Real-Time System for Intrusion Detection Using Optimized Data- Mining Technique

Miss. Karishma Y. Zade.
Mtech CSE Student

Dr. Sanjeev Shrivastava.
Professor

Institute of Engineering & Technology, Nagpur, Maharashtra, India

ABSTRACT

In today's world of fast information technology to maintain the security of network system is important. Network infrastructure devices help greatly in managing various spines of any enterprise like bank accounts, online transaction, transportation, social insurance, protection, correspondence and computer networks systems; but at the same time may suffer from unauthorized access, data theft, network abuse, data modification and DOS (denial of service) attacks which prevent assurance of continuous service to legitimate users of the network.We need a secure and safe network system towards intruders attack. Intrusion detection system is used for identifying the various types of attack in a network. IDS are available in various types network based, host based and hybrid based on the technology detected by them in market. The existing system does not provide that quality of security, so we need a secure and reliable network system. In this paper, we present a review on Real time intrusion detection system (IDS) using data mining and some optimization techniques to efficiently detect various types of intruder attack.

Keywords: IDS, Monitoring, Data Security, Feature selection, Particle Swarm Optimization, Information Gain, Data Mining, Cloud Data

I. INTRODECTION

Intrusion detection is one of the most important techniques for protecting network security. In addition, intrusion detection model can be used to recognize real-time pattern, which has important practical significance for real-time intrusion detection. However, due to the sheer speed and scale of the data, data points must often be analyzed in real time. The one-pass-through requirement and the lack of efficient clustering algorithms to identify intrusion patterns limit the power and scalability of this approach. A data stream clustering algorithm is proposed for real-time network intrusion detection.

Internet technologies and the increase in the number of network attacks, intrusion detection has become a important research issue. Intrusion detection is dynamic research area. Due to remarkable progress and a large amount of work, there are still many opportunities to advance the state-of-the-art in detecting and thwarting network-based attacks. According to Anderson, an intrusion attempt or a threat is a unauthorized access to information, manipulate information, or render a system unreliable or unusable. For example, Denial of Service (DoS) attack attempts to deny a host of its resources, which are essential to work correctly during processing; Worms and viruses exploit other hosts through the

internet and Compromises obtain privileged access to a host by taking advantages of known vulnerabilities. anomaly-based intrusion detection refers to the problem of finding exceptional patterns in network database that do not conform to the expected normal behavior. Intrusion detection has extensive applications in fraud detection for credit cards, intrusion detection for enemy activities, for cyber security, and military surveillance

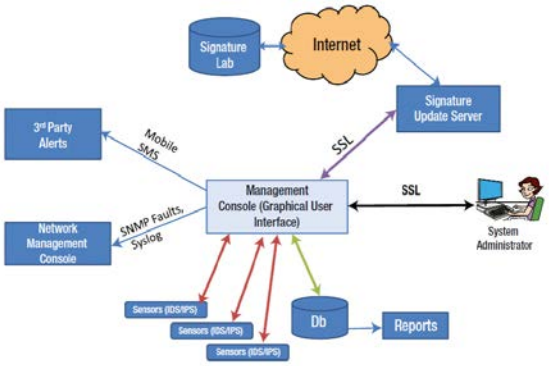

Fig. *IDS Architecture*

Intrusion detection systems (IDSs) have currently attracted the attention of a significant portion of academia, such that their development and enhancement represent a top priority for universities and research and science centers. An IDS is a computer and network security management system which monitors user activities and detects violations, misuses or suspicious activities in computers or computer networks. Lack of system security can naturally lead to performance disruption, deficiency or even temporary failure. In addition to firewalls and intrusion prevention devices, IDSs are also required to provide complete security in computer systems. There are 4 attack types described as fallows :

• Denial of Service Attack (DoS): In this type of attack, access or service provision to authorized users is attempted to be denied.

•User to Root Attack (U2R): In this type of attack, the intruder has access to a local machine but attempts to gain access to the system with administrator privileges.

•Remote to Local Attack (R2L): In this type of attack, an unauthorized user attempts to gain access to the victims system by guessing or cracking the password.

• Probing Attack (Probe): In this type of attack, attempts are made to steal data from the goal machine.

However, IDS is a dynamic one, which can provide dynamic protection to the security of network in monitoring, attack and counter-attack.

II. LITERATURE REVIEW

Priyanka Pawar et al. presents the performance of Neural Network for various values of number of clusters, based on experiments. The optimization of output is done using Particle Swarm Optimization (PSO) by selecting initial through PSO. Particle Swarm Optimization is used to optimize the output of our system, by appropriate selecting the input parameters through PSO.

An algorithm based on the Particle Swarm Optimization and Neural Network for analyzing program behaviour in intrusion detection is evaluated by experiments. Preliminary experiments with KDD cup'99 Data set show that the PSO optimized Neural Network can effectively detect intrusive attacks and achieves a low false positive rate. Ketan Sanjay Desale et al. presents the mechanism to improve the efficiency of the IDS using streaming data mining technique.

They apply four selected stream data classification algorithms on NSL-KDD datasets and compare their results. Based on the comparative analysis of their results best method is found out for efficiency improvement of IDS. Seyed Mojtaba Hosseini Bamakan et al. presents a new method based on multiple criteria linear

programming and particle swarm optimization to enhance the accuracy of attacks detection. Multiple criteria linear programming is a classification method based on mathematical programming which has been showed a potential ability to solve real-life data mining problems. However, tuning its parameters is an essential steps in training phase.

Particle swarm optimization (PSO) is a robust and simple to implement optimization technique has been used in order to improve the performance of MCLP classifier. KDD CUP 99 dataset used to evaluate the performance of proposed method.

The result demonstrated the proposed model has comparable performance based on detection rate, false alarm rate and running time compare to two other benchmark classifiers. Jaina Patel et al. proposed a hybrid model that integrates Anomaly based Intrusion detection technique with Signature based Intrusion detection technique is divided into two stages.

In first stage, the signature based IDS SNORT is used to generate alerts for anomaly data.

In second stage, data mining techniques "k-means + CART" is used to cascade kmeans clustering and CART (Classification and Regression Trees) for classifying normal and abnormal activities.

Architecture of genetic algorithm for IDS: It requires collecting network data for audit which contains normal and abnormal data. After collecting data, network sniffer will analyze the data and will send it to genetic algorithm. After applying fitness function, rules are added to rule set which are stored in rule base .

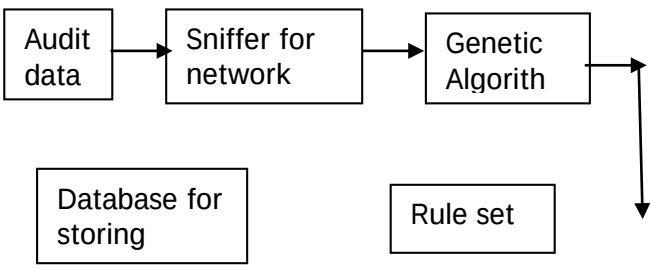

Fig : IDS using genetic algorithm

The hybrid IDS model is evaluated using KDD Cup Dataset. The proposed assemblage is introduced to maximize the effectiveness in identifying attacks and achieve high accuracy rate as well as low false alarm rate. G.V. Nadiammai et al. solved four issues such as Classification of Data, High Level of Human Interaction, Lack of Labeled Data, and Effectiveness of Distributed Denial of Service Attack using the proposed algorithms like EDADT algorithm, Hybrid IDS model, Semi-Supervised Approach and Varying HOPERAA Algorithm respectively. Our proposed algorithm has been tested using KDD Cup dataset. All the propose algorithm shows better accuracy and reduced false alarm rate when compared with existing algorithms

III. IDS WITH OPTIMIZATION TECHNIQUES

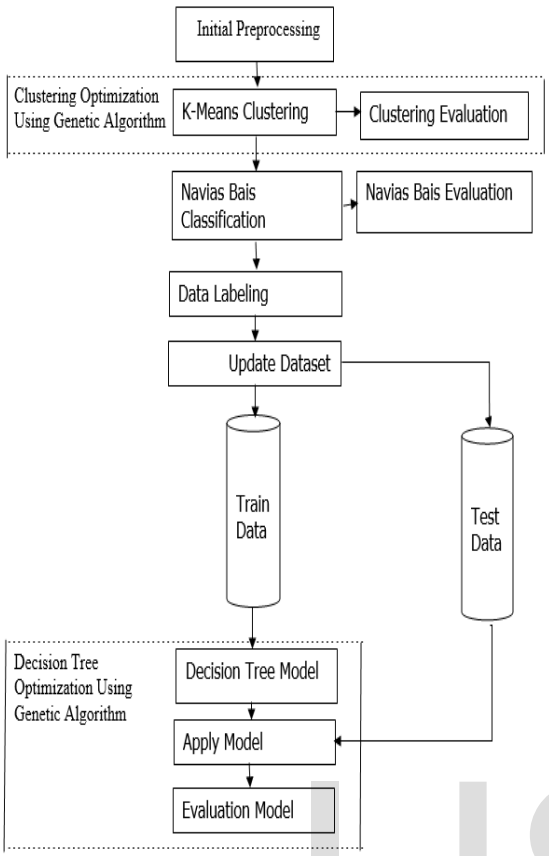

Fig: A Proposed Framework Scheme

The NEC method proposed by Chen et al., was studied in the previous section and an optimized framework is proposed in this section for improving the result of NEC. The suggested framework in this paper was based on the k-means clustering and decision tree (DT) algorithms And Navias Bais classification. The identification of the optimal number of Ks (number of clusters) and max _runs (maximum number of runs) in data clustering and the optimization of the DT algorithm in the data mining process were of interest in this framework. To this end, the genetic algorithm (GA) was used for clustering optimization following the initial preprocessing of the intended dataset. Clusters were used as labels for cluster members.

Thus, the dataset was updated and subsequently used in the classification process. The updated dataset which incorporated cluster labels was then used for data classification. The GA was used to optimize the confidence parameter, hence improving the quality of the DT classification process. Finally, the output of the suggest framework was presented. Given the application of clustering and classification techniques in this framework, it can be used for both labeled and unlabeled datasets, i.e. supervised and unsupervised methods. The suggested framework scheme is presented in Fig

A) Initial Preprocessing:

The initial preprocessing involves the following steps:

1) Uploading the dataset: In this step, the intended dataset should be uploaded to be used in datamining process.

2) Selecting features: In this step, the intended features from the dataset uploaded in the step 1 are selected. The selection basis by include a specific feature, a subset of features or all features.

3) Determining feature roles: In this step, feature roles are specified. “Roles” determine a specific features ID and whether it is regular, special, labeled, etc.

B. K-means Optimization Since the suggested framework was assumed to be applicable to unlabeled data and unsupervised methods, the clustering technique was used for data clustering following the initial data preprocessing.

C. Navias Bais: To detect an attack and detect automatically.

D. Data Labeling & Update Dataset The suggested framework in this paper was assumed to be applicable to supervised and unsupervised methods. In the previous section, the optimized clustering method was used to cluster data.

E. Train & Test Sub-Dataset Following the update of the primary dataset, train and test subsets were then segregated. Decision Tree Optimization the DT

performed a top-down scrolling of the applied records to the model in order to classify data.

IV. RESEARCH METHODOLOGY

Normally, statistical methods use a statistical model for normal behavior to the given data and then apply a statistical inference test to determine if an unseen instance belongs to this model low probability instances generated from the learnt model based on the applied test statistic are declared anomalies. Both parametric and nonparametric techniques have been applied in designing statistical models for anomaly detection. An example of statistical IDS is host based intrusion detection systems (HIDE). HIDE is an anomaly-based network intrusion detection system. It uses statistical models and neural network classifiers to detect intrusions.

B. Classification-based methods and systems Classification techniques are based on establishing an explicit or implicit model. It enables categorization of network traffic patterns into several classes. An example of classification-based IDS is Automated Data Analysis and Mining (ADAM). It provides a test bed for detecting anomalous instances.

C. Clustering and Outlier-based methods and systems Clustering is the task of assigning a set of objects into groups called clusters. The objects in the same cluster are more similar in some sense to each other than to those in other clusters. Clustering is used in data mining. Outliers are that point in a dataset that are highly unlikely to occur given a model of the data, For example, MINDS (Minnesota Intrusion Detection System) is a data mining-based system for detecting network intrusions.

D. Soft computing methods and systems Soft computing techniques are needed for network anomaly detection. Soft computing is generally thought of as encompassing methods such as Genetic Algorithms, Artificial Neural Networks, Fuzzy Sets, Rough Sets, Ant Colony Algorithms and Artificial Immune Systems.

1. Genetic algorithm approaches: Genetic algorithms are population-based adaptive heuristic search techniques. It is based on evolutionary ideas.

2. Artificial Neural Network approaches: Artificial Neural Networks (ANN) is motivated by the recognition that the human brain computes in an entirely different way from the conventional digital computer. An example of ANN-based IDS is RT-UNNID. This system is capable of intelligent real time intrusion detection using unsupervised neural networks (UNN).

3. Fuzzy set theoretic approaches: Fuzzy network intrusion detection systems exploit fuzzy rules. it determine the likelihood of specific or general network attacks. A fuzzy input set can be defined for traffic in a specific network. NFIDS is a neuron-fuzzy anomaly based network intrusion detection system.

V. APPLICATION OF DATA MINING IN INTRUSION DETECTION

In classical IDS, security experts firstly categorize attacking actions and system weakness, select statistical approaches because of detecting kinds, then manually enter 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 increase the effectiveness of detecting the intrusion detecting modes. Safety experts most often predicament in regards to the known attacking aspects and approach weak spot and study on that, which motives the dearth of adaptability of the detecting sample to the unknown intrusion that procedure is set to be dealing with. Meanwhile, lengthy upgrade protection

method cycle, the excessive price, these aren't fine for bettering the adaptability of intrusion detecting pattern.

As the experts' rules and statistical approaches often need hardware and software support, it stops the system from reusing and increasing in novel atmosphere, meanwhile it causes the difficulty of embedding new detecting modules. All of these are not good for gaining scalability of intrusion detecting pattern. Therefore, it has become significant 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 creates IDS can be useful to any computer environment. This method has become a famous research topic, in inter discipline field of network security

Intrusions are the activities that violate the security norms of system. An IDS is Mechanism used to identify, screen network or process actions for malicious hobbies and produces reviews to a administration departments. The development of IDS is influenced through following causes: Most current methods have protection was once that render them susceptible to intrusions, and fixing and finding each these deficiencies aren't viable. Prevention methods cannot be ample. It's close to inconceivable to have an undoubtedly relaxed procedure. Even essentially the most secure systems are prone to insider attacks. New intrusions always emerge and novel ways are required to defend towards them.

VI. CONCLUSION

This paper shows the study about intrusion detection system with its application and drawback. We focus on genetic based intrusion detection and other swarm intelligence based technique so that performance of IDS can improve. Here our system is so efficient then the previous system because we are using two level of dataminig where we get the best , fast and accurate date for the analysis and create the proper alert system. The IDS with genetic algorithm and decision tree is also explained by flow chart. How to represent chromosome in GA is also explained in brief.

VII. REFERENCES

1. T.-S. Chou, Ensemble fuzzy belief intrusion detection design. Florida International University, 2007. S. Mc Elwee, "Active learning intrusion detection using kmeans clustering selection," in South east Con, 2017, 2017, pp. 1–7

2. S. Saha, A. S. Sairam, A. Yadav, and A. Ekbal, "Genetic algorithm combined with support vector machine for building an intrusion detection system," in Proceedings of the International Conference on Advances in Computing, Communications and Informatics, 2012, pp. 566–572.

3. A. P. Muniyandi, R. Rajeswari, and R. Rajaram, "Network anomaly detection by cascading k-Means clustering and C4. 5 decision tree algorithm," Procedia Eng., vol. 30, pp. 174–182, 2012.

4. Y. jun Zhao, M. jun Wei, and J. Wang, "Realization of intrusion detection system based on the improved data mining technology," in Computer Science & Education (ICCSE), 2013 8th International Conference on, 2013, pp. 982–987.

5. H. M. Tahir, A. M. Said, N. H. Osman, N. H. Zakaria, P. N. M. Sabri, and N. Katuk, "Oving K-Means Clustering using discretization technique in Network Intrusion Detection System," in Computer and Information

Sciences (ICCOINS), 2016 3rd International Conference on, 2016, pp. 248–252.

6. A. Sundaram, “An introduction to intrusion detection,” Crossroads, vol. 2, no. 4, pp. 3–7, April 1996. J. P. Anderson, “Computer Security Threat Monitoring and Surveillance,” James P Anderson Co, Fort Washington, Pennsylvania, Tech. Rep., April 1980.

7. A.M. Chandrasekhar, “Intrusion Detection Technique By Using K-Means, Fuzzy Neural And Svm Classifier “, 2013 International Conference on Computer Communication and Informatics (ICCCI - 2013), Jan 04-06, 2013 Coimbatore, India.

8. Hesham Altwaijry, “Bayesian Based Intrusion Detection System “, Journal of King Saud University – Computer and Information Sciences (2012) 24,1–6.

9. Ammar Boulaiche, “A Quantitative Approach For Intrusions Detection And Prevention Based On Statistical N-Gram Models “, Procedia Computer Science 10 (2012) 450 – 457.