ISSN:0975-9646

Reduction and Classification of Feature of Cyber Attack Data: Brief Review

Himanshu Shroti
M.Tech Scholar,Department of CSE,
TRUBA Institute of Engineering and Information
Technology Bhopal (M.P.), India

Kaptan Singh
Professor,Department of CSE,
TRUBA Institute of Engineering and Information
Technology Bhopal (M.P.), India

Abstract— **The detection and prediction of dynamic types of attack is always challenging task for the intrusion detection system. For the detection of these attacks, used distinct data mining technique such as classification and clustering. The process of clustering and classification is very complex due to immense number of feature attribute of network file and data. Now the reduction of feature attributes of intrusion increase the potency of classification and detection of dynamic types of data. Now in current decade various feature reduction method are used such as LDA, Genetic algorithm and other heuristic based optimization technique. In this paper present the review of intrusion detection technique based on the process of feature reduction and classification based technique**.

Keywords— **IDS, Data Mining, Feature Reduction, Swarm Intelligence**

1. INTRODUCTION

The feature reduction is major issue in analysis of network intrusion data. The reduced feature set improved the classification and detection ratio of intrusion detection system. The reduction policy of feature set depends on the basis of static and dynamic. In process of static feature reduction set some feature point for the elimination of data during the process of classification and detection. In dynamic feature reduction, the reduction method decides run time feature selection for the detection mechanism. Some feature reduction and classification technique reduces some meaning full information and faced a problem of detection ratio. Various authors and scientist used pattern recognition based algorithm such as LDA, LDPA and some other technique. The classification technique plays an important role in detection of intrusion. The process of classification offers various data mining technique such as support vector machine, KNN, Decision tree and neural network based classification technique. The data mining based classification technique extended in terms of ensemble classifier. The ensemble based classifier merge two different classifier for the improvement of the classification ratio. Now a day's swarm based feature reduction technique are used for the process of feature optimization. The family of swarm offer various algorithm such as particle swarm optimization, ant colony optimization and many algorithm based on the behaviors of kits. The particle swarm optimization optimized the feature set for the process of classification. The particle swarm optimization work in two modes one is local mode and other is global mode for the selection of feature. The family of swarm algorithm based on the concept of memory and it's a iterative process. Some suspicious attack is very complex in the phase of detection, such types of serious attack estimated by the government of different country for the prevention and detection. The growth of finical and social website creates more Trojan based anomaly software and spread the information of threats around the word. The process of detection technique in these cases affects the performance of intrusion detection system. In section II. Discuss the related work. In section III. Discuss the problem formulation. In section IV. Discuss swarm intelligence and finally conclude in section V.

2. RELATED WORK

In this section discuss the related work in the field of intrusion detection. Now a day's intrusion detection is critical issue in network and web security. The behaviors of intrusion impact the performance of network as well as computer health. Here discuss some related work in concern of intrusion detection.

Alvaro A. Cardenas, Robin Berthier, Rakesh B. Bobba, Jun Ho Huh, Jorjeta G. Jetcheva, David Grochocki and William H. Sanders [1] Et al. The proposed model assists the cost based analysis for measuring security risk factor. The proposed model monitors the process of network and the basis of log information of server. The threats of network security analysis on the basis of risk parameter. The risk parameter estimated the correlation factor for the estimation of attack data for the prevention purpose of network data.

Junho Hong, Chen-Ching Liu and Manimaran Govindarasu [2 Et al. proposed the anomaly detection system for the automation of substation. The anomaly detection system integrated the power of host and network based system. The detection of host and network based system integrated the sharing information and apply mining technique for the processing of intrusion detection system. The proposed ADS system fetches the information from the footprint of malicious software. The extracted feature of malicious software gives the statically information for the estimation of information.

Preeti Singh and Amrish Tiwari [3] Et al. used the process of fast feature reduction technique based on partially ID3 algorithm to find maximum information gain for attribute selection and KNN based GA (genetic algorithm) is applied for classification and revelation of invasion on KDD dataset. The intrusion detection helps security organization accordingly by enhancing the efficiency and it is easy to use. In this analysis they introduce an efficient method for intrusion detection using ID3, KNN (K-nearest neighbor) and genetic algorithm (GA). The ID3 is used for the feature reduction among the large set of data then apply KNNGA together which improve the data categorization due this the performance increases.

Gaby Abou Haidar and Charbel Boustany [4] Et al. Anomaly depend network intrusion detection design are taking the consideration in the majority of networking sites to secure interconnection systems in contrast to awful act, affording a marvelous defense level specifically with the coordination of neural networks in the detection systems which provided advanced methods of threats investigation and detection. Many studies have been elaborated concerning the intrusion inspection and interdict evolving subject. They pin-pointed the dominant profit and key connection behind the network mobile agents models and we clarified its importance in the intrusion detection and prevention. Finally, we introduced by representing the information sharing concept that is considered as one of the almost robust and efficient intrusion detection.

D.P.Gaikwad and Ravindra C. Thool [5] Et al. proposed machine learning ensemble method which is based on intrusion detection technique. Intrusion detection system can be implemented by bagging method of ensemble in which REPTree is used as base class. The significant features from NSL_KDD dataset are elegant to enhance the classification accuracy and decrease the rate of false positive. The Ensemble Bagging method of machine learning for intrusion detection system is presented. The Bagging with REPTree base classifier is proposed for detection of anomaly packet over network. This method is evaluated on test dataset and cross validation of 10-fold. The performances of classifies are calculated in parameters of precision, false positives and building time model.

Robin Berthier and William H. Sanders [6] Et al. suggest a specification-related intrusion invasion analyzer that can be arranged in the held to identify security threats in real time. This sensor supervise the traffic mutually accent and entrance points at the application, transport, and network layers to assure that resources are operative in a secure state and their process favor a stated security policy. The explanation waits on code requirements, precaution demand, and a discursive security plan to recognize security violations and trigger alerts when suspicious activity occurs. The security requirements were developed using a well-ordered approach that combined a threat model, a system model, and a set of network traces that capture valid use case behavior.

Robin Berthier and William H. Sanders [7] Et al. specification-based intrusion detection system that hold tightened control over endorse AMI protocol influence to sustain an efficient network situational observance resolution for AMI operators. A unique aspect of this project is that it leverages an industry-established set of realistic failure scenarios to specify an extensive defense policy. The policy, which includes of 23 rules, was described and translated into machine-checkable constraints. Several months of deployment of Amilyzer at the headend of a large AMI enabled us to increase the scalability and reliability of the IDS. Amilyzer has reached a development phase at which it is starting to consign the pressing need for an efficient and practical monitoring solution for AMIs.

Adel Nadjaran Toosi and Mohsen Kahani [8] Et al. parallel neuro-fuzzy classifiers are used to do an initial classification. The fuzzy inference system would then be related on the turnout of neuro-fuzzy classifiers, making final opinion of either the prevailing action is normal or intrusive. Lastly, in order to realize the best result, genetic algorithm upgrades the anatomy of our fuzzy decision engine. The ANFIS arrangement was utilize as a neuro-fuzzy classifier for intrusion verification. ANFIS is able of yielding fuzzy rules lacking the help of human experts. Also, ablative clustering has been applied to work out the number of rules and associate functions with their beginning locations for more valuable classification.

Nasim Beigi Mohammadi, Jelena Misic, Vojislav B. Misic and Hamzeh Khazaei [9] Et al. They presented security challenges, existing security measures, and IDS solutions for AMI. We proposed a mixed anomaly and signature-based IDS result to scan the smart metering communication network by estimating separate attacks striving physical, MAC, transport, and network layers. They gives the security analysis based tools for the measuring the risk factor for the intrusion detection. The proposed model works on the basis of signature based pattern analysis for the measuring of security threats. The authors designed the smart meter for the detection of attack.

Massimo Ficco, Salvatore Venticinque and Beniamino Di Martino [10] Et al. proposed MOSAIC-depend architecture for specifying appropriate incursion exploration in Cloud Computing. It is a fundamental framework that accumulates knowledge at various Cloud architectural levels, adopting numerous surveillance constituents, which are effectively set up as a distributed architecture. Overlaying the complication of Cloud architecture, this analysis affirms a framework to develop distributed Intrusion Detection model in the Cloud Computing.

Mustafa Amir Faisal, Zeyar Aung, John R. Williams and Abel Sanchez [11] Et al. The proposed system designed for the detection of intrusion detection system in power system substation. The intrusion estimated the information about sensor based digital meter. The sensor data discriminate the behavior of file for the analysis of pattern for the given prediction system. The behavior of traffic is differentiating on the basis of group rule about the information track for the analysis process.

Vincenzo Gulisano, Magnus Almgren and Marina Papatriantafilou [12] Et al. A considerable number of applications surrounding AMIs depend on the data produced by AMIs' devices. It's noisy and lossy nature demands for validation analysis in order to preprocess the data that is later accessed by utilities' or third parties' applications.

Gianluigi Folino, Clara Pizzuti and Giandomenico Spezzano [13] Et al. The authors proposed a distributed model for the collection of the information sharing. Each distributed node act as genetic data collector. The data collector of distributed node used genetic algorithm for the making the decision node is suspicious or normal activate node. The concept of distributed intrusion detection system used the concept of ensemble classification technique. The ensemble based classifier used the process of merging two different independent node data on the basis of decision factor.

Edward Guillen, Jhordany Rodriguez and Rafael Paez [14] Et al. This work is focused on the detection of non-content based attacks and it is possible that the detection of content based attacks requires more complex ML unlike the results shown in this work. Because the correct extraction and selection of attributes are very important, the next step is the implementation of strategies for content based attacks analysis.

Mustafa Amir Faisal, Zeyar Aung, John R. Williams and Abel Sanchez [15] Et al. analyze the possibility of using data stream mining for enhancing the security of AMI through an Intrusion Detection System (IDS), which is a second line of defense after the primary security methods of encryption, authentication, authorization, etc. They recommend a genuine and predictable IDS make up for the entire AMI complex which embodies of separate IDSs for three uncommon leveled of AMI's units: data concentrator, smart meter and AMI headend. They additionally scoop out the effectiveness of certain remaining precocious data stream mining theorem on a commonly feasible IDS dataset, that is, the KDD Cup 1999 dataset.

3. Problem Formulation

Here discuss the problem related to large number of feature attribute. These attribute impact the performance of detection and classification of unknown and dynamic types of attribute. For the reduction of feature attribute several techniques are adopted such as LDP, dynamic programming, swarm intelligence and many more algorithms. The large set of attribute contains two types of feature attribute one is static attribute and another is dynamic attribute. The static attribute used in both cases normal communication plus attack scenario. The dynamic attributes comes only one the case of unknown attack and anomaly attack. Here discuss some common problem related to intrusion detection system.

1. Take more processing time for data conversion
2. Categorization of feature attribute on basis of static and dynamic
3. Selection of feature attribute for the proceeding of classification
4. Feature faced a problem of null value
5. Ambiguity nature of classification

4. Swarm Intelligence

In this section refers the behaviors of swarm intelligence. Basically it is a group of biological inspired kits behaviors. In the bundle of it various feature reduction and feature optimization algorithm are available. In concern of swarm family three most famous algorithms such as particle swarm optimization, ant colony optimization, honey bee and glowworm algorithm. The PSO based on the concept of bird fork. In the sky bird are moving with constant velocity and never collide each other. The particle swarm optimization algorithm stayed on two functions one is Gbest and other is Pbest. The Gbest function is global solution of given problem domain and Pbest is local solution of given problem domain. The value of Pbest=Gbest then the stage of feature optimized. Instead of particle swarm optimization, ant colony optimization stationed on the apprehension of biological ant behaviors. The biological ants continually perceive the unprolonged course form source to food location. The ant colony optimization technique immolates the feature with regard to similarity and dissimilarity. Glowworm optimization based on the concept of glow kits, gives the lighting condition in night and collect neighbors glow for the sharing of information.

5. Conclusion And Future Scope

In view of this study give an introduction about the review of feature reduction and selection based on different data mining technique and swarm intelligence algorithm. The extended amount of feature attribute of intrusion file creates a problem of attribute processing for the detection and classification process. For the reduction of feature used various algorithms such as LDA, genetic algorithm and some other algorithm. the swarm intelligence and neural network based feature reduction technique is also important technique. in future used glowworm optimization technique for the reduction of feature. The glowworm optimization algorithm gives the better reduction possibility instead of another feature reduction technique.

References

[1] Alvaro A. Cárdenas, Robin Berthier, Rakesh B. Bobba, Jun Ho Huh, Jorjeta G. Jetcheva, David Grochocki and William H. Sanders *"A Framework for Evaluating Intrusion Detection Architectures in Advanced Metering Infrastructures"*, IEEE, 2014, Pp 906-915.

[2] Junho Hong, Chen-Ching Liu and Manimaran Govindarasu *"Integrated Anomaly Detection for Cyber Security of the Substations"*, IEEE, 2014, Pp 1-11.

[3] Preeti Singh and Amrish Tiwari *"An Efficient Approach for Intrusion Detection in Reduced Features of KDD99 using ID3 and classification with KNNGA"*, IEEE, 2015, Pp 445-452.

[4] Gaby Abou Haidar and Charbel Boustany *"High Perception Intrusion Detection Systems Using Neural Networks"*, IEEE, 2015, Pp 497-501.

[5] D.P.Gaikwad and Ravindra C. Thool *"Intrusion Detection System Using Bagging Ensemble Method of Machine Learning"*, IEEE, 2015, Pp 291-295.

[6] Robin Berthier and William H. Sanders *"Specification-based Intrusion Detection for Advanced Metering Infrastructures"*, IEEE, 2011, Pp 184-193.

[7] Robin Berthier and William H. Sanders *"Monitoring Advanced Metering Infrastructures with Amilyzer"*, IEEE, 2013, Pp 1-13.

[8] Adel Nadjaran Toosi and Mohsen Kahani *"A new approach to intrusion detection based on an evolutionary soft computing model using neuro-fuzzy classifiers", Computer Communications,* 2007, Pp 2201-2212.

[9] Nasim Beigi Mohammadi, Jelena Misic, Vojislav B. Misic and Hamzeh Khazaei *"A framework for intrusion detection system in advanced metering infrastructure",* SECURITY AND COMMUNICATION NETWORKS, 2012, Pp 195-205.

[10] Massimo Ficco, Salvatore Venticinque and Beniamino Di Martino *"MOSAIC-Based Intrusion Detection Framework for Cloud Computing"*, COMPUTER SYSTEMS SCIENCE AND ENGINEERING, 2012, Pp 1-17.

[11] Mustafa Amir Faisal, Zeyar Aung, John R. Williams and Abel Sanchez *"Securing Advanced Metering Infrastructure Using Intrusion Detection System with Data Stream Mining",* Springer-Verlag Berlin Heidelberg, 2012, Pp 96-111.

[12] Vincenzo Gulisano, Magnus Almgren and Marina Papatriantafilou *"Online and Scalable Data Validation in Advanced Metering Infrastructures",* IEEE, 2014, Pp 1-6.

[13] Gianluigi Folino, Clara Pizzuti and Giandomenico Spezzano *"GP Ensemble for Distributed Intrusion Detection Systems",* Springer Berlin Heidelberg, 2005, Pp 54-62.

[14] Edward Guillen, Jhordany Rodriguez and Rafael Paez *"Evaluating Performance of an Anomaly Detection Module with Artificial Neural Network Implementation"*, International Scholarly and Scientific Research & Innovation, 2013, Pp 1484-1490.

[15] Mustafa Amir Faisal, Zeyar Aung, John R. Williams and Abel Sanchez *"Data Stream-based Intrusion Detection System for Advanced Metering Infrastructure in Smart Grid: A Feasibility Study"*, IEEE, 2014, Pp 1-14.