ISSN 2395-1621

Intrusion Detection System using Soft Computing and Machine learning Algorithm

Abhishek Rameshwar Vaidya,Vikrant Ujjwal Karawande,
Parshwa Bharat Shah,Kunal Sureshkumar Gaikwad,
Asst Prof Anand D. Dhawale.

ABSTRACT

Abstract: Big Data Cyber security Analytics is increasingly becoming an important area of research and practice aimed at protecting networks, computers, and data from unauthorized access by analyzing security event data using big data tools and technologies. Whilst a plethora of Big Data Cyber security Analytic Systems has been reported in the literature, there is a lack of a systematic and comprehensive review of the literature from an architectural perspective. In this paper, we formulate the SVM configuration process as a bi-objective optimization problem in which accuracy and model complexity are considered as two conflicting objectives. System proposes a novel hyper-heuristic framework for bi-objective optimization that is independent of the problem domain. This is the first time that a hyper-heuristic has been developed for this problem. The proposed hyper-heuristic framework consists of a high-level strategy and low-level heuristics. The high-level strategy uses the search performance to control the selection of which low-level heuristic should be used to generate a new SVM configuration. The low-level heuristics each use different rules to effectively explore the SVM configuration search space. To address bi-objective optimization, the proposed framework adaptively integrates the strengths of decomposition- and Pareto based approaches to approximate the Pareto set of SVM configurations. The effectiveness of the proposed framework has been evaluated on two cyber security problems: Microsoft malware big data classification and anomaly intrusion detection.

ARTICLE INFO

Article History

Received: 8th March 2020

Received in revised form :

8th March 2020

Accepted: 10th March 2020

Published online :

11th March 2020

Keywords: Hyper-heuristics, big data, cyber security, optimization.

I. INTRODUCTION

The high-level strategy operates on the heuristic space instead of the solution space. In each iteration, the high-level strategy selects a heuristic from the existing pool of low-level heuristics, applies it to the current solution to produce a new solution and then decides whether to accept the new solution. The low level heuristics constitute a set of problem-specific heuristics that operate directly on the solution space of a given problem. In proposed work we define three different layers for detection the malicious data or connection using SVM and evolutionary base heuristic approach.

The proposed system carried out multi objective heuristic approach to detect the malicious attack from network environment. Initially system deals with training face where the background knowledge has generated from various network dataset. KDD Cup 99 dataset has used to extract the basic features of network attack and stored those features in train model. In strategic approach system evaluate search network packet using support vector machine (SVM), the system works like supervised learning approach for label classification so, it needs to generate a background knowledge before evaluate the test instances. In this work system first execute data preprocessing as well as data normalization. Once the background knowledge has generated buy system it is purely applicable for testing, interesting phase we have written heuristic kernel function for evaluate to each test object. The background knowledge has used to generate the runtime similarity for each known as well as unknown type of attacks.

II. LITARATURE SURVEY

According to Soheily-Khah, Saeid, Pierre-François Marteau[1].Data mining techniques play an increasing role in the intrusion detection by analyzing network data and classifying it as 'normal' or 'intrusion'. In recent years,

several data mining techniques such as supervised, semi-supervised and unsupervised learning are widely used to enhance the intrusion detection. This work proposes a hybrid intrusion detection (kM-RF) which outperforms in overall, according to our experimentation, the alternative methods through the accuracy, detection rate and false alarm rate. A benchmark intrusion detection dataset (ISCX) is used to evaluate the efficiency of the kM-RF, and a deep analysis is conducted to study the impact of the importance of each feature defined in the pre-processing step.

According to Alaei, Parisa, and Fakhroddin Noorbehbahani [2].With the proliferation of the internet and increased global access to online media, cybercrime is also occurring at an increasing rate. Currently, both personal users and companies are vulnerable to cybercrime. A number of tools including firewalls and Intrusion Detection Systems (IDS) can be used as defense mechanisms. A firewall acts as a checkpoint which allows packets to pass through according to predetermined conditions. In extreme cases, it may even disconnect all network traffic. An IDS, on the other hand, automates the monitoring process in computer networks. The streaming nature of data in computer networks poses a significant challenge in building IDS. In this paper, a method is proposed to overcome this problem by performing online classification on datasets. In doing so, an incremental naive Bayesian classifier is employed. Furthermore, active learning enables solving the problem using a small set of labeled data points which are often very expensive to acquire. The proposed method includes two groups of actions i.e. offline and online. The former involves data preprocessing while the latter introduces the NADAL online method. The proposed method is compared to the incremental naive Bayesian classifier using the NSL-KDD standard dataset. There are three advantages with the proposed method: (1) overcoming the streaming data challenge; (2) reducing the high cost associated with instance labeling; and (3) improved accuracy and Kappa compared to the incremental naive Bayesian approach. Thus, the method is well-suited to IDS applications.

According to Falcón-Cardona, Jesús Guillermo et al. [3]in recent years, Indicator-based Multi-Objective Evolutionary Algorithms (IB-MOEAs) have become a relatively popular alternative for solving multi-objective optimization problems. IB-MOEAs are normally based on the use of a single performance indicator. However, the effect of the combination of multiple performance indicators for selecting solutions is a topic that has rarely been explored. A hyper-heuristic which combines the strengths and compensates for the weaknesses of four density estimators based on R2, IGD, and p. The selection of the indicator to be used at a particular moment during the search is done using online learning and a Markov chain. Additionally, a novel framework that aims to reduce the computational cost involved in the calculation of the indicator contributions. Our experimental results indicate that our proposed approach can outperform state-of-the-art MOEAs based on decomposition (MOEA/D) reference points (NSGA-III) and the R2 indicator (R2-EMOA) for problems with both few and many objectives.

According to Rahul, Vigneswaran K., et al. [4]Intrusion detection system (IDS) has become an essential layer in all the latest ICT system due to an urge towards cyber safety in the day-to-day world. Reasons including uncertainty in finding the types of attacks and increased the complexity of advanced cyber-attacks, IDS calls for the need of integration of Deep Neural Networks (DNNs). In this paper, DNNs have been utilized to predict the attacks on Network Intrusion Detection System (N-IDS). A DNN with 0.1 rate of learning is applied and is run for 1000 number of epochs and KDDCup-`99' dataset has been used for training and benchmarking the network. For comparison purposes, the training is done on the same dataset with several other classical machine learning algorithms and DNN of layers ranging from 1 to 5. The results were compared and concluded that a DNN of 3 layers has superior performance over all the other classical machine learning algorithms.

According to Gaied, Imen, Farah Jemili, et.al [5]there is no standard solution we can use to completely protect against computer network intrusion. Every solution has its advantages and drawbacks. Soft computing is considered as a promising paradigm to cope with the dynamic evolution of networks. In previous works, we presented two soft computing approaches of intrusion detection. The first one is based on the neuro-fuzzy and the second one is based on the genetic fuzzy one. In this work, we elaborate an empirical comparative study to highlight the benefits of each method in intrusion detection and exploit their complementarities to enhance the detection rate of all types of attacks as well as decrease the false positives rate.

According to Potteti, Sumalatha, and Namita Parati.[6]The describes how Hybrid IDS is used in Fuzzy Genetic algorithm in wireless networks or networks. These days Intrusion Detection System (IDS) which is define d as a solution of system security is employed to identify the abnormal activities in a computer system or network. The IDS is to detect the attacks and generate the proper response. The drawback of the anomaly based intrusion detection in a wireless network is the high rate of false positive. By designing a hybrid intrusion detection system can solve this by connecting a misuse detection module to the anomaly detection module. In this paper, we propose to develop a hybrid intrusion detection system for wireless local area networks based on Fuzzy Genetic logic. The proposed Fuzzy Genetic logic-based system could be able to detect the intrusive activities of the computer networks as the rule base holds a better set of rules.

According to Mukane, Rohit V. et.al [7] Rotating machine is device that supports an important category of manufacturing industry. Machine fault detection is very advantageous for diagnosis of faults before it occurs and to protect the machinery from catastrophic damages. Vibration monitoring is important for running rotating equipment evenly for years. There are diverse reasons that cause vibrations in the rotating machinery like rotor unbalance, machine looseness, permanent bow (bend shaft or warped shaft), bearing damage, misalignment, eccentricity, oil whirl, cracked tooth, etc. There are several methods used to detect machinery faults such as, model based technique and signal based techniques which includes Hilbert-Huang Transform (HHT), Wavelet Transform (WT), Artificial Neural Networks (ANN), Support Vector Machine (SVM), Shock Pulse Monitoring (SPM), Fuzzy Logic. The proposed work shows a novel approach of classification using fuzzy logic for decision making about the machine condition. The

experimental results show that, algorithm implemented identifies severity of faults to appropriate condition.

According to Behera, Santi Kumari, et al. [8] a suggests a computer vision based system which have ability to identify deformity in the orange fruits and also organize the flaw type appeared on the surface of orange fruit. The symptoms of flaw mark imply the seriousness of the disease and recommend the optimal approach to deal with the disease. It's conjointly required to diagnose the disease properly with prior to great damage by providing proper treatment. Further, estimation of severity of disease is required for applying proper amount of pesticides to avoid the environmental pollution and economic burden. Here we use multi class SVM with K-means clustering for classification of diseases with 90% of accuracy and Fuzzy logic to compute the degree of disease severity.

According to Theresa, W. Gracy, and S. Sakthivelet.al [9] MANETs are potential wireless network, where mobile nodes are connected dynamically in ad hoc basis. This unique characteristic attracted many promising applications and the one among is battlefield communication. The war troops at the edge of the network do not have any computing infrastructure for communication. Therefore MANET plays a vital role in battlefield communication. Since MANET is openness to eavesdropping, routing of information causes vulnerabilities and degrades the performance of network. This necessitated developing a MANET with a novel intrusion detection system to provide reliability and security to the battlefield communication. A Di-Fuzzy logic technique which provides two phase detection for intrusion detection in the network. It works in a cluster based routing environment to co-ordinate and control the entire network. The selection of cluster head is based on the node with the maximum energy to improve the life time of the network. The proposed technique is simulated in network simulator NS 2.8 and the performance is evaluated by comparing with the existing Fuzzy based IDS (F-IDS) & intrusion detection and adaptive response (IDAR) system. From comparison the proposed Di-Fuzzy logic technique improves its performance in all metrics and thus provides a safe environment for battlefield communication.

According to Alqahtani, Saeed M., and Robert John [10] Intrusion detection system (IDS) as one of huge research problem in network security is the most effective tool of protection. It is a method of parsing network traffic data to detect security abuses. Data mining can play a very significant role in evolving an IDS. The dataset of IDSs or soft computing techniques based IDS can be classified into normal and abnormal traffic in order for generated alerts to detect threats. In this paper, we utilized the most common classification algorithms: Decision Tree (J48), Naive Bayes, OneR, and K-Nearest Neighbor (K-NN). These algorithms were chosen after investigating the most effective classification algorithms that are widely used. The aim of this study is to present a comparative study for the performance of each system that was gained from our previous experiments: Snort IDS, Suricata IDS, FL-Snort IDS, and FL-Suricata IDS in order to test which classifier algorithm is the best for our systems results, and investigate which system presents significant results. The performance of these classification algorithms was evaluated using 10-fold cross validation. Experiments and assessments of these methods were performed in the WEKA environment using the ISCX dataset.

III. OBJECTIVES OF SYSTEM

The goal of proposed Bi-objective Hyper-Heuristic system is to maximize the detection accuracy, to minimize false positive rate and detector generation time. The Objective of the proposed application is as follows:

- To design and implement a Bi-objective Hyper-Heuristic system using SVM and FGA in big data environment.
- To improve the performance of overall network
- To detect all types of attacks in online as well as offline environment like NIDS and HIDS. (e,g, DOS, PROBE, U2R, R2L, Unknown)
- Define the security and privacy in wireless network virtualization over the network.

IV. PROPOSED METHODOLOGY

In the proposed research work to design and implement a system for The proposed algorithm identifies the noisy instances and distinguishes them from the instances that are in class boundary. The goal of the algorithm is to identify and eliminate the noisy instances, preserving the class distribution and the classes boundaries such that the neither separate of the classes nor the discriminate power of the classification algorithm is altered.

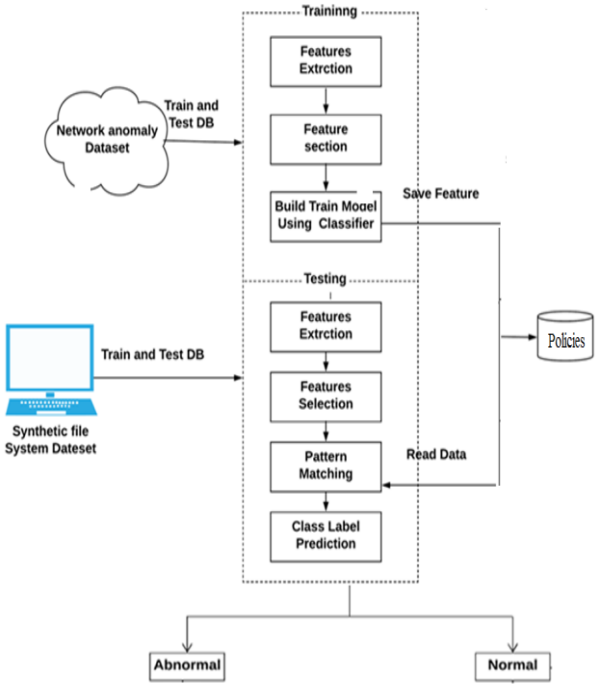

Figure 1 : Proposed System Architecture

In this system we detect the two type of noise in our proposed work after the training phase has successfully done.

Noise in the attributes: It is given by the errors occurred during the entrance of the values of the attributes. Among

the sources of this type of noise are: variables with missing values, and redundant data.

Noise in the classes: It is given by the errors introduced during the assignment of the instances to the classes. The presence of this kind of noise may be due to subjectivity, errors in the data entry process, and incorrect information for assigning a instance to a class. There are two possible sources of class noise.

V. SYSTEM ANALYSIS

Algorithm 1 : SVM base Rule policies generation algorithm

Input: Training set from network log or data packets, Attribute validation policies Background Knowledge (BK) policies, Threshold 'th'.

Output: Rule set as policies or signatures.

Step 1: a. Read values from data file header fields to get Feature Vector.

b. Read data from a network connection to append in Feature []

Step 2: Validate each attribute for the preprocessing phase

Step 3: Normalized irrelevant attribute, and get normalized set NormSet[] □ {Att[i……n]}

Step 4 : for each (Feature into NormSet !=Null**)**

Step 5 : calculate weight w= (Feature, Bk)

Step 6 : if (w>=th)

Ruleset.add □ {Feature,Label}

End if

End for

Step 7 : return Ruleset.

Algorithm 2: Weight calculation for test instances

Input: Feature of BK rules TrainF[], features if test recordTestF[]

Output: highest Similarity weight for class label

Step 1: Read all training rules from DB for each (Rec R into Train[])!=Null

Step 2: items [] split(R)

Step 3: items1 [] split(TestF)

Step 4: CalculateWeight(DB [i], items1)

Step 5: Return w;

Algorithm 3: SVM

Input : Feature of BK rules TrainF[], features if test recordTestF[]

Output : highest Similarity weight for class label

Step 1: for all (T in TrainF [] !=null) do

Step 2: items [] split(T)

Step 3: items1 [] split(TestF)

Step 4: w = classifyToAll(Train,TestF[], Label)

Step 5: Return w;

Mathematical Model

A System has represented by a 5−different phases, each phase works with own dependency System S = (Q, ∑, δ, q_0, F) where −

- **Q** is a finite set of states.
- **∑** is a finite set of symbols called the alphabet.
- **Δ** is the transition function where δ : Q × ∑ → Q
- **q_0** is the initial state from where any input is processed (q_0 ∈ Q).
- **F** is a set of final state/states of Q (F ⊆ Q).

All t(n) policies will return 1 then from training patterns and it generate the similarity weight of fitness function of specific rules.

Q = {Pop[i=0……..n]} set of population size initial set

∑ = {crossoverRate, MutationRate, Fitness etc}

Δ = {Tp*100 / Sun F(x)}

q_0 = {intial population given to crossover function ∑ i=0 }

F = {Attack - Normal}

System has define with 3 stages

Train : if training has successfully done then it returns 0 and system forward to test

Test : Once test has done it will forward for analysis

Analysis : Shows the result state

State => 1 : Under execution state

0 : Process successfully done state

Software Requirements

1. System interfaces: Windows Operating System

2. User interfaces: User interface using Jsp and Servlet

3. Hardware interfaces

Processor :- Intel R-Core i3 2.7 or above

Memory :- 4GB or above

Hard Disk :- 500 GB

4. Software interfaces:

Front End: Jdk 1.7.0, Net beans 7.4

IE 7.0/above

Back-End: Mysql 5.1.

VI. RESULTS AND DISCUSSION

For the system performance evaluation, calculate the matrices for accuracy. The system is executed on java 3-tier architecture framework with INTEL 2.8 GHz i7 processor and 16 GB RAM. In maiden experimentation system show the user multiple existing system accuracy with error rate. Figure 2 shows the accuracy using KDDCUP99 dataset which is taken from DARPA organization.

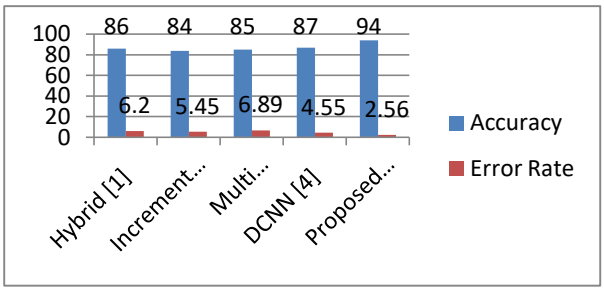

Fig. 2 : System Performance Measures proposed vs Existing approaches

In this work, system proposed a hyper-heuristic SVM optimization framework for big data cyber security problems called IDS. We formulated the SVM configuration process as a bi-objective optimization problem in which accuracy and model complexity are treated as two conflicting objectives. This bi-objective optimization problem can be solved using the proposed hyper-heuristic framework. The framework integrates the strengths of decomposition- and Pareto-based approaches to approximate the Pareto set of configurations. Our framework has been tested on two benchmark cyber security problem instances: Microsoft malware big data classification and anomaly intrusion detection. The experimental results demonstrate the effectiveness and potential of the proposed framework in achieving competitive, if not superior, results compared with other algorithms.

REFERENCES

[1]Soheily-Khah, Saeid, Pierre-François Marteau, and Nicolas Béchet. "Intrusion Detection in Network Systems Through Hybrid Supervised and Unsupervised Machine Learning Process: A Case Study on the ISCX Dataset." Data Intelligence and Security (ICDIS), 2018 1st International Conference on.IEEE, 2018.

[2] Alaei, Parisa, and Fakhroddin Noorbehbahani. "Incremental anomaly-based intrusion detection system using limited labeled data."Web Research (ICWR), 2017 3th International Conference on. IEEE, 2017.

[3] Falcón-Cardona, Jesús Guillermo, and Carlos A. Coello Coello. "A multi-objective evolutionary hyper-heuristic based on multiple indicator-based density estimators." Proceedings of the Genetic and Evolutionary Computation Conference.ACM, 2018.

[4] Rahul, Vigneswaran K., et al. "Evaluating Shallow and Deep Neural Networks for Network Intrusion Detection Systems in Cyber Security." 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT).IEEE, 2018.

[5] Gaied, Imen, Farah Jemili, and Ouajdi Korbaa. "Neuro-fuzzy and genetic-fuzzy based approaches in intrusion detection: Comparative study." Software, Telecommunications and Computer Networks (SoftCOM), 2017 25th International Conference on.IEEE, 2017.

[6] Potteti, Sumalatha, and NamitaParati. "Intrusion detection system using hybrid Fuzzy Genetic algorithm."Trends in Electronics and Informatics (ICEI), 2017 International Conference on.IEEE, 2017.

[7] Mukane, Rohit V., et al. "LabVIEW Based Implementation of Fuzzy Logic for Vibration Analysis to Identify Machinery Faults." 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA).IEEE, 2017.

[8] Behera, SantiKumari, et al. "Disease Classification and Grading of Orange Using Machine Learning and Fuzzy Logic." 2018 International Conference on Communication and Signal Processing (ICCSP).IEEE, 2018.

[9] Theresa, W. Gracy, and S. Sakthivel. "Fuzzy based intrusion detection for cluster based battlefield MANET." Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), 2017 IEEE International Conference on.IEEE, 2017.

[10] Alqahtani, Saeed M., and Robert John. "A comparative analysis of different classification techniques for cloud intrusion detection systems' alerts and fuzzy classifiers."Computing Conference, 2017.IEEE, 2017.

[11] Mohammed Hasan Ali, Bahaa Abbas Dawood AL Mohammed1, Madya Alyani Binti Ismail, Mohamad Fadli Zolkipli, A new intrusion detection system based on Fast Learning Network and Particle swarm optimization.

