 4, April 2013.

[3] D. Shona, A.Shobana "A Survey on Intrusion Detection using Data Mining Technique" International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, Issue 12, December 2015.

[4] Sonam Chourse, Prof. Vineet Richhariya "survey paper on intrusion detection using data mining techniques" International Journal of Emerging Technology and Advanced Engineering, Volume 4, Issue 8, August 2014

[5] Sahilpreet Singh, Meenakshi Bansal " Survey on Intrusion Detection System in Data Mining" International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume No. 2, Issue No. 6, June 2013.

[6] Ms.Radhika S.Landge Mr.Avinash P.Wadhe "Review of Various Intrusion Detection Techniques based on Data mining approach" Vol. 3, Issue 3, May-Jun 2013, pp.430.435, International Journal of Engineering Research and Applications (IJERA).

[7] Anthony Raj.A "A Study on Data Mining Based Intrusion Detection System" International Journal of Innovative Research in Advanced Engineering (IJIRAE) Volume 1 Issue 1 (March 2014).

[8] Darren Foster , Carolyn McGregor , Samir El-Masri "A Survey of Agent-Based Intelligent Decision Support Systems to Support Clinical Management and Research"2011.

[9] Chandrakant Jain, Aumreesh Kumar Saxena "General Study of Mobile Agent Based Intrusion Detection System (IDS)" Journal of Computer and Communications, April 2016.

[10] Saumya Raj, Dr.Rajesh R "Descriptive Analysis of Hash Table Based Intrusion Detection Systems" 2016 IEEE.

[11] Yanjie Zhao "Network Intrusion Detection System Model Based on Data Mining" 2016 IEEE.

[12] Patrie Nader Paul Honeine Pierre Beauseroy "Detection of Cyberattacks In a Water Distribution System Using Machine Learning Techniques" 2016 IEEE.

[13] Hamid Reza Ghorbani, Roya Salek Shahrezaie "Toward a Policy-based Distributed Intruison Detection System in Cloud Computing Using Data Mining Approaches" 2015 IEEE.

[14] Hachmi Fatma, Mohamed Limam "A two-stage process based on data mining and optimization to identify false positives and false negatives generated by intrusion detection systems" 2015 IEEE.

[15] Fang-Yie Leu, Kun-Lin Tsai, Yi-Ting Hsiao, and Chao-Tung Yang "An Internal Intrusion Detection and Protection System by Using Data Mining and Forensic Techniques" 2015 IEEE.

[16] [13]Kailas Elekar M.M. Waghmare, Swami Chincholi, Daund, Pune, India monica.waghmare@gmail.com, Amrit Priyadarshi "Use of rule base data mining algorithm for Intrusion Detection" 2015 IEEE.

[17] Ibéria Medeiros, Nuno Neves, , and Miguel Correia "Detecting and Removing Web Application Vulnerabilities with Static Analysis and Data Mining" 2015 IEEE.

[18] Irina Ioniţă, Liviu Ioniţă "An Agent-Based Approach for Building an Intrusion Detection System" IEEE 2013.