European Journal of Molecular & Clinical Medicine
ISSN 2515-8260 Volume 07, Issue 11 , 2020

A Multi-Objective Hyper-Heuristic Improved Particle Swarm Optimization Based Configuration of SVM for Big Data Cyber Security

PVN RAJESWARI[1], MADDURI SUSMITHA[2], V SRINIVASA MOHAN KUMAR

[1]Associate Professor, Dept of CSE, Visvodaya Engineering College, Kavali, AP, India.
Email id: phdpvnr@gmail.com

[2] PG Scholar, Dept of CSE, Visvodaya Engineering College, Kavali, AP, India.
Email id: susmithareddy.madduri@gmail.com

[2] PG Scholar, Dept of CSE, Visvodaya Engineering College, Kavali, AP, India.
Email id: varigonda.srinu@gmail.com

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 hyperheuristic 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.

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

1. INTRODUCTION

Modern digital information era has created the space for high volume of data to be generated and stored by the advanced technologies and Internet of Things (IoT) [1]. This rapid growth of the Internet data has also exponentially increased the frequency of cyber-attacks. The cyber-attacks cause extensive damages to the networks and hence to tackle them he cyber security systems have been designed and installed. Cyber security techniques and processes are assigned with the

role of thwarting the illegal cyber-attacks to protect the computers and networks from the cyber damages [2]. They perform the major function of protecting the shared information for improving decision making; detecting the vulnerable attacks in applications; prevent unauthorized accessing of networks and secure the confidential network information [3]. Most of the larger companies have their own cyber security network while other organizations make use of such solutions from security organizations like Accenture, IBM, CISCO, etc. [4]. Recent cyber security solutions have inclined more towards the monitoring of network and Internet traffic to identify and avert the bad actions [5]. This is entirely different from the traditional cyber security solutions which focus only on the detection of bad signatures for unauthorized access. While the traditional systems were aimed at detecting the malware by scanning the incoming traffic against the malware signatures, they are relatively weaker with detecting only limited threats [6]. These traditional techniques including the intrusion detection, firewalls and anti-virus software have become ineffective in tackling the hackers as the attack strategies are highly destructive than the older versions [7]. In addition to this, the presence of big data has increased the critical condition as gigabytes of data are transferred between each node of the computer networks; making the hackers job of entering the networks very easier and cause severe damage without getting traced [8]. The big data problems are majorly due to the organizations providing access to their data networks allowing the partners and consumers to access all data and making it vulnerable to the cyber-attacks. Similarly, the big data has also increased the skills of hackers to evade the traditional security systems. Also, the big data has made it difficult to identify the attacks when initiated and the attack is only known after the damage is done to the hardware and software components [9].

2. 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 FakhroddinNoorbehbahani[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 hyperheuristic 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

3. PROPOSED SYSTEM

The proposed hyper-heuristic framework for configuration selection is shown in Figure 1. It has two levels: the high- level strategy and the low-level heuristics. 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. To address the bi-objective optimization problem, we propose a population-based hyper-heuristic framework that operates on a population of solutions and uses an archive to save the non-dominated solutions. The proposed framework combines the strengths of decomposition- and Pareto (dominance) - based approaches to effectively approximate the Pareto set of SVM configurations. Our idea is to combine the diversity ability of the decomposition approach with theconvergence power of the dominance approach. The decomposition approach operates on the population of solutions, whereas the dominance approach uses the archive. The hyper heuristic framework generates a new population of solutions using the old population, the archive, or both the old population and the archive. This allows the search to achieve a proper balance between convergence and diversity. It should be noted that seeking good convergence involves minimizing the distances between the solutions and PF, whereas seeking high diversity involves maximizing the distribution of the solutions along PF. The main components of the proposed hyper-heuristic framework are discussed in the following subsections.

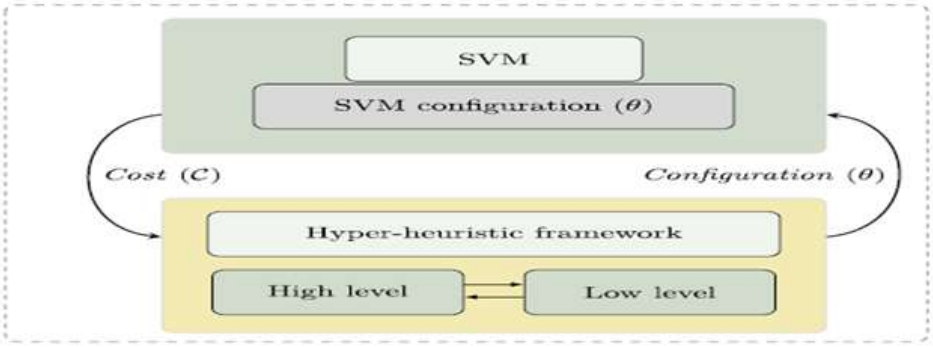

Figure 1 : Proposed system

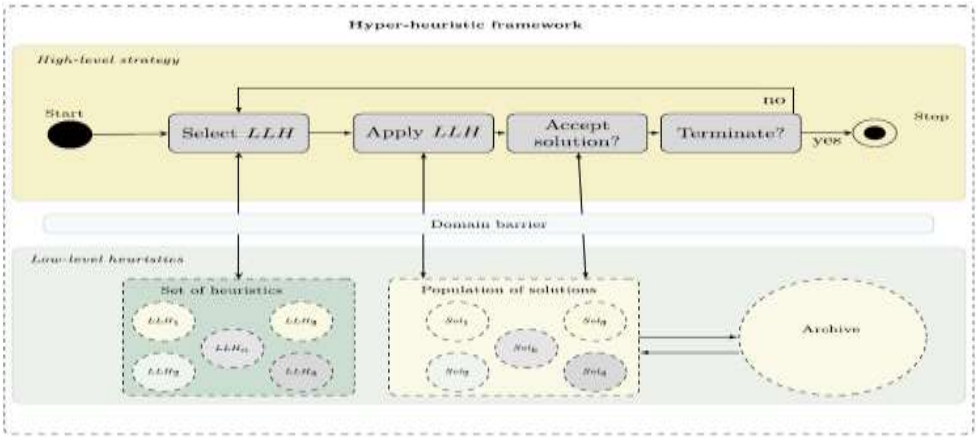

Figure 2 : Hyper Heuristic Framework

3.1 Working Modules

The project is carried out based on the following modules listed:

1. Approved Users

In this system users are not allowed to access resources simply. User need verify their information's with admin. Admin are the authorized and trustworthy to the network. User need to send the request to administrator that they are interested to add the community. Admin views the user request and respond with the pass code to access users account through trusted sources like SSL (Gmail).

2. Security Steps and Upload

This is where the proposed algorithm is going to be effective. The admin can be upload the files with proposed classification algorithm and cryptography in order to classify and upload the encrypted details to network with its tag in the mark of understand to user about the resource.

3. Resource Access

The permissions to access the resource can be sent by users to admin. The requests have been updated by admin with the response to access the resource. Users can decrypt the resource and access the details. The important part is access the resource with the decryption. The passkey to access the details are limited. If the limit of wrong attempts over the threshold value means pass key expires.

4. Graphical Representation

This is graphical notation of the data given by the system. This phase of implementation will shows the effectiveness of the proposed system through pictorially in the order to better understand of proposed system.

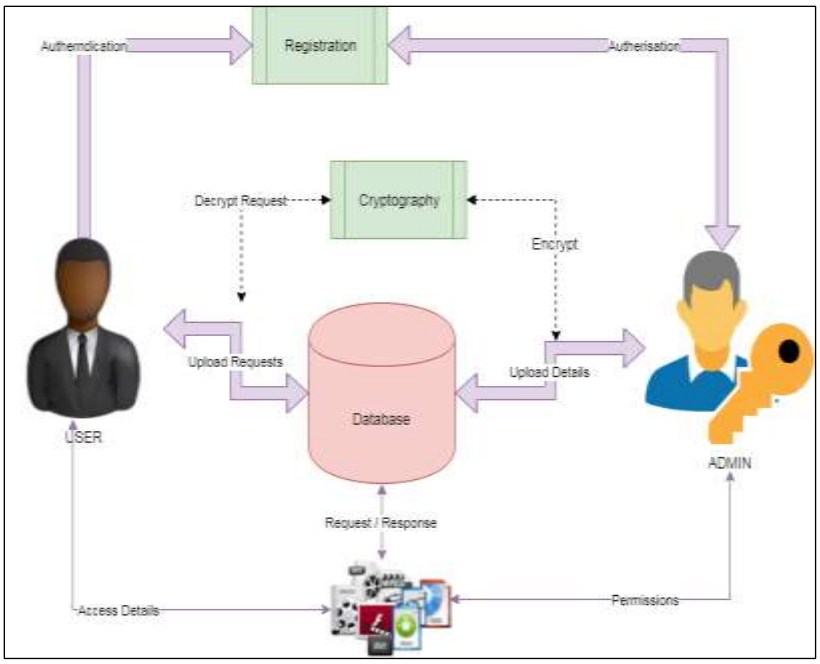

Figure 3:System Architecture

4. RESULTS AND ANALYSIS

Figure 4: Registration Form

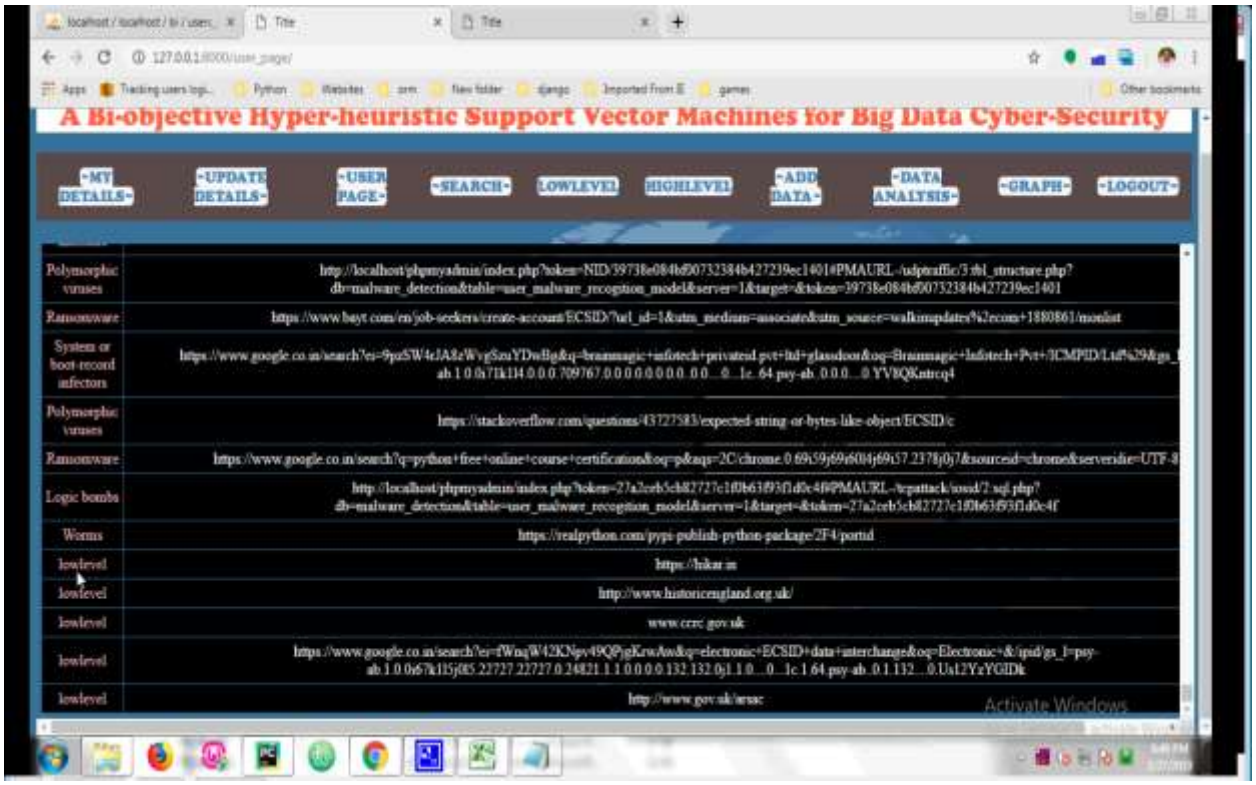

Figure 5:Dataset Values

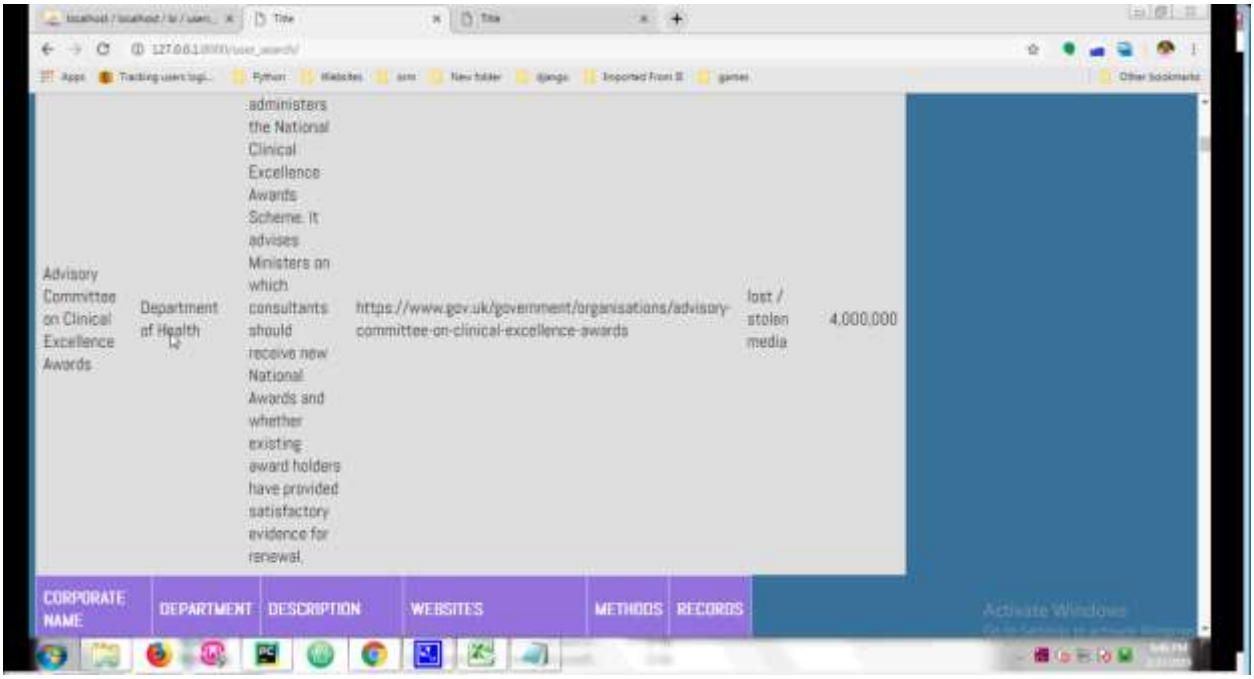

Figure 6:User Search Data

5.CONCLSION

In this work, we proposed a hyper-heuristic SVM optimization framework for big data cyber security problems. 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.

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 FakhroddinNoorbehbahani. "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. CoelloCoello. "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 OuajdiKorbaa. "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.

Author's Profile:

PVN Rajeswari, has received her B.Tech in CSE from Andhra University and M.Tech degree in CSE from Andhra University in 2004 and Allahabad University in 2006 respectively. Presently she is pursuing PhD from Andhra University. She is dedicated to teaching field from the last 14 years. She has guided 22 P.G and 41 U.G students. Her research areas included Artificial Intelligence and Data Mining. At present she is working as Associate Professor in Visvodaya Engineering College, Kavali, Andhra Pradesh, India.

Madduri Susmitha has received her B.Tech Degree in Computer Science & Engineering from RSR Engineering College, affiliated to JNTUA in 2018 and Pursuing M.Tech degree in Computer Science & Engineering in Visvodaya Engineering College, Kavali, affiliated to JNTUA, Ananthapur in 2021.

V Srinivasa Mohan Kumar received his B. Tech Degree in Information Technology from PBR Viswodaya Institute of Technology & Science, affiliated to JNTUA in 2009 and Pursuing M. Tech degree in Computer Science & Engineering in Visvodaya Engineering College, affiliated to JNTUA in 2021.