A Review on Intrusion Detection System in Data Mining

Sweta Sharma[1], Anand Jawdekar[2]
[1, 2] Department of Computer Science Engineering
[1, 2] SRCEM, Banmore M.P, Gwalior, India

Abstract- *With the development of network technology, nowadays more and more people learn various ways of attack through the rich network resources, and carry out extremely destructive attack through simple operation. In recent years, the amount of hackers' attack is growing 10 times per year. Therefore, it has become the urgent topic to ensure the computer systems, network systems as well as the entire information infrastructure security, and it has become the general concern of the computer industry that how to detect and prevent these attacks effectively. Thus to enhance the network security dynamic approach is introduced and called as Intrusion Detection System. IDS collects online information from the network after that monitors and analyzes these information and partitions it into normal & malicious activities, provide the result to system administrator. A hybrid data mining method encompassing feature selection, clustering, filtering. A method for clustering the attribute type of the attacks applying soft set method and then applying clustering the attacks based on the selected attributes. The IDS is introduced for the efficient attacks identification to attain high detection and also accuracy rate as well as low false alarm rate.*

Keywords- Data mining, intrusion detection, Feature Selection, Classification.

I. INTRODUCTION

Data mining is the withdrawal of unseen predictive data or information from a big amount of database. It is strong and novel technology has great prospective to companies focus on the most significant information in their information repository. Data mining tools predict future drift and behaviors through permitting businesses to make knowledge-dive decisions [1].

Data mining mechanism can answer business or profession questions which were classically taking a big amount of time consuming to resolve. In the traditional data set, data does not change with time and they are nature is static, whereas streaming information generated continuously. Continuous data, i.e. streaming data is impossible to store, hence it need to be analyzed in single pass [2] [3] [4].Streaming data can be network data which consists of inbound and outbound traffic of the network.

With the development of network technology, nowadays more and more people learn various ways of attack through the rich network resources, and carry out extremely destructive attack through simple operation. In recent years, the amount of hackers' attack is growing 10 times per year. Therefore, it has become the urgent topic to ensure the computer systems, network systems as well as the entire information infrastructure security, and it has become the general concern of the computer industry that how to detect and prevent these attacks effectively.

There are many methods to strengthen the security of network at moment, for example encryption, VPN, firewall, etc., but all of these are too static to provide an efficient protection.

However, intrusion detection is a dynamic one, which can gives dynamic protection to the network security in monitoring, attack and counter-attack.

II. DATA MINING AND INTRUSION DETECTION

A. Data Mining Technology

Data Mining means extracting or mining the knowledge from the mass data. To be specific, it means processing data so as to gain the implied, prior unknown, potential and useful knowledge, which can be expressed as patterns. The targets of digging include not only data source and file system, but also any data collections such as Web source[5]. Data mining is the newest presented intrusion detection methodology. Its benefit lies in the fact that it can withdraw the needed and unknown information and regularities from the massive network information and host log data. It is a novel attempt to use data mining in achieving network security, both at home and abroad.

At the moment, the research on data mining algorithm is quite mature, and data mining itself is a general knowledge discovering technique. In the field of intrusion detection, we consider it as a data analysis process, which applies certain data mining algorithm to massive safe data in order to construct system of intrusion detection with self-adaptability and scalability. At current, an algorithm of data mining applied to intrusion detection mostly has four basic patterns: association, clustering, sequence and classification.

B. Intrusion Detection Technology

Security had become major concern in all fields of network & system infrastructure[6]. The basic challenge is to authorized user identify & the one who is legitimate to system access without abusing their privileges. Insider threats as well as outsider threats are rigorous to the system/network, known as intruders. Intrusion detection methodology can be describe as a method that classifies and deals with the malicious use of network and computer resources. It contain the exterior method behavior of intrusion and internal user's non-authorized. It is a methodology designed to ensure the computer system security that can discover and inform the non-authorized and abnormal occasions, used to detect the violation of network security.

An Intrusion Detection System (IDS) is critical technology to detect such intruders who are system harmful. Basic aim of the IDS is to protect the system & network from the intruders. IDS keep track of activities behavior; if they are system malicious then it'll be automatically detected through the IDS [7].

Thus IDS is further classified into three categories as follows[8]:

i) NIDS It is an platform independent that classifies intrusions through examining traffic network and monitors numerous hosts. NIDS increase access to network traffic through network hub connecting, port mirroring network switch configured, or network tap.

ii) HIDS It consists of an agent on a host that classifies intrusions through system calls analyzing, application logs, modifications of file-system (Access control lists, binaries, password files, capability databases, etc.) and other different host state and activities. In a HIDS, sensors commonly consist of a software agent.

iii) Hybrid IDS It complements HIDS system through the ability the monitoring network traffic for a particular host; it is various from the NIDS that monitors all network traffic . In computer security, a NIDS is an intrusion detection system that attempts to discover unauthorized access to a computer network through analyzing traffic on the network for signs of malicious activity.

In the case of detecting data target, intrusion detecting system can be classified as host-based, network-based, kernel-based and application-based. In this thesis, we focus on the construction of network -based system.

According to the differences of data analysis methods (that is, detection methods), we can say that there are two types of intrusion detecting system.

1) Misuse detection

Misuse Detection refers to the confirming attack events through matching features by attacking feature library [9] . It advances in the detection high speed and false alarm low percentage. However, it fails in discovering the non-pre-designated attacks in the feature library, so it cannot detect the many novel attacks.

2) Anomaly detection

Anomaly detection refers to features of storing consumer's normal actions into database, then comparing consumer's present behavior with those in the database. If the divergence is big enough, we can say that there is something abnormal. Its merits lie in its comparative irrelevance with the system, its strong versatility and the possibility detect the attack that has never been detected before. But because of the fact that normal contour conducted cannot give a full fill description of each users' behaviors in the system, moreover each user's behavior modifications constantly, its main disadvantage is the false alarm at high rate[10].

Combining these two, we may obtain a better performance. Anomaly detection can detect new, unknown attack or likewise, while misuse detection prevent the occasion that patient hacker gradually change pattern behavior so as to make the anomaly detection legalize the attack, which protects the integrity of anomaly detection. Intrusion Detection knowledge sources can be obtained with some dedicated capture tool. In Windows, data packs are gained with Wincap; in Unix, with Tcpdump and Arpwatch.

	Misuse Detection	**Anomaly Detection**
Characteristics	Use well-known attacks (signatures) patterns to recognize intrusions. Any match with the signatures is reported as a possible attack	Use deviation from normal usage patterns to recognize intrusions. Any significant deviations from expected behavior are reported as possible are reported as possible attacks

Drawbacks	- False negatives - Unable to detect new attacks - Need signatures update - Known attacks has to be hand-coded - Overwhelming security analysts	- False positives. - Selecting the right set of system features to be measured is ad hoc and based on experience - Has to study sequential interrelation between transactions - Overwhelming security analysts

III. NEED OF DATA MINING IN INTRUSION DETECTION

Data Mining refers to the removing hidden procedure, previously unknown and valuable knowledge from big amount databases [11]. It is a convenient extracting patterns way and focuses on issues relating to their feasibility, utility, scalability and efficiency. Thus, data mining methods help to detect patterns in the data set and also use these patterns to detect future intrusions in same data. The following are a some particular things that create the use of data mining significant in system of intrusion detection:

i) Manage rules of firewall for anomaly detection.
ii) Analyze network data big volumes.
iii) Similar data mining tool can be applied to various data sources.
iv) Achieves data visualization and summarization.
v) Differentiates information that can be used for analysis of deviation.
vi) Clusters the data into groups such that it possess high intra-class similarity and also low inter-class similarity.

IV. DATA MINING TECHNIQUES FOR INTRUSION DETECTION SYSTEMS

Data Mining refers to the process of extracting hidden, previously unknown and valuable information from large databases. It is a convenient way of extracting forms and focuses on problems relating to their feasibility, utility, efficiency and scalability. Thus data mining procedures help to detect forms in the data set and use these forms to detect future intrusions in related data. The following are a few specific things that create the usage of data mining important in an intrusion detection system [12]:

i) Arrange firewall rules for anomaly detection.
ii) Analyze large volumes of network data.
iii) Similar data mining tool can be applied to dissimilar data sources.
iv) Performs data summarization and visualization.
v) Dissimilar data that can be used for deviance analysis.
vi) Clusters the data into groups such that it possess great intra-class similarity and low inter-class similarity.

Data Mining Techniques for Intrusion Detection Systems Data mining techniques play an important role in IDS [13]. Different data mining methods like classification, clustering, association rule mining are used normally to acquire information about intrusions by observing and analyzing the network data. The following defines the various data mining methods:

A. Classification:

It is a managed learning technique. A classification based IDS will classify all the network traffic into either common or malicious. Classification technique is normally used for anomaly detection [14]. The classification process is as follows:

i) It receives collection of item as input.
ii) Maps the items into predefined groups or classes define by specific qualities.
iii) After mapping, it outputs a classifier that can correctly predict the class to which a fresh item belongs.

B. Association Rule:

This procedure searches a frequently occurring item set from a large dataset. Association rule mining determines association rules and/or correlation relationships amongst large set of data items [15]. The mining procedure of association rule can be distributed into two steps as follows:

i) Frequent Item set Generation Creates all set of items whose support is better than the identified threshold called as minsupport.
ii) Association Rule Generation From the earlier created frequent item sets, it creates the association rules in the form of ― at that time‖ statements that have confidence better than the identified threshold known as minconfidence.

The commonphase for incorporating association rule for IDS are as follows:

i) The network data is settled into a database table where all row signifies an audit record and all column is an audit records field.
ii) The intrusions and user activities present frequent connections among the network data. Consistent

behaviors in the network data can be captured in association rules.

iii) Rules based on network data can constantly merge the rules from a fresh run to aggregate rule set of totally previous runs.

iv) Thus with the association rule, we get the ability to capture behavior for properly detecting intrusions and hence lowering the false alarm rate.

C. Clustering:

It is an unverified machine learning mechanism for discovering forms in unlabeled data. It is used to label data and assign it into clusters where each cluster comprises of members that are quite similar [16]. Members from dissimilar clusters are different from all other. Hence clustering techniques can be valuable for classifying network data for detecting intrusions. Clustering can be applied on both Anomaly detection and Misuse detection. The basic steps involved in classifying intrusion are follows :

i) Find the biggest cluster, which involves of maximum number of orders and label it as common.

ii) Sort remaining clusters in an ascending order of their spaces to the biggest cluster.

iii) Select first K1 clusters so that various data instances in these clusters sum up to ¼`N and label them as common, where ` is the percentage of common instances.

iv) Label each other clusters as malicious.

v) After clustering, heuristics are used to automatically label all cluster as also common or malicious. The self- labeled clusters are then used to detect attacks in a single test dataset.

V. APPLICATION OF DATA MINING IN INTRUSION DETECTION

In traditional intrusion detecting system, security experts firstly classify attacking actions and system weakness, choose statistical methods due to the detecting types, then manually enter the code and establish the corresponding detecting rules and modes. For complex network system, the limitation of experts' knowledge grows with the change of time and space, so it is not good to improve the effectiveness of detecting the intrusion detecting modes. Security experts usually concern about the known attacking features and system weakness and research on that, which causes the lack of adaptability of the detecting pattern to the unknown intrusion that the system is about to be facing. Meanwhile, the long upgrade cycle of security system, the high cost, these are not advantageous for improving the adaptability of intrusion detecting pattern.

As the experts' rules and statistical methods often require the support of software and hardware, it stops the system from reusing and developing in new environment, meanwhile it causes the difficulty of embedding new detecting modules. All of these are not good for improving scalability of intrusion detecting pattern. Therefore, it has become an important issue how to establish an effective, self-adaptable and scalable intrusion detecting pattern in intrusion detecting field. Considering intrusion detection as a data analysis procedure through using data mining predominance in its effective use of knowledge, this is a technique that can automatically create accurate and applicable intrusion patterns from massive audit data, which createsintrusion detecting system can be applied to any computer environment. This approach has become a popular topic of research, in the field of inter discipline of network security and artificial intelligence. The analysis methods of association, sequence, classification and clustering in data mining has been proved possible [17].

Intrusions are the activities that violate the security norms of system. An IDS is Mechanism used to identify, monitor network or system actions for malicious activities and produces reports to a management departments. The development of IDS is motivated by the following factors: Most existing systems have security was that render them susceptible to intrusions, and finding and fixing all these deficiencies are not feasible. Prevention techniques cannot be sufficient. It is almost impossible to have an absolutely secure system. Even the most secure systems are vulnerable to insider attacks. New intrusions continually emerge and new techniques are needed to defend against them.

VI. DRAWBACKS OF IDS

- Current IDS does not detect the novel intruders: As some of the IDS work on the signature based technology, there are some predefined signatures in IDS, but as the signatures are predefined they fail to detect the novel intruders [18].
- False Positive: It occurs when normal is wrongly classified as intruder.
- False Negative: It occurs when an intruder is wrongly classified as normal.

VII. LITERATURE SURVEY

Sahilpreet Singh (2013) et al show This paper introduces a survey of procedures of IDS utilizing regulated and unsupervised learning. The strategies are sorted based upon various methodologies such as Statistics, Data mining, Neural Network Based and Self Organizing Maps Based methodologies. The location sort is obtained from interruption recognition as either abuse detection or peculiarity detection . It gives the peruser the significant headway in the malware research utilizing these methodologies the components and classes in the surveyed work based upon the above expressed classifications. This served as the significant commitment of this paper [19].

Janmejay Pant (2015) et al present Data vulnerability is a fundamental issue of data science to handle the data and information. Numerous hypotheses handle the instability issue. This paper broke down soft set diminishment and depicted how an information set is changed over into binary information system furthermore examined how it is ideal to decrease the measurement of the data. Like unpleasant set, Soft set and so on., are managing vulnerability. Delicate set hypothesis additionally do basic part to handle the uncertainty issue [20].

Yogita B. (2013) et al introduce that Security and protection of a system is bargained, when an interruption happens. IDS assumes imperative part in system security as it distinguishes different sorts of attacks in the network. So here, propose IDS utilizing data mining strategy: Here, Classification will be finished by utilizing SVM and check with respect to the viability of the proposed framework will be finished by directing a few tests utilizing NSL-KDD Cup'99 dataset which is enhanced rendition of KDD Cup'99 data set. The SVM is a standout amongst the most conspicuous order algorithms in the data mining zone, however its downside is its broad preparing time. In this proposed system, we have completed a few trials utilizing NSL-KDD Cup'99 information set. The trial results demonstrate that we can lessen broad time required to assemble SVM model by performing appropriate data set pre-preparing. Additionally when we do legitimate determination of SVM portion capacity, for example, Gaussian Radial Basis Function, attack detection rate of SVM is expanded and False Positive Rate (FPR) is decline [21].

Chetan R (2012) et al present Network security innovation has gotten to be essential in ensuring government and industry registering foundation. Current interruption discovery applications face complex prerequisites – they should be solid, extensible, simple to oversee, and have cost of low maintenance. In the present years, data mining – based IDSs have exhibited high precision, great speculation to novel sorts of interruption, and strong conduct in a changing environment. Still, noteworthy difficulties exist in outline and usage of creation quality IDSs. Instrumenting segments, for example, information changes, model arrangement, and agreeable appropriated discovery remain a work concentrated and complex building try. This paper portrays a database driven construction modeling that influences information mining with .NET to address these difficulties. It likewise offers various focal points as far as ready foundation, security,scalability, unwavering quality furthermore has information examination apparatuses. The database driven structural engineering is represented with a Data mining Based Intrusion recognition framework application model utilizing .NET [22].

Ming Xue (2009) et al present Intrusion Detection is one of system security zone of innovation principle research bearings. Data mining innovation will be connected to Network Intrusion Detection System (NIDS), might consequently find the new example from the huge network data, to lessen the workload of the manual gathering interruption conduct designs and normal behavior patterns.This article checked on the present intrusion detection innovation and the data mining innovation quickly. Concentrate on data mining algorithm in oddity detection and abuse detection of particular applications. For abuse identification, the principle ponder the characterization algorithm; For abnormality detection, the fundamental study the example correlation and the cluster algorithm In example examination to investigation profoundly the affiliation manages and arrangement rules . At last, has analysised the challenges which the ebb and flow data mining algorithm in applications ofIDS confronted at present , and has shown the following exploration heading [23].

Ketan Sanjay Desale (2014) et al present Recent rising development of data made such a large number of difficulties in data mining. data mining is the procedure of extricating legitimate, beforehand known and exhaustive datasets for the future choice making. As the enhanced innovation by World Wide Web the gushing data come into picture with its difficulties. The data which change with time and redesign its quality is known as gushing data .As the vast majority of the data is spilling in nature, there are such a variety of difficulties need to confront in the suspicion that all is well and good point of view. IDS works in the supposition of recognizing the gatecrashers to ensure the individual system. The exploration in data stream mining and IDS increased high fascination because of the significance of framework's security measure. Algorithms, frameworks and structures that address security challenges have been produced over the previous years. In this study, we show the system to enhance the effectiveness of the IDS utilizing gushing data mining procedure. We apply four chose stream data classification algorithms on NSL-KDD datasets and analyze their outcomes. In view of the relative examination of their outcomes best strategy is discovered for productivity change of IDS[24].

Tutut Herawan (2010) et al present reduct is a subset of properties that are mutually adequate and exclusively essential for protecting a specific property of a given data framework. The current reduct approaches under soft set hypothesis are still in view of Boolean-esteemed information system. Be that as it may, in the genuine applications, the information more often than not contain non-Boolean values. In this paper, an option approach for characteristic lessening in multi-esteemed data framework under delicate set hypothesis is displayed. In view of the thought of multi-delicate sets AND operation, trait diminishment can be characterized. It is demonstrated that the reducts acquired are proportional with Pawlak's harsh decrease [25].

Abhaya (2014)et al present With the significantly improvement of web, Security of system activity is turning into a noteworthy issue of computer network system. Attacks on the system are expanding step by step. The most plugged Attack on network activity is considered as Intrusion. IDS has been utilized for learning interruption and to safeguard the security objectives of data from assaults. data mining systems are utilized to screen and break down extensive measure of network data and group these system information into irregular and normal data. Since data originates from different sources, network traffic is substantial. Data mining strategies, for example, grouping and clustering are connected to construct IDS . A successful IDS requires high discovery rate, low false alert rate and in addition high precision. This paper introduces the survey on IDS and distinctive Data mining methods connected on IDS for the successful recognition of example for both malicious and normal activities in network, which creates secure information system [26].

BVST SAI (2013) et al display that Soft sets hypothesis was initially presented by Molodtsov in 1999. It is implied for managing dubious data in wording ofanalysis and choice making. Delicate sets are an exceptional sort of information systems. They are otherwise called Boolean esteemed information systems. This paper gives bits of knowledge into soft set , the use of the soft set theory, examination with rough sets and the synergetic points of interest of delicate sets utilization alongside other delicate figuring systems. It likewise concentrates on the use of soft sets continuously applications, the difficulties experienced in the process and the conceivable arrangements [27].

Ahmed Youssef (2011) et al present Intrusion location has turned into a basic segment of network administration because of the endless number of attacks steadily undermine our PCs. Conventional IDS are restricted and don't give a complete answer for the issue. They hunt down potential pernicious exercises on network traffics; they here and there succeed to discover genuine security assaults and oddities. Then again, by and large, they neglect to recognize malignant practices (false negative) or they fire alerts when nothing incorrectly in the system (false positive). What's more, they require thorough manual handling and human master impedance. Applying Data Mining (DM) procedures on system movement data is a promising arrangement that grows better interruption location frameworks. In addition, Network Behavior Analysis (NBA) is likewise a powerful approach for intrusion detection .In this paper, we talk about DM and NBA approaches for network intrusion detection and recommend that a mix of both methodologies can possibly distinguish interruptions in networks more effectively[28].

VIII. PROBLEM STATEMENT

Two basic premises of IDS are that activities of system are observable, e.g., via auditing, and there is distinct evidence that can be intrusive activities and distinguish normal. We call evidence extracted from raw audit data features, and use these features for evaluating and building IDS. Feature extraction (or construction) is the procedures of defining what evidence that can be taken from raw audit data is most beneficial for analysis. Feature extraction is thus a critical step in building an IDS. That is, containing features set whose values in normal audit records differ meaningfully from the values in intrusion records is vital for containing good detection presentation. But the problem is

- Present IDS are typically tuned to detect known servicelevel network attacks.
  This abandons them powerless against unique and novel malicious attacks.

- Data overload: Another perspective which does not relate straightforwardly to abuse identification but rather is critical is the amount of data an analyst can effectively analyze. . That measure of data he needs to take a gander at is by all accounts becoming quickly. Contingent upon the interruption discovery instruments utilized by an organization and its size there is the likelihood for logs to achieve millions of records every day.

- False positives: A typical dissension is the measure of false positives an IDS will create. A false positive happens when normal attack is erroneously delegated malicious and treated as needs be.

- False negatives: This is the situation where an IDS does not create a ready when an interruption is really occurring. (Classification of malicious traffic as normal).

IX. CONCLUSION

Now the people are more dependent on internet technology for their needs like online transactions, communications, emails etc. Due to this number of threats, even hackers and intruders, attack on the integrity and confidentiality of the system is increases. Thus the field of information security needs more security and safety. The high security can be achieved using authentication system, firewalls, encryption system, IDS etc. Data mining is a procedure of finding the unknown pattern from given set of patterns. In case of intrusion detection system, we use the data mining concept we will find out the pattern which will track all users activity to find out the intruders. In existing system we are focusing on knowledge engineering processes in which the decisions are taken on the basis of some fixed rule. Mainly IDS is divided in two broad categories i.e. IDS applying association rule mining and IDS using event correlation data mining. Our a hybrid data mining method encompassing feature selection, clustering, filtering. A method for clustering the attribute type of the attacks applying soft set method and then applying clustering the attacks based on the selected attributes. The IDS is introduced for the efficient attacks identification to attain high detection and also accuracy rate as well as low false alarm rate.

REFERENCES

[1] Trupti Phutane, Apashabi Pathan,” A Survey of Intrusion Detection System Using Different Data Mining Techniques”, International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol. 2, Issue 11, November 2014, pp: 6801-6807

[2] 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), pp: 21-25

[3] Harshna and NavneetKaur,” Survey paper on Data Mining techniques of Intrusion Detection”, International Journal of Science, Engineering and Technology Research (IJSETR) Volume 2, Issue 4, April 2013, pp: 799-802

[4] Changxin Song, Ke Ma, “Design of Intrusion Detection System Based on Data Mining Algorithm,” Proceedings of 2009 International Conference on Signal Processing Systems, IEEE 2009, pp. 307-373.

[5] Yogendra Kumar Jain and Upendra, “An Efficient Intrusion Detection based on Decision Tree Classifier Using Feature Reduction,” International Journal of Scientific and Research Publication, Volume 2, Issue 1, pp. 1-6, January 2012.

[6] Manikandan R, Oviya P and Hemalatha C, “A New Data Mining Based Network Intrusion Detection Model,” Journal of Computer Application, Volume 5, Issue EICA2012-1, pp. 1-10 February 10, 2012.

[7] Daejoon Joo, Taeho Hong and Ingoo Han, “The neural network models for IDS based on the asymmetric costs of false negative errors and false positive errors,” Expert System with Applications 25, 2003, pp.69-75.

[8] Wenke Lee and Salvatore J.Stolfo, “Data Mining Approaches for Intrusion Detection,” Proceedings of the 7th USENIX Security Symposium San Antonio, Texas, January 26-29, 1998.

[9] Harshna and NavneetKaur,” Survey paper on Data Mining techniques of Intrusion Detection”, International Journal of Science, Engineering and Technology Research (IJSETR) Volume 2, Issue 4, April 2013, pp: 799-802

[10] Zhao, Y., Luo, F., Wong, S.K.M. and Yao, Y.Y. “A general definition of an attribute reduct”. Proceeding of Second International Conference on Rough Sets and Knowledge Technology, RSKT 2007, LNAI 4481, 101–108.

[11] Molodtsov, D. “Soft set theory-first results”. Computers and Mathematics with Applications. 37, 1999, 19–31.

[12] Yao, Y.Y. “Relational interpretations of neighbourhood operators and rough set approximation operators, Information Sciences, 111, 1998, 239–259.

[13] Maji, P.K., Roy, A.R., and Biswas, R. “An application of soft sets in a decision making problem”. Compututer and Mathematics with Application, 44, 2002.

[14] Zou, Y. and Xiao, Z. “Data analysis approaches of soft sets under incomplete information”. Knowledge Based Systems, 21, 2008, 941–945.

[15] Saakshi Saraf,” Survey or Review on Soft Set Theory and Development”, The SIJ Transactions on Computer Science Engineering & its Applications (CSEA), Vol. 1, No. 3, July-August 2013, pp: 59- 66

[16] J.Daxin, C.Tang and A. hang (2004) Cluster Analysis for Gene Expression Data: A Survey, IEEE Transactions on

Knowledge and Data Engineering, Vol. 16, Issue 11, pp. 1370-1386.

[17] Sunita Jahirabadkar and Parag Kulkarni (2013) Clustering for High Dimensional Data: Density based Subspace Clustering Algorithms, International Journal of Computer Applications (0975 – 8887) Vol 63– No.20, pp. 29-35.

[18] Amandeep Kaur Mann and Navneet Kaur,” Survey Paper on Clustering Techniques”, International Journal of Science, Engineering and Technology Research (IJSETR) Volume 2, Issue 4, April 2013

[19] Sahilpreet Singh,Meenakshi Bansal,”A 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, pp: 2190- 2194

[20] Janmejay Pant, Amit Juyal and Shivani Bahuguna,” Soft set, a soft Computing Approach for Dimensionality Reduction”, IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 4, April 2015, pp: 728- 735

[21] Yogita B. Bhavsar1, Kalyani C.Waghmare,” Intrusion Detection System Using Data Mining Technique: Support Vector Machine”, International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 3, March 2013), pp: 581- 586

[22] Chetan R & Ashoka D.V.,” Data Mining Based Network Intrusion Detection System: ADatabase Centric Approach”, 2012 International Conference on Computer Communication and Informatics (ICCCI -2012), Jan. 10 – 12, 2012, Coimbatore, INDIA

[23] Ming Xue and Changjun Zhu,” Applied Research on Data Mining Algorithm in Network Intrusion Detection”, 2009 International Joint Conference on Artificial Intelligence

[24] Ketan Sanjay Desale, Chandrakant Namdev Kumathekar and Arjun Pramod Chavan,” Efficient Intrusion Detection System using Stream Data Mining Classification Technique”, 2015 International Conference on Computing Communication Control and Automation, pp: 469- 473

[25] Tutut Herawan, Rozaida Ghazali, Mustafa Mat Deris,” Soft Set Theoretic Approach for Dimensionality Reduction”, International Journal of Database Theory and Application Vol. 3, No. 2, June, 2010, pp: 47- 60

[26] Abhaya, Kaushal Kumar, Ranjeeta Jha, Sumaiya Afroz,” Data Mining Techniques for Intrusion Detection: A Review”, International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 6, June 2014, pp: 6938- 6942

[27] BVST SAI,” SOFT SET BASED TECHNIQUES FOR MINING UNCERTAIN DATA”, International Journal of Mathematics and Computer Applications Research (IJMCAR) ISSN 2249-6955Vol. 3, Issue 3, Aug 2013, 57-64

[28] Ahmed Youssef and Ahmed Emam,” NETWORK INTRUSION DETECTION USING DATA MINING AND NETWORK BEHAVIOUR ANALYSIS”, International Journal of Computer Science & Information Technology (IJCSIT) Vol 3, No 6, Dec 2011, pp: 87- 98