INTERNATIONAL JOURNAL · ENGINEERING RESEARCH ·

International Journal for Innovative Engineering and Management Research

A Peer Revieved Open Access International Journal

www.ijiemr.org

IJIEMR Transactions, online available on 20th Nov 2019. Link :http://www.ijiemr.org/downloads.php?vol=Volume-08&issue=ISSUE-11

Title **ANOMALY BASED INTRUSION DETECTION USING SIMPLE K-NEAREST NEIGHBOR CLASSIFIER ON KDD-CUP 99**

Volume 08, Issue 11, Pages: 75–86.

Paper Authors

NALIGALA VIJAYA RANI, NAGA MALLESWARA RAO PURIMETLA

Chintalapudi Engineering College, Ponnur, A.P, India-522124

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

ANOMALY BASED INTRUSION DETECTION USING SIMPLE K-NEAREST NEIGHBOR CLASSIFIER ON KDD-CUP 99

[1]NALIGALA VIJAYA RANI,[2]NAGA MALLESWARA RAO PURIMETLA

[1]((M.Tech), Dept of CSE,Chintalapudi Engineering College, Ponnur, A.P, India-522124

[2] Assoc. Professors, Dept of CSE, Chintalapudi Engineering College, Ponnur, A.P,India-522124.

Abstract

Nowadays there is an incredible escalation of the usage of computers over various networks and application domains, which in turn increases the security threats in terms of intrusions and in recent years the security in data mining applications has become crucial in protecting the public and private computing systems. Contemporary computing applications are facing numerous complex problems related to various internal or external attacks. An intrusion may happen either internally or externally and the traditional approaches used in intrusion detection are unable to meet the requirements of preventing and detecting an intrusion. For the detection of different attacks, intrusion detection occupied important work for the maintaining of privacy and reliability in network resource. In the modern world of security many researchers have proposed various new approaches among those techniques application of data mining for intrusion detection is one of the best suitable approaches for detection and prevention of intrusions. These data mining approaches will provide better results by using highly reliable and cost control mechanisms. The intrusion detection system (IDS) is an essential network protection device or software for guarding computing systems and it is proficient to identify and monitor network traffic data packets. Snort IDS is a free open-source network protection tool. Though, the Snort tool can detect only acknowledged attacks. In the proposed system, the methodologies of Data Mining has been used for increasing the performance of the IDS, and to handle Some of the problems like data Preparation, pre-processing of the data, data classification and Intrusion detection are being solved using different techniques like Dynamic Data Preparation (DDP), Hybrid Rule-based Pre-processing, and Simple K Nearest Neighbours Classification (SKNN) respectively.

Keywords: Intrusion, SKNN, NIDS, Preprocessing, Classification

1. Introduction

In the present computing world, usage of internet and application developed based on the internet has been increased rapidly and there is the proportionate growth of intrusions in the form of cyber attacks. Handling of new forms of intrusions is a severe task to the administration and it became the global issue. The main goal of intrusion detection is to monitor resources to detect abnormal behavior and misuses. In the year 1980, the concept was projected by James P. Anderson [3] by providing various ways to improve security [2, 6] auditing and surveillance at customer sites. During the period 1984 and 1986 Peter Neumann and Dorothy Denning developed the first real-time IDS, named as Intrusion Detection Expert System (IDES). Initially, IDES was trained to detect known malicious behavior using rule-based approach and further it was refined and named as Next-Generation IDS (NIDES) in the year 1988, university of California and U.S. Government-funded for the research projects like Haystack (US Air Force). Research work was done by comparing audit with known patterns, Host-based pattern matching system evolved and it was included in the Distributed atmosphere (i.e. Distributed IDS) In 1990, NIDS (Networks bases Intrusion Detection) was introduced by UC Davis's Todd Heber lien and contributed in DIDS and deployed NSM (Network Security Monitoring) and in early 90's Commercial IDS are developed like CMDS (Computer Misuse Detection System) host-based approach. In 1994, ASM (Automated Security Measurement system) came into the market.Despite the wide development of data innovation, security has stayed one testing territory for PC and systems. The quantities of hacking and interruption episodes are expanding year on year as innovation takes off. Security danger comes from outer gatecrashers as well as from inner clients as abuse. The firewall will be able to break the system and it can open the framework into the system and is unable to differentiate between good or bad activity. Consequently, if there is a requirement to permit an opening to a system, then a firewall which is a static rule-based, unable to protect from intrusion attempts. In contrast, Intrusion Detection Systems (IDS) can examine the hostile action on these openings. Conversely, Intrusion Detection Systems (IDS) can screen for threatening movement on these openings. The generic aspect of the IDS is represented in figure 1.1.In the present computing world, the necessary and important elemental of IDS is the network security architecture. Before characterizing the performance of IDS, it is important to know the behavior of an intrusion. The intrusion can be categorized in terms of integrity, confidentiality, and accessibility. An action or event causes a violation of confidentiality of the system. An action or event causes a violation of integrity if it permits shifting the circumstances of locating the resources, in a machine in an unlawful aspect. Likewise, the action or the event may cause a violation of the accessibility Sometimes the real users may be prohibited for the accessing of the services or its resources which are there on a computer. IDS have the options to track what actions are being performed in the system or on the web and observing it and to analyze the cipher of attacks. For monitoring or analyzing the attacks, IDS will act like a software or hardware which automatically processes its events. Due to the fast escalation of attacks, numerous intrusion detection systems anticipated in research. A few fundamental components are similar to the existing system and the rest vary from the proposed system. Figure 1.1 exhibits the generic design of IDS. Figure1.1 shows some of the detected like misuse and anomaly units, etc. Audit Trail Dataset collects the data to find events and processes the data to convert in the proper format. The Feature Extraction unit is the key aspect of

IDS. For detecting several intrusive behaviors, an alarm is set to detect.

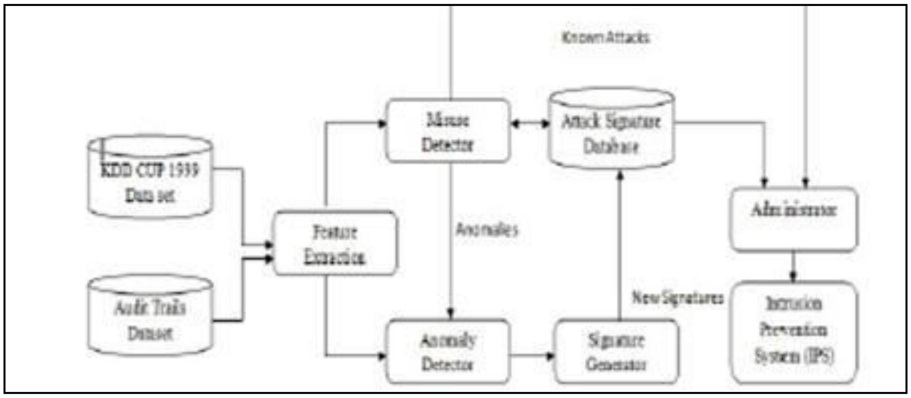

Figure 1: The generic view of traditional intrusion detection system

1.1 IDS-Detection Approaches

Intrusion detection is the process of monitoring the events occurring in a computer system or network and analyzing them for signs of possible attacks [1, 15, 28], which are violations or pending threats of violation of security [30] policies of a machine, satisfactory use of policies and standard security practices. Intrusion prevention is the process of performing intrusion detection and attempting to stop detected possible attacks. Intrusion detection and prevention systems are primarily focused on identifying possible attacks, logging data about attacks, trying to stop them and make recording them for security administrators. An intrusion detection system is disturbed mainly by the detection of aggressive actions. Technically intrusion detection methodologies are categorized mainly as two (as per the newly proposed schemes, the classifications can be more than two types) types. (i) Signature-based detection [3, 27] and (ii) Anomaly-based detection [2, 19]. In a broadway, the types of intrusion detections are classified based on their role and scope of working. As per technical view, the following are the different types of IDS,

Host Based Intrusion Detection (HIDS) System: A host-based intrusion detection system which monitors a computer system on which it is installed to detect an intrusion or misuse and responds by classification of the activity and notifies to the chosen authority. HIDS [15] can be measured as an agent, who monitors and analyzes whether attackshappened by anything or anyone. Generally, the attacks can be either an internal or an external attack, which has violated the security policies of the system.

Network-based intrusion detection (NIDS) system: Itis mostly used to observe and scrutinize network traffic to shield a system from threats related to the network. The NIDS [1, 16] examines all input packets and searches for any misbehaving or suspicious patterns. When attacks are revealed, based on its thoroughness [17], the machine can take necessary actions such as informing to administrators, or excluding the source IP address from the current network.

Finally, the data from the network is examined against a database and it flags those who look doubtful. Audit data from one or more hosts may be used as well to detect signs of intrusions [18, 25].

1.2 SIGNATURE-BASED DETECTION

A signature can be viewed as a pattern representing a well-known attack or hazard and it tries to compare patterns with captured events for detecting possible intrusions and accumulates with all specific attacks and system threats [4]. It is also known as knowledge-based detection. The IDS has an understanding of doubtful behavior and looks for activity which violates acknowledged policies [20]. It also looks for known malicious or unnecessary behavior. In fact, its major features are its effectiveness and comparably low false alarm rate [21].

1.3 ANOMALY-BASED DETECTION

An anomaly can be viewed as a variation or deviation to a known behavior. Anomaly profile represents expected or normal behavior obtained by monitoring regular activities [5, 6], connections in a network, and no. of hosts, network routers and users for a period of time. Generally, profiles can be categorized as either dynamic or static and it can be termed as behavior-based detection. For an example, Denial-of-Service attacks (DOS) [8, 12], SYN Flooding [9], usage of the processor [10, 13], failed log in attempts, Trojan horse, etc. Furthermore, hybrid detection methodologies have been evolved, such as (SPA) State full Protocol Analysis [14]. The role of Anomaly Detection Model (ADM) is the identification of data points, substance, event, and observations or attacks that do not conform to the expected (relevant) pattern of a given collection. These anomalies occur very infrequently but may indicate a huge and important threat such as cyber intrusions or fraud. Anomaly detection is a great deal used in behavioral analysis and other forms of analysis in order to assist in knowledge concerning the detection, recognition, and forecast of the occurrence of these anomalies or attacks. Anomaly detection can also term as outlier detection. The IDS has knowledge of normal behavior so it looks for anomalous behavior or deviations from the recognized baseline. While anomaly detection's most obvious drawback is its high false positive [19], it does offer detections of unidentified intrusions [26] and new exploits.

1.4 Data for Experimental study

The dataset selected for the fifth International Conference on KDD Process of Knowledge Discovery and Data Mining tools. The aim of the contest task was to frame an intrusion detector for network security, a foretelling IDS model proficient of differentiating among intrusion or attacks, called as bad connections, and normal connections called as good connections. This standard database consists of audited data, designed using a large range of attacks which have been simulated in the environment of the military network. The datasets are obtained from DARPA- 98 network data. Every connection in the network is described using 41 features, which provide information regarding BF-Basic Features, CF-Content Features, TTF-Time-based Traffic Features and HTF-Host-based Traffic Features. The attack classification is done by using class label considered as a 42nd feature, and it is used to distinguish the connection as normal or attack (the type of attack). About five million records are used for designing the training dataset and more than half million records are used for creating the testing dataset. Four categories of attacks are used for both testing and training datasets; they are Denial of Service, Remote-2-Local, User-2-Root, and Probe. The Majority of Pattern reorganization and classification techniques tested and trained on KDD IDS datasets are unable to identify major U2R and R2L attacks. These observations are taken to investigate further to identify the limitations and shortcomings of the KDD-99 dataset to dispute that these datasets should not be used in pattern reorganization or classification techniques used for detecting misuse activities of these two U2R and R2L attack categories.

It is similar hypothetical results for U2R and R2L. These techniques are analyzed by cross switching of the roles of both training and testing datasets, and relative and subjective analytical rules are generated separately on testing and training datasets through the decision tree approaches in data mining classification. The 1999 KDD Dataset is utilized to accept the adequacy of the Hybrid IDS. The originators of interruption discovery dataset mainly depend on the 1998 DARPA activity for to assess of frameworks in distinctive philosophies. The Military system consists of three machines with different frameworks and administrators. For the parody distinctive IP locations are used to produce activity. To record all the movement activities for the TCP dump position we use a sniffer. The reenacted period for the system is given as seven weeks. And now the attacks in the system are categorized into four types which are as below:

Denial of Service (DoS): The intruder tries to prohibit genuine customers from using network services.

Remote to Local (R2L): The intruder does not record in the machine, which results to get entrance.

Sender to Root (U2R): The intruders have a neighborhood for casualty machines and to increase master client benefits.

Probe: Intruder trying to get data from the objective host. The following parameters categorize the attributes of IDS. The 41 attack feature set is categorized based on the above-mentioned parameters. BF-Basic Features- 9, CF-Content Features- 13, TTF-Time-based Traffic Features-9, HTF-Host-based Traffic Features-10.

2. Related work

Intrusion Detection System was principally proposed by J. Anderson in the year 1980 [3]. W. R. Cheswick has ordered existing firewalls into three kinds dependent on the entryways they are application door, bundle sifting, and circuit separating and these sorts can be more than each one in turn [2]. Both SVM and C4.5 are analyzed by Ektefa the classifier execution does not suit for ongoing complex issues. The execution of C4.5 is better contrasted and different strategies [14, 15]. To enhance intrusion identification utilizing unlabeled information, Ching-Hao et al. proposed Co-preparing system. The proposed strategy demonstrated less mistake rate than existing techniques; the proposed technique has indicated upgraded precision [4]. Denning, D.E has proposed detecting and checking system on anomalous examples of review information to counteract security infringement. The Proposed strategy utilizes profiles for conduct portrayal regarding factual models and measurements [5]. To manage the multidimensional dataset, cross breed highlight determination is proposed by Sethuramalingam. S. The proposed strategy has evacuated a conflicting and repetitive component that diminishes the execution of characterization. For choosing huge highlights of the dataset hereditary system has joined with data gain. The proposed technique has demonstrated better precision when highlights are consolidated [24].

Berchtold et. al. [7] suggests pre-finding out, approximating and requesting the course of action space for the nearest neighbor issue in dimensional spaces. Pre-figuring the course of action space suggests choosing the Voronoi diagram of the data centers. The right Voronoi cells in d space are regularly incredibly staggering, in this way, the makers propose requesting estimation of the Voronoi cells. This procedure is fitting for first nearest neighbor issue in high dimensional spaces. John Mchugh has proposed a system of intrusion identification with the blend of the savage power strategy which is utilized to assess the intrusions and the proposed technique manages abuse location dependent on mark and inconsistency recognition [12]. Prof. Ujwala Ravale et al. proposed intrusion location component utilizing k-implies grouping and kernal elements of SVM utilized in the characterization display plan. The proposed framework has created a diminished number of ascribes identified with every datum point [18].

Gao Xiang, Wang Min has proposed unsupervised technique; it utilizes a huge dataset as preparing information and has recorded less exactness. To vanquish this issue, a semi-regulated methodology has been proposed [25]. The J48 algorithm is proposed by Panda, the proposed strategy arranges information into isolated classes like Attack or Normal. Both proposed strategies indicate more blunder inclined and Root Mean Squared Error [18].

3. SKNN Classifier

Classification is the process of finding a group of models, which differentiate and depict data, classes and their concepts. The goal of classification is to predict the class objects and assigning class labels for unknown class labels. The major challenge in classification is to build a data mining models with anomalous and unreliable datasets. In this proposed research a supervised Simple k-Nearest Neighbours algorithm has been used. The Simple k-Nearest Neighbours algorithm (SKNN) uses non-parametric technique for classification and regression. The major advantages of SKNN algorithm are less computation time, High predictive power, due to these aspects SKNN mostly used to solve classification problems. The input for SKNN algorithm is a set of k-closest training examples selected from feature space. SKNN can make predictions directly using the training dataset. The classification method is derived from standard KNN function, and the gamma value of the function decides the kernel activity to be carried on the training and testing datasets.

3.2. Notations of SKNN function

Simple K Nearest Neighbours function is utilized for basic knn characterization. Picking the quantity of nearest neighbors for example deciding the estimation of k assumes a noteworthy job in deciding the adequacy of the model. Subsequently, determination of k will decide how well the information can be used to sum up the aftereffects of the kNN calculation. Substantial k esteem has benefits which incorporate diminishing the fluctuation because of the boisterous information; the symptom being building up a predisposition because of which the student will in general disregard the littler examples which may have valuable bits of knowledge. The calculation is exceptionally impartial in nature and makes no earlier presumption of the hidden information. Being basic and powerful in nature, it is anything but difficult to execute and has increased great prominence. SKNN function has drawn a great deal of attention among the research community, due to the effectiveness and capable of classifying the large amount of data. The basic working principle of SKKN is that it forms the initial K nearest neighbors based on the best separable value among the attribute values. If numbers of neighbors are more, there will an increase in the processing time. The performance of the SKNN is measured by using gamma value.

4. Implementation Classification model

The Pseudo code for the proposed classification model is as follows.

```
Step 1: load the Misuse or Anomaly Dataset
Step2: Initialize the k-value
Step3: to obtain the predicted class,
    Perform iteration from 1 to total number of training data points
    Step3.1: Calculate the Euclidean distance between training data
             And each row of test data
    Step3.2: Based on the measured values, sort calculated distances
    Step3.3: Select top k, rows from the sorted dataset
    Step3.4: Identify most frequent data items
    Step3.5: Return the predicted class
```

The proposed of classification scheme is executed on the training dataset and process is depicted as follows.

```
Input: Training Dataset
Output: Classified data
Step 1: Select the training dataset.
Step 2: Proposed Kernel function is deployed
Step 3: The SKNN training is executed on training Data
Step 4: The trained dataset is loaded for testing.
Step 5: The testing data, structured fields are given for Classification of test data
Step 6: The SKNN classifier works based upon the proposed training structure
Step 7: The classification results are obtained.
Step 8: The classification result contains
```

the detected attacks for the protocols

The proposed SKNN is an improved version of the traditional Nearest Neighbours Classification approach which does classification using supervised learning approaches. Proposed technique maps linear vectors into non-linear space. Derived kernel function is used to construct hyper plane space by splitting features space. Semi supervised approach is used in the proposed SKNN technique in which prediction is done by setting target attribute values. The proposed technique is carried out in an iterative approach for generating decision function by using training dataset. The training dataset is combination both target and predictor values. If the proposed technique is able to predict an attack values for the chosen target value, then it is called the function of classification.

```
Algorithm for Anomaly Dataset
1. Perform read operation using read.csv - function
2. Read the table
3. aRow = nrow (function (x), dim(x) [1L] )
4. aCol = ncol (function (x) , dim(x) [2L])
5. Sub=Sampling of records
6. Generate anomalyTrainingSet
7. Generate anomalyTestSet
8. SKNN Classifier, specify the Gamma value, number of Nearest Neighbours and
    Anomaly Dataset
9. Anomaly Prediction
10. Generate Confusion Matrix
```

```
Algorithm for Misuse Dataset
1. Perform read operation using read.csv - function
2. Read the table
3. mRow = nrow ( function (x) ,dim(x) [1L])
4. mCol = ncol (function (x), dim(x) [2L])
5. Sub=Sampling of records
6. Generate misuseTrainingSet
7. Generate misuseTestSet
8. SKNN Classifier, specify the Gamma value, number of Nearest Neighbours and
    Misuse Dataset
9. Misuse Prediction
10. Generate Confusion Matrix
```

5. Results

The Proposed model has developed using SKNN Classification model and Statistical analysis tool, R programming language is used for analytical and classification activities. The KLAR library package is capable of adapting varied class labels used in the classification. The Results of Anomaly and Misuse attacks detection is presented in Figure 1.2. The existing system was developed from the concept of Hybrid PSO and C4.5. In this study, The IDS system is resided in the concepts of SKNN Classifier implemented in R. In this work “klaR” package available in R. In this research paper, data mining methodologies have been used for intrusion detection. The Proposed method will distinguish the features of Known features and unknown attacks. This work of intrusion detection is carried out using data mining tools with a sample of 6212 records of KDD Cup 1999 dataset to estimate and analyze the effectiveness among the existing traditional methods and our proposed methods. Each and every attack related features are measured and the count of observed results of each attack is depicted as in figure 1.3.

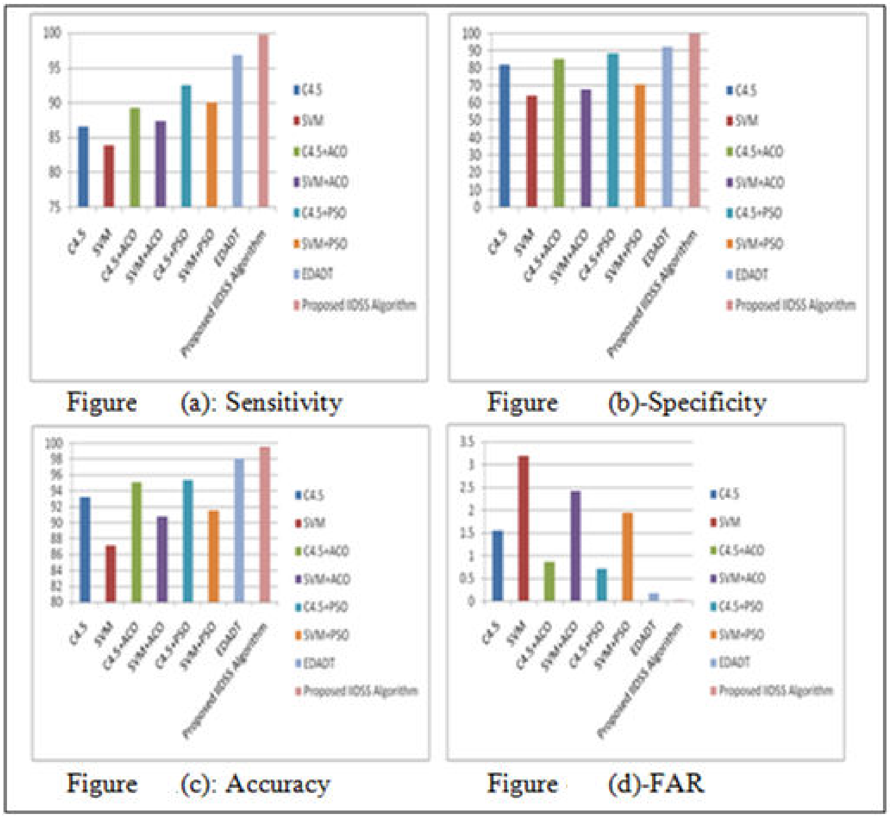

Figure (a): Sensitivity

Figure (b)-Specificity

Figure (c): Accuracy

Figure (d)-FAR

Figure 1.2: (a) Sensitivity, (b) Specificity, (c) Accuracy and (d) FAR

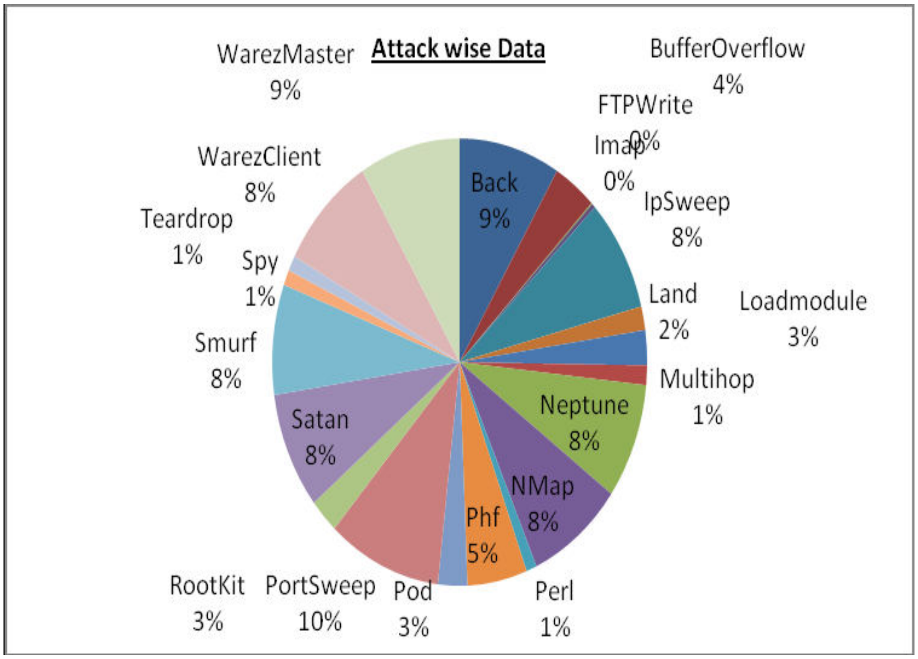

Figure 1.3-Attack wise count

6. Conclusion

Contemporary mechanisms for designing advanced intrusion detection system have been projected in this thesis; supervised learning schemes are presented and evaluated using proposed schemes. The efficiency and precision of the algorithms are confirmed by testing on the standard KDD cup 99 datasets. Initially, to carry out the experimental activities 10 percent of KDD cup 99 corrected dataset is taken. The Dynamic data preparation approach is used, in preprocessing of the data; to remove the redundant records the record count is finally minimized. The size of the datasets used in the preparation phase is extremely large, and the time taken for the data preparation is minimized. During the preprocessing phase, low variant or high variant attributes are identified by using predefined threshold value, which results high coherent datasets are generated.The KDD cup records are categorized into five groups, namely Normal, DoS, U2R, R2L and Probe. Each data record is associated with a total of 42 attributes, out of which 41 are the functional attributes and the last column of the record is the class label. Sampling mechanism is adapted to develop training and testing models. To handle missing, noisy and inconsistent values, preprocessing approach is used. This phase of preprocessing reduced the feature set count from 42 to 39. For classifying the unlabelled record, IIDSS methodology is used. Using random sampling approach a sample record count of 6212 is taken for the first iteration and the classification process is repeated and observed that accuracy has improved.

The IIDSS approach has recorded better accuracy values over the existing methodologies. The false alarm rate has been minimized; this reduced FAR will directly affect the reduction of administrator workload. The IIDSS method produced 14.24% sensitivity is increased over C4.5 when compared with the SVM approach the sensitivity is raised by 14.96%. On the other hand the proposed system has recorded an amount of 11.45%, inclination in sensitivity over C4.5+ACO, and over EDADT approach its incremented values is 2.69%. The classifier model has shown high accuracy values and low false alarm rate is recorded. The size of training and testing datasets is also higher compared with existing methodologies. Time taken for the building of classifier is minimized.

References

1. A.Saidi et al.:, The functional of A Mobile Agent System to Enhance DoS and DDoS Detection in Cloud, International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 6 (2016) pp 4615-4617
2. Adeeb Alhomoud et al., Performance Evaluation Study of Intrusion Detection Systems, The 2nd International Conference on Ambient Systems, Networks and Technologies, (ANT), Procedia Computer Science 5 (2011) 173–180, 1877–0509 © 2011 Published by Elsevier Ltd. Selection and/or peer-review under responsibility of Prof. Elhadi Shakshuki and Prof. Muhammad Younas. doi:10.1016/j.procs.2011.07.024

3. Anderson, James P., "Computer Security Threat Monitoring and Surveillance," Washing, PA, James P. Anderson Co., 1980.
4. Anna L. Buczak. (2015). A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection, 10.1109/COMST.2015.2494502, IEEE Communications Surveys & Tutorials, 1553-877X (c)
5. Aymen Abid et al.:, Outlier detection for wireless sensor networks using density-based clustering approach, IET Wireless. Sens. Syst., 2017, Vol. 7 Iss. 4, pp. 83-90, The Institution of Engineering and Technology 2017, ISSN 2043-6386
6. Bellovin, S.M. “Network Firewalls”, IEEE Communications Magazine, Vol. 32, pp. 50- 57, 1994.
7. Berchtold, B. Ertl, D. A. Keim, H.-P. Kriegel, and T. Seidl. Fast nearest neighbor search in highdimensional space. In Proceedings of the Fourteenth International Conference on Data Engineering, ICDE ’98, pages 209–218, Washington, DC, USA, 1998. IEEE Computer Society.
8. Blum, Avrim L. & Pat Langley (1997). Selection of relevant features and examples in machine learning. Artificial Intelligence, 97(1-2), 245–271
9. Catania Carlos A, Garino Carlos. Automatic network intrusion detection: current techniques and open issues. Elsevier Comput Electr Eng 2012; 38(5):1062–72.
10. Chien-Yi Chiu, Yuh-Jye Lee, Chien-Chung Chang. Semi-supervised learning for false alarm reduction. In: Industrial conference on data mining, no. 10; 2010. p. 595–605.
11. Ching-Hao, Hahn-Ming L, Devi P, Tsuhan C, Si-Yu H. Semi-supervised co-training and active learning based approach for multi-view intrusion detection. In: ACM symposium on applied computing, no. 9; 2009. p. 2042– 7.
12. Claude Turner et al. (2016). A Rule Status Monitoring Algorithm for Rule-Based Intrusion Detection and Prevention Systems, Complex Adaptive Systems, Conference Organized by Missouri University of Science and Technology 2016 - Los Angeles, CA, Procedia Computer Science 95 (2016) 361 – 368, 1877-0509, doi: 10.1016/j.procs.2016.09.346
13. Das, S. (2001). Filters, Wrappers and a Boosting-Based Hybrid for Feature Selection. Proc. 18thInt’l Conf. Machine Learning, 74-81
14. Dasgupta, D. and F. A. Gonzalez, ”An intelligent decision support system for intrusion detection and response”, In Proc. Of International Workshop on Mathematical Methods, Models and Architectures for Computer Networks Security (MMM-ACNS), St.Petersburg. Springer- , 21-23 May,2001.
15. Denning, D.E. “An Intrusion-Detection Model”, in IEEE Transactions on Software

Engineering, Vol.13, No. 2, pp. 222-232, 1987.

16. Dickerson, J. E. and J. A. Dickerson, ”Fuzzy network profiling for intrusion detection”, In Proc. of NAFIPS 19th International Conference of the North American Fuzzy Information Processing Society, Atlanta, pp. 301306. North American Fuzzy Information
17. Divya and Surendra Lakra, “SNORT: A Hybrid intrusion detection system using artificial intelligence with a snort”, International journal computer technology & application, Vol 4(3), 466-470, 2013.
18. E.Kesavulu Reddy, Member IAENG, V.Naveen Reddy, P.Govinda Rajulu. A Study of Intrusion Detection in Data Mining. Proceedings of the World Congress on Engineering 2011 Vol III WCE 2011, July 6 - 8, 2011, London, U.K.
19. Eskin, E., Arnold, A., Prerau, M., Portnoy, L., and Stolfo, S. J., A Geometric Framework for Unsupervised Anomaly Detection: Detecting Intrusions in Unlabeled Data, In D.Barbarà and S. Jajodia (eds.), Applications of Data Mining in Computer Security, Kluwer Academic Publishers, Boston, MA, 2002, pp. 78-99.
20. Fayyad, U. M., G. Piatetsky-Shapiro, and P. Smyth, ”The KDD process for extracting useful Knowledge from volumes of data,” Communications of the ACM 39 (11), November 1996, 2734.
21. Feng, J. Gao, K. Feng, L. Liu, and Y. Li. Active and passive nearest neighbor algorithm: A newly developed supervised classifier, pages 189–196. Springer Berlin Heidelberg, 2012.
22. G. J. Klir, ”Fuzzy arithmetic with requisite constraints”, Fuzzy Sets and Systems, 91:165175, 1997.
23. G. V. Nadiammai and M. Hemalatha, “Handling intrusion detection system using a snort based statistical algorithm and semi-supervised approach”, Research Journal of Applied Sciences, Engineering and Technology 6(16): 2914-2922, 2013.
24. G.V. Nadiammai, M. Hemalatha. The effective approach toward Intrusion Detection System using data mining techniques In Egyptian Informatics Journal (2014) 15, 37–50, ISSN: 1110-8665.
25. Gao Xiang, Wang Min. Applying semi-supervised cluster technique for anomaly detection. In: IEEE international symposium on information processing, no. 3; 2010. p. 43–5.