



SWARM OPTIMIZATION BASED INTRUSION DETECTION SYSTEM

¹ Vivek M, ² Dharani N S, ³ Guruabinaya K, ⁴ Harish Kumar R, ⁵ Jeevanandham B

¹ Assistant Professor, Department of CSE, Jansons Institute of Technology, Coimbatore, India

²⁻⁵ UG Students, Department of CSE, Jansons Institute of Technology, Coimbatore, India

Abstract:

Network security plays a critical role in our lives based on the threats and attacks to which we are exposed and that increase daily; these attacks result in the need to develop different protection methods and techniques. Network intrusion detection systems are a way to detect several malicious network attacks. Many existing systems have focused on developing intrusion detection based on machine learning (ML) approaches to detect variants of attacks. ML approaches can automatically discover the essential variances between normal and abnormal data by analysing the features of a large dataset. Indeed, many features are extracted without discrimination, Intrusion detection is needed as another level of security to protect Wireless Network systems. Signature-based analysis is a technique that was proposed earlier. It was widely used in intrusion detection community to protect a system by using a combination of an alarm that sounds whenever the security sites has been compromised, with site security officer (SSO). SSO can respond to the alarm and take the appropriate action, for instance by ousting the intruder, calling the proper external authorities, and so on. Many complete systems have been constructed and operated on live computer systems. However, despite over 25 years of research, the topic is still popular, partly due to the rapid development of information processing

increasing the computational complexity. By applying a feature selection method, a subset of features is selected from the whole feature set with the aim of improving the performance of ML based detection methods. The SALP swarm algorithm (SSA) is a nature-based optimization algorithm that has demonstrated efficiency in minimizing processing challenges to perform optimization for feature selection problems. The proposed system investigates the impact of the SSA on improving ML-based network anomaly detection using naïve bayes classifier.

I. INTRODUCTION

systems and the consequent discovery of new vulnerabilities, but also due to fundamental difficulties in achieving an accurate declaration of an intrusion. Intrusion systems are noted for high false alarm rates and considerable research effort is still concentrated on finding effective intrusion, non-intrusion discriminates.

However, there are many problems with present intrusion detection systems. One of the major problems is the high number of false positives alarm. False alarms are high and the potential to recognize an attack is uncertain.

KEYWORDS

SSA, Intrusion, PSO, Naive Bayes, NIDS, FS.

RELATED WORKS

Network intrusion detection system (NIDS) is a commonly used tool to detect attacks and protect networks, while one of its general limitations is the false positive issue. On the basis of our comparative experiments and analysis for the characteristics of the particle swarm optimization (PSO) and Xgboost, this paper proposes the PSO-Xgboost model given its overall higher classification accuracy than other alternative models such like Xgboost, Random Forest, Bagging and Adaboost. Firstly, a classification model based on Xgboost is constructed, and then PSO is used to adaptively search for the optimal structure of Xgboost. The benchmark NSL-KDD dataset is used to evaluate the proposed model. Our experimental results demonstrate that PSO-Xgboost model outperforms other comparative models in precision, recall, macro-average (macro) and mean average precision (mAP), especially when identifying minority groups of attacks like U2R and R2L. This work also provides experimental arguments for the application of swarm intelligence in NIDS.

An Approach for the Application of a Dynamic Multi-Class Classifier for Network Intrusion Detection Systems

Currently, the use of machine learning models for developing intrusion detection systems is a technology trend which improvement has been proven. These intelligent systems are trained with labelled datasets, including different types of attacks and the normal behaviour of the network. Most of the studies use a unique machine learning model, identifying anomalies related to possible attacks. In other cases, machine learning algorithms are used to identify certain type of attacks. However, recent studies show that certain

models are more accurate identifying certain classes of attacks than others. Thus, this study tries to identify which model fits better with each kind of attack in order to define a set of reasoner modules. In addition, this research work proposes to organize these modules to feed a selection system, that is, a dynamic classifier. Finally, the study shows that when using the proposed dynamic classifier model, the detection range increases, improving the detection by each individual model in terms of accuracy.

Building an Efficient Intrusion Detection System Based on Feature Selection and Ensemble Classifier

Intrusion detection system (IDS) is one of extensively used techniques in a network topology to safeguard the integrity and availability of sensitive assets in the protected systems. Although many supervised and unsupervised learning approaches from the field of machine learning have been used to increase the efficacy of IDSs, it is still a problem for existing intrusion detection algorithms to achieve good performance. First, lots of redundant and irrelevant data in high-dimensional datasets interfere with the classification process of an IDS. Second, an individual classifier may not perform well in the detection of each type of attacks. Third, many models are built for stale datasets, making them less adaptable for novel attacks. Thus, we propose a new intrusion detection framework in this paper, and this framework is based on the feature selection and ensemble learning techniques. In the first step, a heuristic algorithm called CFS-BA is proposed for dimensionality reduction, which selects the optimal subset based on the correlation between features. Then, we introduce an ensemble approach that combines C4.5, Random Forest

(RF), and Forest by Penalizing Attributes (Forest PA) algorithms. Finally, voting technique is used to combine the probability distributions of the base learners for attack recognition. The experimental results, using NSL-KDD, AWID, and CIC-IDS2017 datasets, reveal that the proposed CFS-BA-Ensemble method is able to exhibit better performance than other related and state of the art approaches under several metrics.

New Hybrid Method for Attack Detection Using Combination of Evolutionary Algorithms, SVM, and ANN

Intrusion detection systems (IDS) have been playing an important role for providing security of computer networks. They detect different types of attacks and malicious software usage, which sometimes cannot be identified by firewalls. Based on machine learning algorithms, many IDS have been extended to classify network traffic as normal or abnormal. This paper describes a new hybrid intrusion detection method with two phases - a feature selection phase and an attack detection phase. In the feature selection phase, a wrapper technique, namely MGA-SVM, is used. This technique combines features of support vector machine (SVM) and the genetic algorithm with multi-parent crossover and multi-parent mutation (MGA). In the attack detection phase, an artificial neural network (ANN) is used to detect attacks. For improving its performance, a combination of a hybrid gravitational search (HGS) and a particle swarm optimization (PSO) is used to train the classifier. The proposed hybrid method is thus called MGA-SVM-HGS-PSO-ANN. It's performance is compared with other popular techniques such as Chi-SVM, ANN based on gradient descent (GD-ANN) and decision tree (DT), ANN based on genetic algorithm (GA-

ANN), ANN based on combining gravitational search (GS) and PSO (GSPSO-ANN), ANN based on PSO (PSO-ANN), and ANN based on GS (GS-ANN). Using the NSL-KDD dataset as a standard benchmark for attack detection evaluation, the obtained test results show that the proposed MGA-SVM-HGS-PSO-ANN method can attain a maximum detection accuracy of 99.3%, dimension reduction of NSL-KDD from 42 to 4 features, and needs only 3 seconds as maximum training time.

Research on Network Intrusion Detection Based on Incremental Extreme Learning Machine and Adaptive Principal Component Analysis

Recently, network attacks launched by malicious attackers have seriously affected modern life and enterprise production, and these network attack samples have the characteristic of type imbalance, which undoubtedly increases the difficulty of intrusion detection. In response to this problem, it would naturally be very meaningful to design an intrusion detection system (IDS) to effectively and quickly identify and detect malicious behaviours. In our work, we have proposed a method for an IDS-combined incremental extreme learning machine (I-ELM) with an adaptive principal component (A-PCA). In this method, the relevant features of network traffic are adaptively selected, where the best detection accuracy can then be obtained by I-ELM. We have used the NSL-KDD standard dataset and UNSW-NB15 standard dataset to evaluate the performance of our proposed method. Through analysis of the experimental results, we can see that our proposed method has better computation capacity, stronger generalization ability, and higher accuracy.

TSE-IDS: A Two-Stage Classifier Ensemble for Intelligent Anomaly-Based Intrusion Detection System

Intrusion detection systems (IDSs) play a pivotal role in computer security by discovering and repealing malicious activities in computer networks. Anomaly-based IDS, in particular, rely on classification models trained using historical data to discover such malicious activities. In this paper, an improved IDS based on hybrid feature selection and two-level classifier ensembles are proposed. A hybrid feature selection technique comprising three methods, i.e., particle swarm optimization, ant colony algorithm, and genetic algorithm, is utilized to reduce the feature size of the training datasets (NSL-KDD and UNSW-NB15 are considered in this paper). Features are selected based on the classification performance of a reduced error pruning tree (REPT) classifier. Then, a two-level classifier ensemble based on two meta learners, i.e., rotation forest and bagging, is proposed. On the NSL-KDD dataset, the proposed classifier shows 85.8% accuracy, 86.8% sensitivity, and 88.0% detection rate, which remarkably outperform other classification techniques recently proposed in the literature. The results regarding the UNSW-NB15 dataset also improve the ones achieved by several state-of-the-art techniques. Finally, to verify the results, a two-step statistical significance test is conducted. This is not usually considered by the IDS research thus far and, therefore, adds value to the experimental results achieved by the proposed classifier.

Improved SALP swarm algorithm based on particle swarm optimization for feature selection

Feature selection (FS) is a machine learning process commonly used to reduce the high

dimensionality problems of datasets. This task permits to extract the most representative information of high sized pools of data, reducing the computational effort in other tasks as classification. This article presents a hybrid optimization method for the FS problem; it combines the slap swarm algorithm (SSA) with the particle swarm optimization. The hybridization between both approaches creates an algorithm called SSAPSO, in which the efficacy of the exploration and the exploitation steps is improved. To verify the performance of the proposed algorithm, it is tested over two experimental series, in the first one, it is compared with other similar approaches using benchmark functions. Meanwhile, in the second set of experiments, the SSAPSO is used to determine the best set of features using different UCI datasets. Where the redundant or the confusing features are removed from the original dataset while keeping or yielding a better accuracy. The experimental results provide the evidence of the enhancement in the SSAPSO regarding the performance and the accuracy without affecting the computational effort.

A dynamic locality multi-objective SALP swarm algorithm for feature selection

Developing intelligent analytical tools requires pre-processing data and finding relevant features that best reinforce the performance of the predictive algorithms. Feature selection plays a significant role in maximizing the accuracy of machine learning algorithms since the presence of redundant and irrelevant attributes deteriorates the performance of the learning process and increases its complexity. Feature selection is a combinatorial optimization problem that can be formulated as a multi-objective optimization problem with the purpose of maximizing the classification

performance and minimizing the number of irrelevant features. It is considered an NP hard optimization problem since having a number of (n) features produces a large search space of size (2^n) of different permutations of features. An eminent type of optimizer for tackling such an exhausting search process is evolutionary, which mimic evolutionary processes in nature to solve problems in computers. SALP Swarm Algorithm (SSA) is a well-established metaheuristic that was inspired by the foraging behaviour of salps in deep oceans and has proved to be beneficial in estimating global optima for optimization problems. The objective of this article is to promote and boost the performance of the multi-objective SSA for feature selection. Therefore, it proposes an enhanced multi-objective SSA algorithm (MODSSA-lbest) that adopts two essential components: the dynamic time-varying strategy and local fittest solutions. These components assist the SSA algorithm in balancing exploration and exploitation. Thus, it converges faster while avoiding locally optimal solutions. The proposed approach (MODSSA-lbest) is tested on 13 benchmark datasets and compared with the well-regarded Multi-Objective Evolutionary Algorithms (MOEAs). The results show that the MODSSA-lbest achieves significantly promising results versus its counterpart algorithms

SALP Swarm Algorithm: A bio-inspired optimizer for engineering design problems

This work proposes two novel optimization algorithms called SALP Swarm Algorithm (SSA) and Multiobjective SALP Swarm Algorithm (MSSA) for solving optimization problems with single and multiple objectives. The main inspiration of SSA and MSSA is the swarming behaviour of SALPS when navigating and foraging in oceans. These two algorithms are tested on

several mathematical optimization functions to observe and confirm their effective behaviours in finding the optimal solutions for optimization problems. The results on the mathematical functions show that the SSA algorithm is able to improve the initial random solutions effectively and converge towards the optimum. The results of MSSA show that this algorithm can approximate Pareto optimal solutions with high convergence and coverage. The paper also considers solving several challenging and computationally expensive engineering design problems (e.g., air foil design and marine propeller design) using SSA and MSSA. The results of the real case studies demonstrate the merits of the algorithms proposed in solving real-world problems with difficult and unknown search spaces.

UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)

One of the major research challenges in this field is the unavailability of a comprehensive network-based data set which can reflect modern network traffic scenarios, vast varieties of low footprint intrusions and depth structured information about the network traffic. Evaluating network intrusion detection systems research efforts, KDD98, KDDCUP99 and NSLKDD benchmark data sets were generated a decade ago. However, numerous current studies showed that for the current network threat environment, these data sets do not inclusively reflect network traffic and modern low footprint attacks. Countering the unavailability of network benchmark data set challenges, this paper examines a UNSW-NB15 data set creation. This data set has a hybrid of the real modern normal and the contemporary synthesized attack activities of the network traffic. Existing and novel methods are

utilised to generate the features of the UNSWNB15 data set. This data set is available for research purposes and can be accessed from the link.

EXISTING SYSTEM

Intrusion detection systems (IDS) have been playing an important role for providing security of computer networks. They detect different types of attacks and malicious software usage, which sometimes cannot be identified by firewalls. Based on machine learning algorithms, many IDS have been extended to classify network traffic as normal or abnormal. This paper describes a new hybrid intrusion detection method with two phases - a feature selection phase and an attack detection phase. In the feature selection phase, a wrapper technique, namely MGA-SVM, is used. This technique combines features of support vector machine (SVM) and the genetic algorithm with multi-parent crossover and multi-parent mutation (MGA). In the attack detection phase, an artificial neural network (ANN) is used to detect attacks. For improving its performance, a combination of a hybrid gravitational search (HGS) and a particle swarm optimization (PSO) is used to train the classifier. The proposed hybrid method is thus called MGA-SVM-HGS-PSO-ANN. Its performance is compared with other popular techniques such as Chi-SVM, ANN based on gradient descent (GD-ANN) and decision tree (DT), ANN based on genetic algorithm (GA-ANN), ANN based on combining gravitational search (GS) and PSO (GSPSO-ANN), ANN based on PSO (PSO-ANN), and ANN based on GS (GS-ANN). Using the NSL-KDD dataset as a standard benchmark for attack detection evaluation, the obtained test results show that the proposed MGA-SVM-HGS-PSO-ANN method can attain a

maximum detection accuracy of 99.3%, dimension reduction of NSL-KDD from 42 to 4 features, and needs only 3 seconds as maximum training time.

PROPOSED SYSTEM

Computer network protection plays an essential role regarding internal and external threats; there are various gaps that attackers can exploit to break into and access these networks to manipulate or steal sensitive information and cause considerable damage. One of the ways to isolate and protect an environment from outside attacks is to use firewalls and traditional rule-based security protection techniques. Furthermore, to increase the security protection level, another system is needed to support traditional security techniques in protecting from different types of malicious attacks. Moreover, advanced and sophisticated technologies are needed to examine and analyse enormous amounts of data from network infrastructure transactions. Robust network intrusion detection systems (NIDSs) have been produced, which play a crucial role in ensuring network security and require analysis of the generated complex data. A NIDS protects the computer system or administrators when dealing with various threats and attacks. Accordingly, when using a NIDS and its protection ability, it must be up to date since there are many gaps to be determined, addressed, and filled in for the network detection model. NIDSs work by analysing and extracting the abnormal behaviour to be detected. After these suspicious behaviours are detected, alerts are sent to notify parties of the abnormal behaviour that must be considered before sensitive information is accessed and reached or these data manipulated and leaked. Two methods are used by a NIDS. First, anomaly detection is

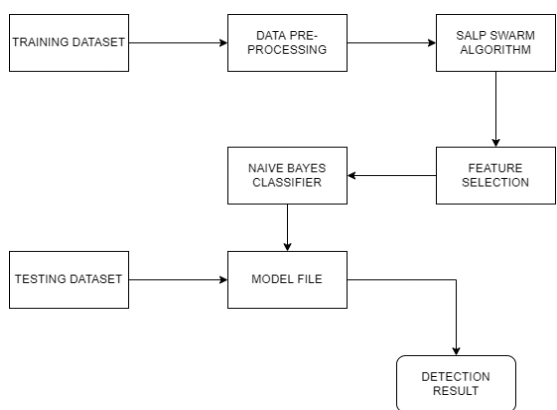
used to detect unknown attacks that pass through the network or the system, which are new attacks that have not been addressed in the past. Second, signature detection, also known as misuse detection, detects abnormal behaviour through prior knowledge, which must be defined with a rule or pattern as knowledge-based detection. Different techniques are applied in anomaly detection, such as machine learning (ML) techniques, to achieve a high detection rate and accurate results. ML-based detection plays an essential role in building an effective model with different algorithms and approaches to analyse big data consisting of traffic flows in networks that prompt intrusions. Moreover, one way to develop effective ML-based anomaly detection and reach a desired goal is to use feature selection (FS) techniques. The importance of FS is that it improves the model performance and obtains an accurate result. Therefore, producing an intelligent analytical tool requires pre-processing data as one of the essential steps in finding relevant features that best reinforce the predictive algorithm performance. FS plays a meaningful role in maximizing the performance of a ML model considering redundant and irrelevant attributes that degenerate the learning process performance and increase its complexity. FS is an optimization problem that can be expressed as a multiobjective optimization problem and is an NP-hard optimization problem since it has several (n) features, producing an ample search space of size (2^n) of various permutations of the features. In particular, numerous search methods can be utilized to detect the optimal subset of features. The first is to apply the greedy search method by producing and assessing all of the features, making this scheme time consuming. The second is to apply a random search method by exploring the

domain randomly, which has some drawbacks and limitations. There is a chance of stagnation issues, including a very high time complexity. One way to address the drawbacks and gaps of previous FS methods that have been proposed by researchers is to use meta-heuristic paradigms. Meta-heuristic methods are global optimization approaches that accompany all physical, biological, and animal activities. They can explore the search space globally and locally when applied to FS problems. One meta-heuristic algorithm is the swarm intelligence (SI) technique. This technique comes from the concept of the intelligence of swarms, herds, schools, or flocks of creatures in nature. SI algorithms have been widely used to address different optimization problems and reach suboptimal or optimal solutions. In particular, we investigate the influence of the SALP swarm algorithm (SSA), as one of the newest SI algorithms, to determine its effectiveness and its capability to address FS problems as one of the essential pre-processing steps to enhance the ML-based anomaly detection model. Our contributions to the network security field are as follows. We have proposed network anomaly detection based on three phases. (i) In the first phase of efficient network anomaly detection, to achieved a high classification accuracy and increase the detection rate while reducing the false alarm rate without using excessive computational resources. To accomplish these goals, we have adopted the SSA-FS method to obtain the most relevant and accurate feature representation. According to our limited knowledge, no research has focused on the impact of SSA as an FS method on the network anomaly detection problem. (ii) The second phase used classifier algorithms named Naïve Bayes. Based on FS, we have monitored and test the effectiveness of

the SSA and its impact on these two different algorithms. (iii) The last phase was to conduct an extensive experimental evaluation test of the proposed method using two datasets: UNSW-NB15 and NSL-KDD. Additionally, the method was compared to several of the newest state-of-the-art techniques.

MODULE DESCRIPTION

In order to classify the network intrusion quickly the number of features has to be reduced so SALP swarm algorithm is implemented This work proposes two novel optimization algorithms called SALP Swarm Algorithm (SSA) and Multiobjective SALP Swarm Algorithm (MSSA) for solving optimization problems with single and multiple objectives. The main inspiration of SSA and MSSA is the swarming behaviour of SALPS when navigating and foraging in oceans. These two algorithms are tested on several mathematical optimization functions to observe and confirm their effective behaviours in finding the optimal solutions for optimization problems. The results on the mathematical functions show that the SSA algorithm is able to improve the initial random solutions effectively and converge towards the optimum.



Block Diagram

Needless to say, the mathematical model for simulating SALP chains cannot be directly employed to solve optimization problems. In other

words, there is a need to tweak the model a little bit to make it applicable to optimization problems. The ultimate goal of a single-objective optimizer is to determine the global optimum. In the SSA swarm model, follower SALPS follow the leading SALP. The leading SALP also moves towards the food source. If the food source be replaced by the global optimum, therefore, the SALP chain automatically moves towards it. However, the problem is that the global optimum of optimization problems is unknown. In this case, it is assumed that the best solution obtained so-far is the global optimum and assumed as the food source to be chased by the SALP chain



Data Flow Diagram

SALP SWARM ALGORITHM

The importance and effectiveness of SI have been proven in many applications and research. In the global optimization framework, SI addresses the optimization problem, which means improving and identifying the suboptimal solutions for a problem from among a group of alternative solutions, among which is the optimal solution. One of the newest SI algorithms is the SSA, which is considered new and is being tested and analysed to determine its efficiency and ability to solve

REFERENCES

- [1] H. Jiang, Z. He, G. Ye, and H. Zhang, "Network intrusion detection based on PSO-XGBOOST model," *IEEE Access*, vol. 8, pp. 58 392–58 401, 2020.
- [2] A. Husain, A. Salem, C. Jim, and G. Dimitoglou, "Development of an efficient network intrusion detection model using extreme gradient boosting (XGBOOST) on the unsw-nb15 dataset," in *2019 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pp. 1–7, 2019.
- [3] Y. Zhou, G. Cheng, S. Jiang, and M. Dai, "Building an efficient intrusion detection system based on feature selection and ensemble classifier," *Computer Networks*, vol. 174, p. 107247, 2020.
- [4] S. Hosseini and B. M. H. Zade, "New hybrid method for attack detection using combination of evolutionary algorithms, SVM, and ANN," *Computer Networks*, vol. 173, p. 107168, 2020.
- [5] J. Gao, S. Chai, B. Zhang, and Y. Xia, "Research on network intrusion detection based on incremental extreme learning machine and adaptive principal component analysis," *Energies*, vol. 12, no. 7, 2019.
- [6] B. A. Tama, M. Comuzzi, and K.-H. Rhee, "TSE-ids: A two-stage classifier ensemble for intelligent anomaly-based intrusion detection system," *IEEE Access*, vol. 7, pp. 94 497–94 507, 2019.
- [7] R. A. Ibrahim, A. A. Ewees, D. Oliva, M. Abd Elaziz, and S. Lu, "Improved SALP swarm algorithm based on particle swarm optimization for feature selection," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 8, pp. 3155–3169, 2019.
- [8] I. Aljarah, M. Habib, H. Faris, N. Al-Madi, A. A. Haidari, M. Mafarja, M. A. Elaziz, and S. Mir Jalili, "A dynamic locality multi-objective SALP swarm algorithm for feature selection," *Computers & Industrial Engineering*, vol. 147, p. 106628, 2020.
- [9] S. Mirjalili, A. H. Gandomi, S. Z. Mirjalili, S. Saremi, H. Faris, and S. M. Mirjalili, "SALP swarm algorithm: A bio-inspired optimizer for engineering design problems," *Advances in Engineering Software*, vol. 114, pp. 163– 191, 2017.
- [10] N. Moustafa and J. Slay, "Unsw-nb15: a comprehensive data set for network intrusion detection systems (unsw-nb15 network data set)," in *2015 Military Communications and Information Systems Conference (MILCIS)*, pp. 1–6, 2015.