A Survey on Intrusion Detection System Based on K-Means and RBF Kernel Function

Vandana Shakya[1], Rajni Ranjan Singh Makwana[2]
[1, 2] Department Of Computer Science and Information Technology
[1, 2] Madhav Institute of Technology and Science, Gwalior (M.P.) India

Abstract- *In these days an increasing number of public and commercial services are used through the Internet, so that security of information becomes more important issue within the society records Intrusion Detection System (IDS) used against attacks for protected to the Computer networks. Feature choice strategies are explored. These are analyzed to see what effect they have on the accuracy of a simple SVM. Several filter and wrapper techniques are investigated. Hybrid methods which use combinations of filter and wrapper techniques are also investigated. Data mining method has been extensively carried out in the network intrusion detection device via extracting useful knowledge from massive wide variety of network data.*

Keywords- DM;IDS; k-means algorithm; RBFK; FS etc.

I. INTRODUCTION

Data mining is utilized to extricate verifiable and in the past obscure information from data. Data mining is the system which affords a concept to draw interest of customers because of high availability of large quantity of data and want to convert such facts into beneficial data. So, many human beings use the term “knowledge discovery device” or KDD for data mining. Knowledge extraction or discovery is completed in seven sequential steps used in data mining:

1) Data cleaning: we get rid of noise data and irrelevant data from collected raw data, at this step.
2) Data integration: At this step, we integrate multiple data sources into single data keep called target data.
3) Data Selection: Here, data applicable to evaluation mission are retrieved from data base as pre-processed data.
4) Data transformation: Here, data is consolidating into widespread codecs suitable for mining via summarizing and aggregated operations.
5) Data Mining: At this step, various smart strategies and tools are implemented as a way to extract data pattern or rules.
6) Pattern evaluation: At this progression, perceive see tree patterns speaking to comprehension.
7) Knowledge representation: This is the last stage wherein, visualization and information illustration techniques are used to assist customers to apprehend and secure and translate the data mining learning or result.

The aim of knowledge discovery and data mining procedure is to find the patterns that are hidden among the big set of data and interpret useful understanding and data.

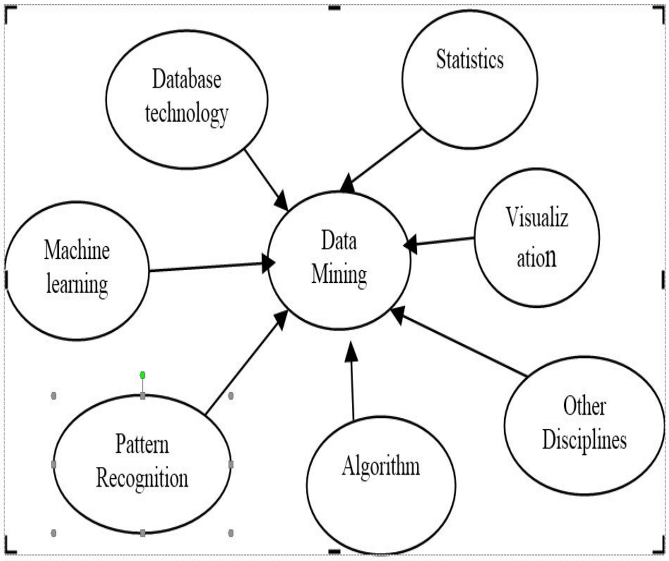

Figure 1. Data mining Concept and Techniques

In the diagram data mining is the principle part of expertise discovery system. Data mining programs:

- Marketing: Customer profiling, maintenance, character of capacity client, advertise segmentation.
- Fraud detection: Identify credit card fraud and intrusion detection.
- Scientific data analysis: Identify the examination decision making data.
- Text and web mining: used to search textual content or data on web or given raw data.
- Any other applications that contain large quantity of data [1]

II. INTRUSION DETECTION SYSTEM

The concept of IDS was proposed by Denning (1987), to become aware of, come detect and trace the intrusion. It gathers and analyzes the network traffic & detect the malicious patterns and finally alert to the proper authority. The main function of IDS includes: Monitoring and analyzing the facts gathered from each user and system sports. Analyzing configurations of system and comparing the record integrity and system integrity. For static records, it unearths out the peculiar sample. To capture irregular example, it utilize static realities and caution to system administrator.

A. Classification of IDS

As indicated by methodologies utilized for intrusion detection in view of regardless of whether attack's patterns are known or obscure, IDS arranged into two classes.
(1) Misuse location
(2) Anomaly location

1) Misuse location:

It is Signature based IDS where detection of intrusion depends on the practices of known attacks like antivirus software. Antivirus software compares the data with known code of virus.. In Misuse detection, pattern of acknowledged malicious pastime is saved within the dataset and perceive suspicious facts by means of evaluating new Times with the stored pattern of attacks.

2) Anomaly detection:

It isn't the same as Misuse detection. Here baseline of regular records in network data in network e.g. 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 occasions are contrasted and the pattern, if there is any deviation from baseline, data is advised as intrusion. For this motive, it's also called behavior primarily based IDS.

III. ADVANTAGES AND DISADVANTAGES OF ANOMALY DETECTION AND MISUSE DETECTION

The basic disadvantage of misuse detection methods is that they will recognize best the attacks for which they are educated to detect. Novel attacks or unknown attacks or maybe versions of common attacks regularly go undetected. The primary advantage of anomaly detection processes is the potential to come across novel attacks or unknown attacks against software program structures, editions of regarded attacks, and deviations of regular usage of packages no matter whether the supply is a privileged internal user or an unauthorized outside person. The drawback of the anomaly detection approach is that famous attacks might not be detected, in particular if they in shape the set up profile of the person. Once detected, it's far regularly hard to signify the character of the attack for forensic functions. Finally an excessive false positive rate might also end result for a narrowly trained detection algorithm, or conversely, a high false negative fee might also result for a broadly trained anomaly detection method [2].

IV. NEED OF DATA MINING IN INTRUSION DETECTION

Data Mining refers back to the method of extracting hidden, formerly unknown and useful statistics from huge databases. It is a handy manner of extracting patterns and makes a speciality of problems regarding their feasibility, utility, performance and scalability. Thus data mining techniques assist to come across patterns inside the data set and use these patterns to discover destiny intrusions in similar data. The following are some unique matters that make the use of facts mining important in an IDS:

i) Manage firewall rules for anomaly detection.
ii) Analyze large volumes of network data.
iii) Same data mining device may be carried out to specific data sources.
iv) Performs data summarization and visualization.
v) Differentiates data that can be used for deviation analysis.
vi) Clusters the data into groups such that it possess high intra-magnificence similarity and coffee inter-magnificence similarity [3].

V. TYPES OF IDS

There are several kinds of IDS; they're characterized on the idea of different monitoring and analysis technique. Another way of classifying IDS is to institution them by statistics source. Some IDS analyze statistics sources generated with the aid of the application software program or Operating device for signs and symptoms of intrusion. Other analyzes the community packet captured from network link to discover attackers. Protected systems of IDS are Network based totally system and Host based totally system. Host based system video display units an individual host machine. Network primarily based device monitors the traversing of packet on network hyperlink. People want to apply the IDS with the intention to pick out attacks in host based machine and network based machine. A. Network Based System Network Based IDS video display units the packet that traverses through LAN segment and analyzes the network activity to become aware of attacks.

Listening on a LAN segment, network based totally IDS can reveal the network traffic affecting more than one host which might be linked to the network section. So that It could defend the ones hosts. Network-based totally IDS often include hosts or a hard and fast of single -purpose sensors placed at numerous factors in a LAN. Most of those Sensors are layout to run in stealth mode, for the motive of making it more difficult for an attacker/intruder to decide their presence and vicinity. It is most typically deployed at a boundary among networks, consisting of in virtual private network servers, wireless networks and remote access servers.

The accompanying are the benefits of the use of network based IDS:

- Network-based IDSs can be made imperceptible to numerous attackers to give insurance against attack.
- A few network based IDSs can monitor a huge network.
- Network-based IDSs are normally passive devices that pay attention on a network cord without interfering with the normal operation of a community. Thus, it also includes easy to healthy in an existing network to consist of network-based IDSs with minimum attempt. Disadvantages of the usage of community based totally IDS are:
- Network-based totally IDSs is unable to analyze encrypted records because most of the business endeavor makes utilization of virtual non- public networks.
- Most of the advantages of network based absolutely IDS don't practice to little section of network i.e. Switch based network. Monitoring variety of switches are not universal, this boundaries the network based totally IDS tracking range to unmarried host.
- Some network based IDS have additionally hassle in managing network based attacks which involve the packet fragmentation. This anomalously shaped packets thought process the IDS to develop as precarious and crash.
- Host primarily based System A host-based totally IDS video display units sports associated with a specific host and geared toward amassing statistics approximately activity on a number system or inside an individual pc machine. In host based IDS separate sensors could be wanted for a person computer system Sensor monitors the event takes place on the system. Sensors collect the data from system logs, logs generated by operating system processes, application activity, file access and modification. The following are the advantages of using Host based IDS:
- Host based IDS can detect attacks which can't be seen by network based IDS due to the fact they screen neighborhood occasions of a number.
- Host based IDS operate on operating system audit trails, that can help to detect attacks involve in software integrity breaches.
- Host-based IDSs stays unaffected by utilizing switched networks. Detriments of the utilization of Host based IDS are:
- Host based IDS can be incapacitated by methods for beyond any doubt DOS attacks.
- Host based IDS aren't well alluring for detecting attacks, those objectives a whole network.
- Host based IDS are tough to manipulate, as for each person machine; information is configured and managed [4].

VI. METHODOLOGIES AND TECHNIQUES OF IDS

a) Methodologies of Intrusion Detection

There are three number one methodologies of Intrusion detection framework together with signature- based, anomaly based, and stateful protocol assessment.

- Signature Based Intrusion Detection Signature based totally detection is the best detection approach. A signature of packet is a sample or string that corresponds to a recognized chance. It is styles of well-known attacks. Signature-based detection is the system of comparing sample against located activities to discover viable incidents. It is effective for detecting recognized threats however it isn't always effective for detecting unknown threats. It is also incapable to song and recognizes the state of complex communications. It cannot come across the multiple occasions as it lacks the capacity of remembering previous request.

- Anomaly Based Detection

  Anomaly Based Detection is profile based absolutely detection which exhibits the standard conduct of client, hosts, network associations, or applications. The computer or network profiles can be produced by checking the attributes of commonplace movement over a timeframe. The current profile is matched with normal profile to identify significant deviations. Anomaly based detection is the way toward contrasting meanings of what action is viewed as typical against watched occasions to distinguish critical deviations. The Anomaly based totally intrusion detection system makes use of statistical methods to examine the traits of present activity to thresholds related to the profile. Profiles for anomaly-based totally detection can both be static or dynamic. It might be extremely viable at recognizing beforehand obscure dangers and produces numerous false positives.

- Stateful Protocol Analysis Based Intrusion Detection its protocol profile based. Anomaly based detection uses user, host, or network specific profiles. It is predicated on vendor advanced conventional profiles that explain how specific protocols must and have to not be used. Stateful protocol analysis is the system of evaluating predetermined profiles of typically standard meanings of favorable protocol enthusiasm for every protocol country against decided events to see deviations. It is capable to recognize and track the country of delivery, network, and application protocols which have a belief of nation. It can pick out surprising sequences of instructions and perform authentication [5].

VII. K-MEANS ALGORITHM

K-means is a centroid-based clustering with low time multifaceted nature and fast convergence, which is extremely basic in interruption identification because of the gigantic size of the network traffic review dataset. Each cluster in profile can be truly communicated as a centroid and an impact radius.

So a profile file may be represented as the following format (Centroid, radius, kind) Centroid is a centric vector of the cluster, radius refers to persuade range of a data point (represented as the Euclidean distance from the centroid), and type refers to the cluster's category, e.g. normal or attack. We can decide whether or not a vector is in the cluster or not simplest through computing the distance among the vector and the centroid and comparing the distance with the radius. If the space is much less than radius, we don't forget that the vector belongs to the cluster. And then we will label the vector because the cluster's kind. Therefore the whole search in the profile only includes several simple distance calculations, which means we can deal with the data rapidly. of course, no longer all clusters can serve as the profile. Some perhaps consist of each everyday and attack examples and now not healthy for the profile apparently. It is necessary to select some clusters in step with an approach. A majority example is an instance that belongs to the most common class within the cluster. The higher the purity is, the better the cluster is served as a profile. A cluster with small purity means that there are many attacks with different types in the cluster, so we don't select such cluster as our profile. Instead, we use them because the training set for classifier. After the clusters are decided on for the profile, we put them into the profile repository. The basic contents include centroid, radius and type. Here, we use the type of majority examples in one cluster as the whole cluster's type regardless of the minority examples [6].

VIII. RADIAL BASIS FUNCTION KERNEL FUNCTION

The Radial basis characteristic kernel, additionally referred to as the RBF kernel, or Gaussian kernel, is a kernel this is inside the shape of a radial basis function (more specifically, a Gaussian function). The RBF kernel is defined as

$$k_{RBF}(x,x') = \exp[-y\|x - x'\|^{2}]$$

Where γ is a parameter that sets the "spread" of the kernel. The RBF kernel as a projection into limitless dimensions. Recall a kernel is any function of the form:

$$k(x,x') =< \oint(x), \oint(x') >$$

where ψ is a function that projections vectors x into a new vector space. The kernel feature computes the inner -product among two projected vectors.

As we show under, the ψ function for an RBF kernel initiatives vectors into an infinite dimensional space. For Euclidean vectors, this area is an countless dimensional Euclidean space. That is, we show that
ψRBF : R ^{n} → R ∞

Without loss of generality, let γ = 1 2.

$$K_{RBF}(x,x) = exp\left[-\frac{1}{2}||x-x||^{2}\right]$$
$$exp\left[-\frac{1}{2}(x-x,x-x)\right]$$
$$exp\left[-\frac{1}{2}((x,x-x)-(x,x-x))\right]$$
$$exp\left[-\frac{1}{2}((x,x-x)-(x,x-x))\right]$$
$$exp\left[-\frac{1}{2}((x,x)-(x,x)-(x,x)+(x,x))\right]$$
$$exp\left[-\frac{1}{2}(||x||^{2}+||x||^{2}-2(x,x))\right]$$
$$exp\left[-\frac{1}{2}||x||^{2}-\frac{1}{2}||x||^{2}\right] exp\left[-\frac{1}{2}-2(x,x)\right]$$
$$= c_{e}^{(x,x)} \quad c = exp\left[-\frac{1}{2}||x||^{2}-\frac{1}{2}||x||^{2}\right] \textit{ is a constant}$$
$$= c\sum_{n=0}^{\infty}\frac{(x,x)^{n}}{n!} \quad \textit{Taylor expansion of } e^{x}$$
$$= c\sum_{n=0}^{\infty}\frac{k_{poly(n)}(x,x)}{n!}$$

We see that the RBF kernel is shaped through taking an infinite sum over polynomial kernels.

As proven previously, recall that the sum of two kernels

$$k_{c}(x,x') := k_{a}(x,x') + k_{d}(x,x')$$
$$k_{c}(x,x') := k_{a}(x,x') + k_{d}(x,x')$$

Implies that the ψc function is defined so that it forms vectors of the form

$$\psi_c(x) := (\psi a(x), \psi b(x))$$

That is, the vector ψc(x) is a tuple where the first element of the tuple is the vector ψa(x) and the second element is ψb(x). The inner-product on the vector space of ψc is defined as

$$< \psi c(x), \psi c(x') > := < \psi a(x), \psi a(x') \geq < \psi b(x), \psi b(x') >$$

For Euclidean vector spaces, this means that ψc(x) is the vector formed by appending the elements of ψb(x) onto the ψa(x) and that

$$< \psi c(x), \psi c(x') > := \sum_{i=1}^{dimension(a)} \psi a,i(x)\,\psi a,i(x') + \sum_{j=1}^{dimension(b)} \psi b,j(x)\,\psi b,j(x')$$
$$= \sum_{i=1}^{dimension(a)+dimension(b)} \psi c,i(x)\,\psi c,i(x')$$

Since the RBF is an infinite sum over such appendages of vectors, we see that the projections are into a vector space with infinite dimension [6].

IX. FEATURE SELECTION

Feature selection is the method that selects a subset of unique attributes and reduces the feature space

b) Relevant:

These are elements which have an impact on the output and their part can't be accepted by the rest. Feature selection is necessary either due to the fact it's miles computationally infeasible to apply all available capabilities, or due to troubles of estimation when limited data samples (however a massive range of features) are present [7].A “feature” or “attribute” or "variable" refers to a component of the data. Usually earlier than gathering facts, functions are targeted or selected. Features can be discrete, persistent, or nominal. For the most part, components are described as:

1) Relevant: These are elements which have a power at the output and their position can't be accepted through the rest.
2) Irrelevant: Irrelevant functions are described as the ones capabilities no longer having any have an impact on at the output, and whose values are generated at random for each instance.
3) Redundant: A redundancy exists on every occasion a characteristic can take the role of every other (possibly the most effective manner to model redundancy) [8].

X. GENERAL ALGORITHMS FOR FEATURES SELECTION

The basic feature selection algorithm is shown in the following.

```
Input:
S - data pattern with functions X,|X|  = n
J - evaluation degree to be maximized
GS – successor generation operator

Output:
Solution – (weighted) characteristic subset
L: = Start Point(X);
Solution: = {best of L according to J ;
repeat
L: = Search Strategy (L, GS (J), X);
X’:= {best of L according to J};
If J (X’) ≥J (Solution) or (J(X’)=J(Solution) and |X’| < |Solution|) then Solution :=X’;
Until Stop (J, L) [8]
```

XI. LITERATURE SURVEY

Hatim Mohamad Tahir et al. [2016] The purpose of this research is to improve on the existing Anomaly Based Intrusion Detection (ABID) method using K Means clustering technique as to maximize The problem with outliers may disturb the K-Means clustering process as it might be avoided in the clustering process from mixing with the normal data that make the NIDSs become less accurate. Thus this research aims to improve the performance of the ABID systems that balance the loss of information or ignored data in clustering. An integrated machine learning algorithm using K-Means Clustering with discretization technique and Na¨ıve Bayes Classifier (KMC-D+NBC) is proposed against ISCX 2012 Intrusion Detection Evaluation Dataset. The outcome depicts that the proposed method generates better detection rate and accuracy up to 99.3% and 99.5% respectively and reduces the false alarm to 1.2% with better efficiency of 0.03 seconds time taken to build model [9].

Dikshant Gupta et al. [2016] This research paper consists of the implementation of various data mining algorithms together with Linear regression and K-Means Clustering to mechanically generate the guidelines for classify network activities. A comparative analysis of those techniques to discover intrusions has also been made. To learn the patterns of the attacks, NSL-KDD dataset has been used. There are many risk of network attacks in the Internet environment.. Nowadays, Security on the net is a essential problem and consequently, the intrusion detection is one of the major research problem for commercial enterprise and personal networks which resist external attacks. A Network (IDS) is a software program utility that monitors the network or system sports for malicious activities and unauthorized get entry to devices. The goal of designing NIDS is to protect the data's confidentiality and integrity. Our project focuses on these issues with the help of Data Mining [10].

Qiuwei Yang et al. [2016] our propose an progressed particle swarm optimization algorithm ICPSO, which use chaos operator periodicity, randomness, sensitivity to initial situations and different qualities and the ICPSO is utilized to make the confusion into the dormancy weight consider parameters and The disarray is connected to the enhancement of the RBF kernel characteristic parameter g and the penalty aspect C, and to improve the convergence velocity and precision of the particle swarm optimization. The exploratory outcomes demonstrate that: with respect to the PSO-SVM calculation and GA-SVM algorithm, ICPSO-SVM enhances the productivity of intrusion detection, and is a successful intrusion detection version [11].

Angela Denise Landress et al. [2016] in this paper, Network intrusion detection means to reveal unapproved get admission to to PC structures. Anomaly intrusion detection utilizes unsupervised figuring out how to recognize attacks in light of profiles of ordinary user practices. If the gadget is being used otherwise, it triggers an alarm. Current methods of intrusion detection are not able to provide indicators without a high range of false positives. The proposed studies will utilize a set of artificial intelligence gadget studying strategies to lower the range of fake positives in anomalous intrusion detection facts. This technique combines facts clustering the usage of the simple K- means algorithm, feature selection that employs the J48 Decision Tree algorithm, and self organizing maps to effectively reduce false positives using the KDD CUP 99 data set [12].

Ait Tchakoucht Taha et al. [2015] This method help decrease U2R(exploring vulnerabilities to gain root access to the system) and R2L(obtaining access to remote system without having a user account) attacks that exploit operating system or software vulnerabilities and which are the most dangerous attacks towards confidentiality and integrity. Our experimental results will be applied to the hospital information system (IDS) [13].

Zhiguo Shi et al. [2015] In this paper, considering that the serious network protection state of affairs we're dealing with and the hassle of an increasing quantity of facts generated via the network, we proposed an IDS based on Hadoop, due to the lack of the conventional K-Means set of rules exists at we advise an advanced K-Means set of rules, we examine the performance of the K-Means algorithm and the progressed K-Means algorithm with KDD '99 facts units with the aid of the use of the IDS based on Hadoop. The experimental results display that the Accuracy Rate can attain 0.96, the Detection Rate can reach 0.89, the False Alarm rate the minimal is simplest zero.018. Three intrusion detection execution markers are superior to anything the conventional K-Means algorithm [14].

Xu Yang det al. [2015] in this paper, A data mining era is proposed in this newsletter in an effort to lessen the quantity of the fake alarms generated with the aid of intrusion detection systems and meanwhile improve the detection accuracy, in which such data mining era is an unsupervised clustering technique primarily based on hybrid ant colony algorithm and can be used to locate intruders' collective behaviors, without the want to know the earlier information. In the mean time, we embrace K- means clustering algorithm to quicken the meeting rate of the ACO Actually, the exploratory outcome demonstrates that the strategy proposed subsequently has higher discovery rate however bring down false alarm rate [15].

Theyazn hassn hadi et al. [2015] In this paper, our demonstrate the use of supervised partition membership preprocessing method to identify ambiguous packets. We propose an integrated version those effects in stepped forward type accuracy through explicitly clustering ambiguous packets to overcome its misclassification. The curiosity of our procedure lies being used of non- crisp clustering methods like fuzzy c- means (FCM) and rough k-means (RKM) that could adaptation vagueness. Further, we additionally examined whether or not FCM clustering and RKM clustering can assist to decide elegance of ambiguous packets exactly or approximately. The support vector machine (SVM) and J48 Classifiers consequences received on general facts sets are presented and as compared [16].

XII. CONCLUSION

Data mining strategies are able to extracting patterns automatically and adaptively from a massive amount of data. The safety of computer networks plays a planning role in present day computer system Detection of intrusion attacks is the most important issue in computer network security. Feature selection may be used to optimize the classifiers used to identify attacks through doing away with redundant or inappropriate features while enhancing the quality.

REFERENCES

[1] Aarti Sharma,Rahul Sharma,Vivek Kr. Sharma,Vishal Shrivatava “Application of Data Mining – A Survey Paper” (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 2023-2025.

[2] J.S. SHANTHINI, Dr. S. RAJALAKSHMI “Data mining techniques for efficient intrusion detection system: a survey” International Journal On Engineering Technology 2012.

[3] Ms.Radhika S.Landge Mr.Avinash P.Wadhe “Review of Various Intrusion Detection Techniques based on Data mining approach” International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com Vol. 3, Issue 3, May-Jun 2013, pp.430.435

[4] Sonam Chourse, Prof. Vineet Richhariya “Survey Paper on Intrusion Detection using Data Mining Techniques” International Journal of Emerging Technology and Advanced Engineering (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 8, August 2014).

[5] D.P.Gaikwad, R.C.Thool “A Framework for Simulation of Intrusion Detection System using Support Vector Machine” International Journal of Computer Applications (0975 – 8887) Volume 76– No.2, August 2013.

[6] Arvind Mewada, Shamaila Khan and Prafful Gedam “Network Anomaly Detection via Clustering and Custom Kernel in MSVM” (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 1, o. 1, 2010.

[7] Prof. Ujwala Ravale, Prof. Nilesh Marathe, Prof. Puja Padiya, “Feature Selection Based Hybrid Anomaly Intrusion Detection System Using K Means and RBF Kernel Function”, 1877-0509 © 2015.

[8] L.Ladha, “FEATURE SELECTION METHODS AND ALGORITHMS”, ISSN : 0975-3397 Vol. 3 No. 5 May 2011

[9] Hatim Mohamad Tahir, Abas Md Said ,Nor Hayani Osman, Nur Haryani Zakaria , Puteri Nurul ’Ain M Sabri5 and Norliza Katuk “OVING K-MEANS CLUSTERING USING DISCRETIZATION TECHNIQUE IN NETWORK INTRUSION DETECTION SYS” 2016 3rd International Conference On Computer And Information Sciences (ICCOINS).

[10] DikshantGupta, SuhaniSinghal, Shamita Malik, Archana Singh “Network Intrusion Detection System Using various data mining techniques” International Conference on Research Advances in Integrated Navigation Systems (RAINS - 2016), April 06-07, 2016, R. L. Jalappa Institute of Technology, Doddaballapur, Bangalore, India.

[11] Qiuwei Yang, Hongjuan Fu and Ting Zhu “An Optimization Method for Parameters of SVM in Network Intrusion Detection System” 2016 International Conference on Distributed Computing in Sensor Systems, 2325-2944/16 $31.00 © 2016 IEEE.

[12] Angela Denise Landress “A Hybrid Approach to Reducing the False Positive Rate in Unsupervised Machine Learning Intrusion Detection” 978-1-5090-2246-5/16/$31.00 ©2016 IEEE.

[13] AIT TCHAKOUCHT Taha, EZZIYY ANI Mostafa, JBILOU Mohammed, SALAUN Mikael “Behavioral appraoch for intrusion detection Application field: "E-Health" 978-1-5090-0478-2/15/$31.00 ©2015 IEEE

[14] Zhiguo Shi, Jianwei An “An Intrusion Detection System Based on Hadoop” 978-1-4673-7211-4/15 $31.00 © 2015 IEEE.

[15] Xu Yang, Zhao Hui “Intrusion Detection Alarm Filtering Technology Based on Ant Colony Clustering Algorithm” 2015 Sixth International Conference on Intelligent Systems Design and Engineering Applications.

[16] Theyazn hassn hadi and Manish R. Joshi “Handling Ambiguous Packets in Intrusion Detection” 2015 3rd International Conference on Signal Processing, Communication and Networking (ICSCN), 978-1-4673-6823-0/15/$31.00 ©2015 IEEE.