المجلة العراقية للبحوث الانسانية والاجتماعية والعلمية
Iraqi Journal of Humanitarian, Social and Scientific Research
Print ISSN 2710-0952 Electronic ISSN 2790-1254
العدد 13 آيار 2024
No.13A May 2024

Detection System Using Dragonfly in Feature Selection and the Herd Optimization Algorithm (HOA) Combination

Mohammed sami Ibrahim mohammed Alghaloom
Department of Computer Science, Urmia University, Iran
Corresponding author :hamdsami40@gmail.com

Regardless of numerous endeavors in battling Web of Things network assaults in the previous year's such goes after, for example, DDoS are as yet a serious danger to the security of the internet particularly in Web of Things organizations. As an ever-increasing number of fundamental administrations rely upon the Web as a feature of their interchanges foundation the results of disavowal of-administration assaults can destroy. Interruption identification framework is one of the accessible innovations to safeguard organizations and Web of Things gadgets. This framework screens the organization and identifies the interruption of malevolent hubs with the point of detached (listening in) or positive assault to upset the organization. Meanwhile giving a reasonable assault recognition framework in Web of Things networks is straightforwardly connected with the choice of highlights and proficiency of the pre-owned characterization strategy. In such manner in this exploration an interruption recognition model in view of component extraction and determination and order of brain networks has been proposed. In the proposed technique IoT-BoT datasets are gathered first and afterward the information are pre-handled to eliminate commotion missing information and expansions and in the continuation of the component determination process in view of the dragonfly improvement calculation the elements of the dataset diminish. From that point onward, by utilizing the MLP brain network characterization strategy, which its weight coefficients and boundaries are advanced by the multitude calculation, an interruption identification model can be assembled. To assess the proposed strategy, it is reproduced alongside a correlation technique involving MATLAB programming language and executed on normal boundaries in this field. The assessment results affirm that the proposed strategy has given fundamentally improved results than the correlation technique. These assessments, which have been performed in light of changes in the size of the preparation set of the dataset, show that the proposed technique has improved separately Review Exactness Accuracy and F-Measure measurements by 15%, 14%, 12%, and 14%.

Keywords: Internet of things, Intrusion detection system, Neural network classification, Dragonfly algorithm feature selection, Horse herd algorithm.

المجلة العراقية للبحوث الانسانية والاجتماعية والعلمية العدد 13 آيار 2024
Iraqi Journal of Humanitarian, Social and Scientific Research No.13A May 2024
Print ISSN 2710-0952 Electronic ISSN 2790-1254

نظام الكشف باستخدام Dragonfly في اختيار الميزات ودمج خوارزمية تحسين القطيع HOA

محمد سامي ابراهيم محمد الغلوم

المستخلص

على الرغم من الجهود العديدة المبذولة لمكافحة هجمات شبكة ويب الأشياء في العام الماضي، فإن مثل هذه الهجمات، على سبيل المثال، لا تزال هجمات DDoS تمثل خطرًا كبيرًا على أمن الإنترنت وخاصة في مؤسسات ويب الأشياء. نظرًا لأن عددًا متزايدًا من الإدارات الأساسية تعتمد على الويب كجزء من أساس التبادلات الخاصة بها، فإن نتائج هجمات التنصل من الإدارة يمكن أن تدمر. يعد إطار تعريف الانقطاع أحد الابتكارات التي يمكن الوصول إليها لحماية المؤسسات وأدوات .Web of Things يقوم هذا الإطار بفحص المنظمة وتحديد انقطاع المحاور الخبيثة بدرجة الانفصال (الاستماع) أو الاعتداء الإيجابي لإزعاج المنظمة. وفي الوقت نفسه، يرتبط توفير إطار عمل معقول للتعرف على الاعتداءات في شبكات Web of Things بشكل مباشر باختيار النقاط البارزة وكفاءة استراتيجية التوصيف المملوكة مسبقًا. بهذه الطريقة، في هذا الاستكشاف، تم اقتراح نموذج التعرف على الانقطاع في ضوء استخراج المكونات وتحديد وترتيب شبكات الدماغ. في التقنية المقترحة، يتم جمع مجموعات بيانات IoT-BoT أولاً وبعد ذلك تتم معالجة المعلومات مسبقًا للتخلص من الضجة المفقودة في المعلومات والتوسعات وفي استمرار عملية تحديد المكون في ضوء حساب تحسين اليعسوب، تتضاءل عناصر مجموعة البيانات. من تلك النقطة فصاعدًا، من خلال استخدام استراتيجية توصيف شبكة الدماغMLP ، والتي يتم تطوير معاملات الوزن والحدود الخاصة بها عن طريق الحساب المتعدد، يمكن تجميع نموذج تحديد الانقطاع. ولتقييم الاستراتيجية المقترحة، تم إعادة إنتاجها إلى جانب تقنية الارتباط التي تتضمن لغة برمجة MATLAB وتنفيذها على الحدود الطبيعية في هذا المجال. وتؤكد نتائج التقييم أن الإستراتيجية المقترحة أعطت نتائج أفضل بشكل جوهري من تقنية الارتباط. تظهر هذه التقييمات، التي تم إجراؤها في ضوء التغييرات في حجم مجموعة إعداد مجموعة البيانات، أن التقنية المقترحة قد تحسنت بشكل منفصل دقة مراجعة الدقة وقياسات F بنسبة 15% و14% و12% و14.%

الكلمات المفتاحية: إنترنت الأشياء، نظام كشف التسلل، تصنيف الشبكات العصبية، اختيار ميزة خوارزمية اليعسوب، خوارزمية قطيع الخيول.

1.1 Introduction

Today, many pieces of human life have been impacted by the Web of Things (IoT). This development permits various clients to store, backing and move a lot of delicate and individual data anyplace and whenever [1]. By accomplishing a conclusive objective in the Trap of Things, all individuals and things will reserve the option to trade data and, if important, cycle data as per predefined plans. In such an association, not exclusively will data be accessible across clients, yet every apparatus office related with this data will be open all over the place, and the significance of the public web will be plainly apparent. . Meanwhile, expanding clients' confidence in the Trap of Things and its apparatuses is a significant issue to accomplish promising outcomes. The European Commission Organization Security Philosophy and Confirmation Act is likewise obviously attempting to increment trust later in the utilization of the Trap of Things. Meanwhile, the fundamental inquiries, for example, apparatus assurance, security and interoperability ought to likewise be considered [1].

As Web of Things contraptions have created and high level, their unfortunate security has made them truly defenseless against a wide assortment of assaults, including DDOS assaults. Advanced assaults on the Web of Things disturb admittance to instruments or upset the whole association by causing signal tenacity or depleting resources' center point batteries. With the expansion in assaults completed by fraudsters during the trading of data over the web, IoT apparatuses face security issues. In this survey, a cross breed approach utilizing meta-heuristic improvement computation and brain network regularization is proposed to oversee such assaults. [2].

In this part, at the outset, a short depiction of the essential point is given, in the going with test questions, the objectives and motivations of this examination are determined, lastly, an overall clarification of the improvement of the test and it is given. The overall format, at last, the segments and the plan presented in the investigation are portrayed.

The Internet of Things is a network of devices or physical objects such as embedded electronic devices, sensors, wireless connections and software that enable the collection and exchange of data. The development and growth of the Internet of Things for data exchange has made human life significantly easier. But cyber security is a big challenge in this field which affects further development and applications [3]. Poor security and a large number of IoT devices that are vulnerable to hacking may suffer from a distributed denial of service attack. [4].

In general, an attack refers to a process where an attacker enters the system or the system server and sends malicious packets to the user's system in order to steal, change or damage any confidential or important information. Attack also refers to the illegal sending of network packets through the network. Intrusions may occur as a result of vulnerabilities in the system, such as user abuse, system misconfiguration, or application defects on the server or system. In a global network, many online services and many large servers are running in this system. Such networks become more attractive to attackers and hence the network needs intelligent intrusion detection systems (IDS) to defend these systems [5].

A generative discovery framework is usually built using data mining methods because they can detect outages well. However, implementing and creating such frameworks can be inherently confusing. The inherent complexities of the framework can be separated into a set of undeniable issues with respect to the value of accuracy and convenience boundaries, in most cases those procedures that depend

on feature recognition in contrast to past recognition strategies that depend on manual signs. have shown a higher level of false positives. As a result, it is difficult for these procedures to handle information and detect online interruptions. The interrupt detection framework works in different steps such as information classification, salient selection (FS) and sequence pre-management. [6].

In this research, a comprehensive solution for identifying and reducing cyber attacks based on neural networks is presented, which combines the Horse Herd Optimization Algorithm (HOA) [8] to improve the accuracy of the MLP neural network as a classifier along with the Dragonfly Algorithm (DA). [9] as a feature selector.

Among machine learning techniques, MLP is one of the most widely used methods for intrusion detection systems. Artificial neural networks have many advantages including the capacity to adapt to fuzzy and basic models, the framework has a high adaptability that can learn the complex relationships within the information index without raising suspicion about the specific idea of each relationship. In addition, having the possibility of self-organization and the ability to learn makes these networks especially effective and efficient in predicting and classifying Internet traffic in high-dimensional datasets [10]. In recent studies, meta-heuristic algorithms have been used to improve and reduce classification errors in multilayer perceptrons of artificial neural networks. In most of these studies, reducing the MLP output error is considered the main challenge

In the same direction and in order to reduce the classification error and increase its accuracy, in this research, the focus is on DDoS attack in the environment of Internet of Things network devices based on dragonfly meta-heuristic algorithms for feature selection along with horse herd algorithm to improve MLP coefficients. Finally, after modeling and implementing the proposed method, its results will be compared with the basic method [7].

1.2 The importance of research

Considering the speed and size of digital attacks, human techniques are not sufficient to easily investigate the attack and respond appropriately to these attacks. Subsequently, the existence of free and intelligent programmed techniques that can quickly and accurately detect and respond to digital attacks has become a major test. These aspiring methodologies should have the option to deal with the entire attack response process in a timely manner, or at the very least, they should have the conclusion of what type of attack has occurred, what its goals are, and what is the appropriate response to it. attack, as well as how to focus and avoid Decide auxiliary attacks [11].

Breach detection is the most common way to investigate what is happening in an organization and examine it for signs of potential breaches or imminent threats to the organization's security arrangements. Interrupt countering is a method by performing interrupt detection and then stopping distinct attacks. Considering that a discontinuity detection framework generally handles a lot of information, the element determination step present in a large part of these frameworks is a difficult step because it only considers selected highlights as a subset of the full elements for Grouping thinks. This expands the accuracy of the characterization. Based on this, perhaps one of the main tests in the planning of the break location framework is the selection and reduction of the subset of suitable elements [10].

Recently, selection strategies and simplification methods have been highlighted to select the most important and appropriate elements. Since the characterization in breakpoint frameworks depends on the class mark, extra and redundant data output is quite expected in this context, and the component selection cycle assumes part of the fog. As a result, in this context, conventional computations (techniques in the light of hard and precise logic at the selection level) are meaningless in the frameworks of interrupt detection and against digital attacks that are gradually developing. For this purpose, imaginative methods such as the use of man-made consciousness techniques (artificial intelligence) and meta-heuristic calculations have become very common in this field today. because it provides acceptable flexibility and learning ability for these systems.

Research purposes

We seek after the accompanying objectives in our exploration

1) highlight extraction utilizing the computerized reasoning methodology and element determination utilizing the dragonfly metaheuristic calculation fully intent on decreasing the component in mix with the brain network grouping procedure in the field of interruption discovery frameworks, which covers the shortcomings of the past techniques and which gives great outcomes gets

2) Information preprocessing and include extraction and choice, which can decrease the elements and lessen the issue space, in this manner diminishing the intricacy.

3) Preparing the MLP brain organization and working on its coefficients with the pony group calculation that involves the information being referred to lessen the impact of assaults.

1.3 Exploration suspicions

The speculation of the examination is that the proposed interruption recognition framework ought to have an extraordinary element to be successful against serious digital assaults. These details incorporate the accompanying:

1) Convenient interruption discovery while the assault is underway or following

2) Misleading positive cautions ought to be limited

3) Human management ought to be limited and consistent activity guaranteed.

4) It is normal that the exactness of the grouping system will increment by utilizing the dragonfly calculation to choose the component.

1.4 Exploration questions

Is it conceivable to propose a technique in light of the dragonfly meta-heuristic calculation for highlight choice notwithstanding the pony group calculation to further develop the MLP coefficients to identify a DDoS assault in the climate of Web of Things network gadgets, which will expand the exactness?

1.5 Exploration advancement

As per the items communicated in this examination, an interruption recognition model in light of element extraction and choice and order of brain networks is proposed. In the proposed strategy, the BoT-IoT informational index is gathered first, and afterward the information is pre-handled to eliminate the commotion and missing information and additional items. In the accompanying, the information will be refined, accordingly, in this step, the elements have been chosen utilizing the dragonfly calculation. At this stage, a subset containing helpful highlights is chosen from the arrangement of accessible elements. In the subsequent stage, the information is separated into two sections, preparing and testing. In this step, the preparation informational collection is placed into a MLP brain organization and it is prepared. In any case, we are not happy with preparing the MLP network utilizing conventional methodologies. Rather, we have worked on the heaviness of neurons by utilizing the pony crowd calculation. At last, we have estimated the exhibition nature of the proposed technique by utilizing the information test segment. The development of the examination is in the component determination and characterization. The essential paper utilized the milk improvement strategy and the dragonfly calculation, and we utilize the pony group and dragonfly calculation. Among the as of late presented keen calculations, the Dragonfly Calculation, which is known for its quick discovery speed, has tracked down wide application in data

procurement as well as data assessment, and the Horse Gathering Advancement (HOA) estimation is a creative forward computation. which is easy to carry out and is for the most part utilized. In the field of progress, it has been utilized to manage circuitous systems and can follow the genuine and complex underpinnings of a structure and moreover comprehend its fluctuation to deal with the exactness of the survey to cycle the component determination. improve on utilizing it. HOA and DA coordinated processing disclosure. It very well may be accepted that it has not been utilized in that frame of mind to date

1.6 The idea of the Web of Things (IoT)

The Trap of Things is an association of genuine things, apparatuses, vehicles, structures, and other electronic gadgets embedded in these circumstances, the result of sensors and hierarchical affiliations that permit information assortment and business. The development of the Snare of Things is a central development towards current development. The power and improvement of the Snare of Things will fundamentally influence different areas of day to day existence and anticipated client conduct. Application circumstances where the Trap of Things will before long take a significant part incorporate shrewd homes, brilliant transportation systems, splendid metropolitan regions, horticulture, crisis center stock organization structures, vibration area savvy network structures, and so on [14] . As a general rule, the Trap of Things comprises of three fundamental parts:

1) Utilized things

2) Correspondence networks that speak with them.

3) PC structures for utilizing data stream from/to objects.

The Trap of Things is an association of unequivocal genuine things or things that incorporate programming, hardware, sensors, and trades that empower them to exchange data with makers, chiefs, and the different devices and organizations they speak with. to oversee Be that as it may, aside from everything, these associations have restrictions of limit, correspondence, energy and perception capacities. These limitations make the snare of things and various improvements consolidate and, as a development, increment the chance of expanding the requirements and execution of the current structure.

1.7 Related work

The main interruption discovery framework that has been enrolled so far depends on the exploration of Eduardo Delahoz et al. [23]. With a blend of self-sorting out maps (SOM) and measurable strategies, they have introduced a grouping strategy to

distinguish network irregularities. Their element choice incorporates the utilization of Fisher Error Proportion (FDR) notwithstanding Head Part Examination (PCA). The association's trades are then allocated utilizing expected automatic and skirmish departure directs typically or strangely. In this survey, they utilize a mix of versatile and SOM techniques to recognize network oddities. Subsequently, Head Part Examination (PCA) and Fisher Inconsistency Proportion (FDR) are utilized to decide the striking nature and denoising of the parametric self-arranging not entirely set in stone to address the component space and empower capability among typical and abnormal organizations. The area limits of the structure can be changed without guide retraining solely by changing the execution probabilities of the unit, and it deals with the quick execution of the hinder identification system to oblige the ceaseless association move rate.

Raval et al. [24] have proposed an intermingling strategy that utilizes a blend of nearness mining data. How much components related with every information point is decreased utilizing the K-Imples gathering computation. Additionally, BitFundamental holds the Winding Piece Capacity (RBF) utilized for portrayal. The component choice statement is utilized to indicate and choose a subset of the remarkable places and moreover to decrease the striking space, and is possibly fundamental for two reasons: computationally, it contains any of the accessible components or issues. In the assessment as test information is unfathomable. They appear to be restricted, generally utilized, and this data contains incalculable features that endlessly analyze them. Thus, the determination of significant components for classifiers and data decrease is a central question. Classification is the most common way of isolating a huge assortment of data into a bunch of things considering normal characteristics and afterward gathering them into critical bunches so the people in the gathering share some, yet not all, qualities. Whenever summation is performed on an informational collection considering one or more normal components between them, every information point contains n numbers. To arrange data in sets, comparable characteristics of every data point are thought of. Regardless, in the proposed system, just quantifiable components ought to be chosen to coordinate the data in a single collection and a couple of additional quantifiable features to be totaled in another conglomeration. These features are chosen by the different assault classes in the dataset and their attributes. These features are different for each assault class. One more norm for change determination incorporates assessed changes for the relating assault class. For instance, all traffic characteristics are considered for DOS assault recognition. In the first place, the K-concentrate on grouping estimation is applied to the preliminary data file. How much components related with every information point is decreased utilizing the K-thing characterization computation. To close individual informational collections are the principal bunches that work

with the primary assaults. Second, the wide suspicion of spot capacity (RBF) is applied to check whether or not a hinder happens. With the assistance of this part, the makers of the assurance stage can utilize this component to clean the data record of additional features. In this way, the boundless time intricacy of ordinary administration is decreased and the engineers utilize a standard dataset (KDD CUP 99). Sunil Pawar et al. [25] introduced a hereditary calculation based discovery network that has chromosomes of various lengths. Chromosomes with related highlights are utilized to produce the standard. The wellness capability of each standard is characterized utilizing the wellness capability. Every chromosome contains at least one standards to distinguish anomalies really. The determination of chromosomes is done arbitrarily involving roulette wheel choice as one of the choice procedures, determination of race, determination of positioning, choice of fixed conditions, and so forth. The chose chromosomes are restored along these lines.

1) It isn't ensured that every one of the ideal guidelines will be made.

2) It causes a misuse of figuring time.

Alduiri et al. [26] utilized counterfeit honey bee province (ABC) for irregularity location. They utilized tree characterization and Bayesian organization relapse and Markov chain for highlight choice. Swarm Information is a smart and composed machine of human-made thinking to determine and satisfy tricky responsibilities. It impersonates the exercises of bumble bee states and conglomerations like gatherings of birds, fish and underground bug territories, bumble bees and grasshoppers. In their investigation, the old KDD Cup 99 dataset was utilized for testing and planning.

Review [31] provides a comprehensive review of SDN-based DDoS attack detection and existing countermeasures arrangements. Software-defined organization (SDN) is an emerging worldview observed by many professionals to address the issues of current server farms. In this review, the arrangement is finished with regard to DDoS attack detection methods and the recognition of the necessities of a powerful arrangement brought to life by SDN capacities. It is also proposed to use the highlights of SDN for network security as well as an SDN-based protection system that has various elements involved in detecting and mitigating attacks. These elements include the separation of the control layer from the information layer, coherent integrated control, the opportunity to organize the organization by external projects, and programming-based traffic monitoring. Figure 3-2 shows the proposed architectural framework. The distributed control layer increases the reliability and scalability of the solution to meet the needs of a large data center.

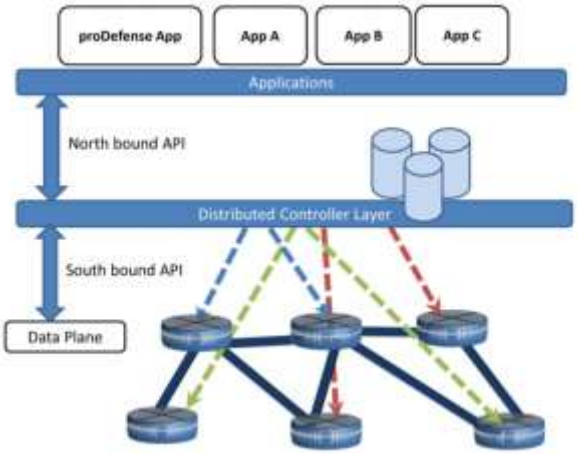

Figure (2-2) Architectural design [30] Pro-Defense

In the exploration [35], another mutt model is acquainted with assess the level of the info edge as per the ideal features of the association's data trade for planning. Recognizable confirmation requires flimsy information assortment. The assessments are amazing because of the quirk of meta-heuristics in the exploratory examination of organization trade data. These assessments are significant for giving and anticipating the likelihood of interference as per the subjective nuances in the trading of the association. The creators of this paper had the choice to utilize the NSL-KDD dataset of much number and multiclass with a 20% test dataset. The exploratory consequences of this investigation show that the semibreeding procedure is basically connected with restricting the computational intricacy and time in concluding the relating impact size of a part. The exactness of the proposed model was 99.81% and 98.56% for the second class and exploratory informational collections through the security method of the association and the research center equipment of the informational collections independently. As per the outcomes acquired by utilizing the proposed model, better precision of enormous misdirecting negative rate and low bogus positive rate is noticed. The proposed method is carried out in WEKA. WEKA is a computational information mining instrument used to perform request tests on the NSL-KDD dataset. The dataset comprises of various assault classes. Producers will keep on growing this field by doing a completely flowing organization ID. Moreover, these various procedures are utilized to work with correspondence between NIDS.

In the survey [36], the similarity of representative systems as for AI models has been researched. In this survey, two computations are proposed for the component of making a specific request of the foe, the principal estimation can choose the notable

focuses that can't be changed by the foe. While the subsequent calculation permits exact control. Their work has proposed a manner by which a foe can take advantage of a framework in view of AI methods by changing some framework inputs. Endeavors have been made to make a conventional technique for estimating the ill-disposed adaptability of various AI models, which prompts its correlation with these changes. Producers furnish the right feeling of force with exceptional lighting in multi-contact cross breed structures. The significance of model assessed impeding (MRB) presented by them is an action to assess the general strength of various models, and they likewise propose two new part determination estimations to construct a prepared classifier against assaults. At last, they assess the technique utilizing a true illustration of dynamic malware grouping utilizing a huge and excellent arrangement of noxious and pernicious malware. The creators show that the authenticity of utilizing foe data includes the assurance to construct more reliable classifiers and contend for working on the intrinsic consistency of outfit calculations instead of single-model calculations. Their strategy relies on two significant thoughts: first, the expense plan of a hard and fast assault and the expense of basic control. Together, these two clear conversations make the way for the showcase of ill-disposed limits in a pre-arranged execution circumstance.

3-2- An overview of the proposed method

Most datasets used in networking sessions have many features. This issue causes the complexity of the data and, by nature, the complexity of the machine learning algorithms used in intrusion detection problems to be very high, and as a result of this issue, it reduces the quality of the outputs of these algorithms. Therefore, in the proposed method, after the initial pre-processing, it is tried to reduce the features of the input data by using one of the dimensionality reduction methods, which includes the extraction and selection of the dragonfly feature. Then, in the classification part, an attempt will be made to learn a space similar to the learned space by using a neural network (MLP) whose weight parameters were improved by the horse herd algorithm. The three steps of the proposed method are as follows:

1) Pre-processing: This step usually includes methods of normalization, cleaning and aggregation of the input data.

2) Feature selection stage: For feature selection, it is necessary to have a method that can select important features in the layer with appropriate efficiency. In this dragonfly stage, the features with the highest score are selected among the features that are ranked based on the dragonfly approach.

3) Classification stage: In the second stage, the dimensions of the data are reduced and normalized, then in this stage the neural network is trained based on these data. In the following, an MLP multilayer neural network method is used for classification.

3-3- Suggested method

The framework of the proposed model used for intrusion detection in IoT networks is given in Figure 1-3. In the first step, the raw data needed for forecasting is collected. Then, the most relevant features that affect the detection of healthy leaks from attack sessions are selected by the objective function of the dragonfly algorithm. In the last stage of the proposed model, the MLP network is built as a classifier, and in the proposed model, the horse head algorithm is used to improve the training of the MLP network and, as a result, find the best set of weights and biases that provide accurate prediction for diagnosis. Finally, the prediction results obtained are evaluated.

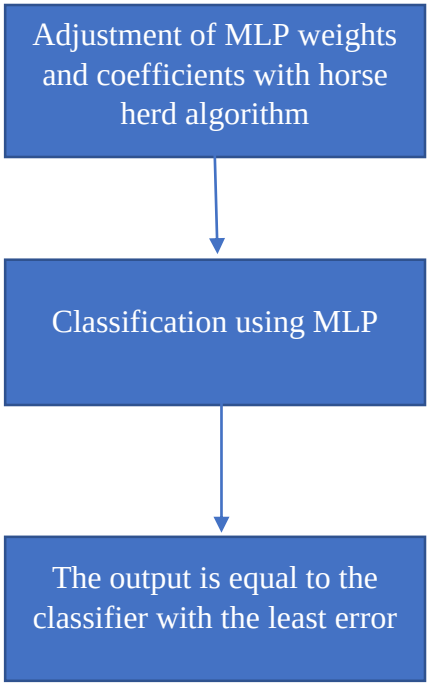

Figure 3-1. The framework of the proposed model

- First step: data preprocessing

In this step, the data is refined. In fact, this step is considered as data pre-processing. This step includes two parts including removing incomplete data and normalizing

the data. The existence of incomplete data in the data set is always inevitable. This issue can be due to reasons such as human errors, system failure, problems in information transmission, or other cases. But the presence of incomplete data can impose bad consequences on the system. This problem becomes more acute when the system used is based on machine learning. Because in such systems, the correctness of the information is assumed, and any defect in the received data can lead to the wrong output of the system. In the proposed method, the data received from the input is checked and the presence of any incomplete data is discovered and these data are destroyed. But the second part also includes data normalization. The data used in any numerical algorithm can be either binary or continuous. Therefore, in pre-processing, the main issue is converting the data into binary form or normalized continuous form. To convert the numbers and to normalize the characteristics, first a statistical study is done on each of the characteristics based on the data in the data set and a maximum value is determined for the values of each characteristic. Then, based on the following relationship, the numbers are converted into normalized form.

- The second step: selecting the feature based on the dragonfly algorithm

In this step, the pre-processed data is entered and the information of different meetings is extracted and selected using their characteristics, which helps to improve the accuracy of detecting attacks. Unselected feature values such as irrelevant, redundant and unnecessary are not useful for attack classification. Therefore, the feature selection technique is used to select the appropriate features to determine the accuracy in the search space.

Basically, the key goal of feature selection methods is to maximize the classification ability of a learning model by selecting an approximately optimal feature subset. Therefore, in order to select an effective subset of features, a strong algorithm is needed that correctly identifies the subset of features required to classify the examined patterns. It should be noted that the goal of feature selection is not to find an individual feature that is relevant to the classification problem, but rather a combination of different features that when added together reveal the pattern deeply. Feature selection makes classification problems computationally efficient by reducing the cost of classifying patterns with large feature dimensions. This means that the data storage and computational resources required for the training phase can be reduced. Among the feature selection methods, comprehensive search method is impractical due to high computational cost, especially for large data sets. Population-based meta-heuristic methods cannot guarantee the desired results. However, they can achieve an acceptable feature subset in a reasonable amount of time. Since in recent years population-based algorithms have shown their advantages in solving

feature selection tasks, various population-based methods (metaheuristics) and even some recent algorithms have been used to solve feature selection problems.

The third step: improving the MLP coefficients with the horse herd algorithm

In this step, the data that have been normalized and feature reduced are analyzed using the MLP neural network. These classifiers try to be trained using the meetings of each cluster of information. After the additional features of the data are removed, they will be provided to the multi-layer perceptron neural network.

A multilayer perceptron (MLP) is an artificial neural network with a feedforward architecture that has multiple layers of nodes or neurons. A typical MLP architecture has an input layer, one or more hidden layers, and an output layer. The input layer of the MLP is where the input data is presented, which can be in the form of a vector image or any other type of data that can be represented numerically. An attribute or feature of the input data is represented by each node in the input layer. Using a network of weighted connections between nodes in the adjacent levels of the hidden layers, an MLP is calculated on the input data. Each node in a hidden layer is connected to each node in the upper layer, and each connection has a weight assigned to it. Hidden layers add nonlinearity to the output of nodes using activation functions. The final output of the network is generated by an MLP output layer. The number of nodes in the output layer is determined by the solved problem. For example, a node in the output layer produces a binary output for a binary classification task. There will be multiple nodes in the output layer for a multi-class classification topic, each corresponding to a class label. During training, an MLP learns the weights of connections between nodes. During training, a set of input data and their corresponding output labels are sent to the network. It modifies the connection weight to reduce the difference between the expected output and the actual output of the network. This process is repeated in several cycles until the network performance in a validation set reaches an acceptable level.

In this step, in order to find the optimal input parameters for MLP, the horse herd algorithm with archive-based concepts is used. HOA is a population-based optimization algorithm based on crowd intelligence that mimics the cooperative herd behavior style of horses for grazing and living. Every HOA implementation starts with a random set of solutions. During the iterative improvement loop HOA uses the strong features of current solutions to create a new solution through its smart operators. Typically, the results of any HOA enforcement will vary from enforcement to enforcement. The run is repeated several times to ensure that the results are approximately stable. The horse herd optimization algorithm to improve MLP-NN, which we abbreviate as AHOA-NN in the following work, uses an archive to store the best solutions obtained in the current run and use them as part of the

initial population of the next run. Uses. The number of best solutions stored in the archive depends on the chosen archiving rate (Ar) where (A_r∈[01]). This process is repeated during the evolution of the algorithm. This section presents the detailed steps of HOA and how to implement and improve it to enhance the performance of MLP by providing an optimal input vector (weights and biases) that shows better prediction results for seat changes.

1.8 Data collection

The BoT-IoT dataset was the one used for the evaluations [45]. At UNSW Canberra's Cyber Range Lab, a genuine network environment was designed in order to build this dataset. Botnet and regular traffic coexist in the network environment. A variety of file formats for the dataset source files are available, such as the original PCAP files, produced Argus files, and CSV files with session information. To aid in the labeling process, these files have sessions divided into attack categories and subcategories. The acquired PCAP files include about 72,000,000 records and a size of 69.3 GB. DDoS DoS OS assaults, Service Scan Keylogging, and Data Exfiltration are all included in the dataset. This data set's features are listed in Table (1-4).

Table (4-1) characteristics of the data set

Feature	Description
pkSeqID	Row Identifier
Stime	Record start time
flgs	Flow state flags seen in transactions
flgs_number	Numerical representation of feature flags
Proto	Textual representation of transaction protocols present in network flow
proto_number	Numerical representation of feature proto
saddr Source	IP address
sport Source	port number
daddr	Destination IP address
dport	Destination port number
pkts	Total count of packets in transaction
bytes	Totan number of bytes in transaction
state	Transaction state
state_number	Numerical representation of feature state
ltime	Record last time
seq	Argus sequence number

dur	Record total duration
mean	Average duration of aggregated records
stddev	Standard deviation of aggregated records
sum	Total duration of aggregated records
min	Minimum duration of aggregated records
max	Maximum duration of aggregated records
spkts	Source-to-destination packet count
dpkts	Destination-to-source packet count
sbytes	Source-to-destination byte count
dbytes	Destination-to-source byte count
rate	Total packets per second in transaction
srate	Source-to-destination packets per second
drate	Destination-to-source packets per second
attack	Class label: 0 for Normal traffic 1 for Attack Traffic
category	Traffic category
subcategory	Traffic subcategory

1.9 Evaluation criteria and simulation parameters

In data mining science, the four criteria of Accuracy, Precision Recall and F-Measure are always used for data classification. These criteria, which are among the main criteria for determining the efficiency of classifications, can well analyze the performance of a classification. These criteria and their equations are shown below

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1\text{-}4)$$

$$Precision = \frac{TP}{TP+FP} \quad (2\text{-}4)$$

$$Recall = \frac{TP}{TP+FN} \quad (3\text{-}4)$$

$$\text{F-Measure} = \frac{2*Precision*Recall}{Precision+Recall} \quad (4\text{-}4)$$

It should be noted that in these equations, True Positive (TP) represents the correct prediction of a healthy session. True Negative (TN) indicates a correct prediction

about a carry. A false positive (FP) indicates an incorrect prediction of an attack session. A False Negative (FN) indicates an incorrect prediction of a healthy session.

Before presenting the evaluation results, it is necessary to have a look at the parameters used in the evaluations.

1.10 Feature selection and evaluation results

In this section, the evaluation results will be evaluated in the form of data sets. In fact, in this section, evaluations have been done separately for each of the data sets, and the results can be seen in different forms. It should be noted that in the proposed system, the data sets are divided into two parts, training and testing. In the next step, when feature selection is done using the dragonfly algorithm, a subset of features is selected and the rest of the features are removed. In the third step, when it comes to using classification, we have used a multilayer perceptron neural network. This neural network uses horse herd algorithm to improve the weight of neural network neurons trained by the dataset training department. In the next step, we will verify the performance of the proposed approach using the test data set. Considering that the size of the training data set can have a great effect on the performance of the compared methods, therefore, in our experiments, we have used two different sizes for the training data set, i.e. 50% and 80%. It is clear that when the size of the training data set was 80%, the size of the test data set was considered to be 20%, and when the size of the training data set was 50%, the size of the test data set was also considered to be 50%.

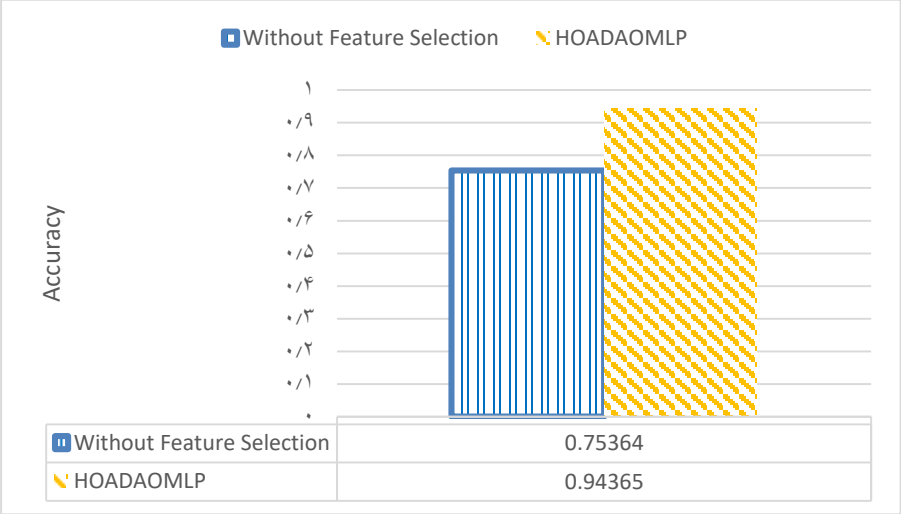

Figure (4-1) Examination of component choice strategies

The consequences of the above figure demonstrate the prevalence of the proposed strategy that involves the dragonfly calculation for highlight choice. Likewise, this

diagram shows that without choosing the element, the nature of the proposed strategy is extraordinarily diminished. Taking into account that in this segment it was found that the proposed technique performs better in the method of utilizing highlight choice, we will keep on introducing the consequences of the proposed strategy as the models introduced in the past area.

The principal standard we assess is Precision. This model, which shows the precision of the presentation of the looked at strategies, is vital. As should be visible in the situation of this action, the denominator incorporates the quantity of right forecasts and the denominator incorporates all expectations. Truth be told, this action shows the number of rates of forecasts that have been made accurately. Figure (4-2) shows how much precision got for every one of the informational collections.

But before presenting the evaluation results, it is necessary to conduct experiments on the proposed method and feature selection stage. Therefore, we have tested the proposed method for two different situations. These modes include the proposed method without using the feature selection step and feature selection using the dragonfly algorithm. In order to know how much effect the feature selection algorithm had on the performance of the proposed method, Figure (1-4) shows the results of the evaluations for the stated states according to the Accuracy criterion.

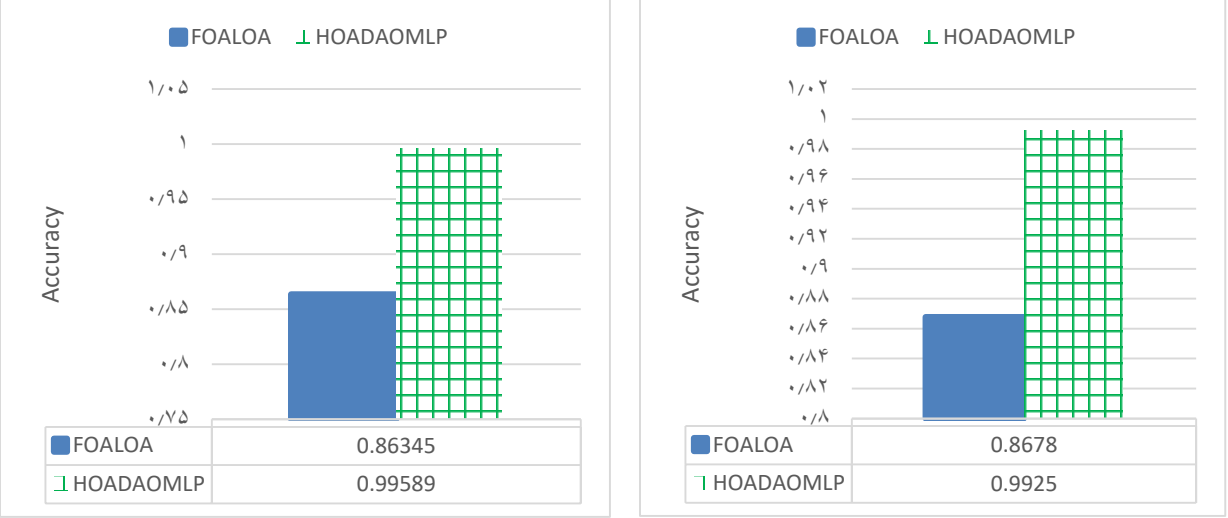

Figure (2-4) Accuracy obtained by each of the compared methods

As it is clear in this figure, the proposed approach has been able to improve compared to the compared approach for both sizes of the training dataset. In fact, the results of this figure state that for healthy meetings, the proposed approach has provided better predictions than the compared approach. The fact that the proposed approach has been able to achieve higher precision shows that it has made better

decisions about the sessions that were attacked and healthy. That is, if it detects a session as healthy or attacked, there is a high probability that the session in question was healthy or attacked. Because the focus of the Precision equation is on the correctness of the prediction, i.e. TP.

The third criterion to evaluate is recall. This criterion, like the Precision criterion, is related to the response of each of the compared methods to healthy and unhealthy sessions. Figure (4-4) contains the results of the proposed method and the compared method for both data sets.

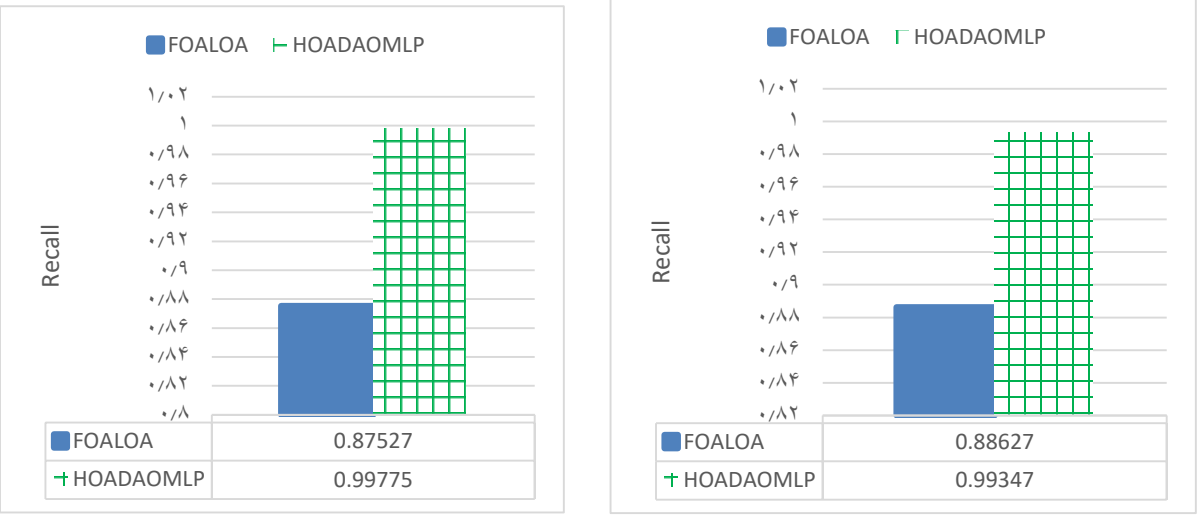

Figure (4-4) The amount of recall obtained by each of the compared methods

1.11 Conclusions

With the ever-increasing expansion of computer systems and then computer networks, it is the turn of a new generation of these networks called IoT networks. These networks make it possible to communicate between people and things. This issue leads to people and things being able to find a mutual understanding of each other. This understanding gives humans new abilities that provide countless benefits. However, these networks, like all computer networks, are not safe from the influence of profiteers. Therefore, intrusion detection has become a challenging issue for these networks. In this thesis, we have presented an approach for intrusion detection in IoT networks. The proposed system consists of several different steps. In the first step, we will refine the data. Therefore, in this step, feature selection using an evolutionary algorithm has been put on the agenda. At this stage, feature selection has been done using the dragonfly algorithm. At this stage, a subset containing useful features is selected from the set of available features. In the next step, the data is divided into two parts, training and testing. In this step, the training data set is entered

into an MLP neural network and it is trained. But we are not satisfied with training the MLP network using traditional approaches. Rather, we have improved the weight of neurons by using the horse herd algorithm. Finally, we have measured the performance quality of the proposed method by using the data test section. MATLAB programming language has been used for evaluations. Also, in order to measure the quality of the proposed method, the approach presented in [7] has also been implemented along with the proposed method and measured using the Bot-IoT dataset. The evaluation criteria including Accuracy, Precision Recall and F-Measure show that the proposed method has been able to show much better performance than the compared method. Future works The issue of intrusion detection is always a hot and interesting issue. Therefore, we recommend interested researchers to continue research on this issue. Therefore, our proposal to new researchers to use the framework of this thesis is a feature selection approach combined with improved classifiers. For this, they can use new evolutionary algorithms, such as point hyena hunting dolphins or weed, and for classification, they can go to new classifications, such as Feed Forward neural networks.

1.12 Reference

[1] Dwivedi S. Vardhan M. & Tripathi S. (2020). Defense against distributed DoS attack detection by using intelligent evolutionary algorithm. *International Journal of Computers and Applications* 1-11

[2] Alaba F. A. Othman M. Hashem I. A. T. & Alotaibi F. (2017). Internet of Things security: A survey. *Journal of Network and Computer Applications 88* 10-28.

[3] Liu G Quan W Cheng N Zhang H Yu S (2019) Efficient ddos attacks mitigation for stateful forwarding in internet of things. J Netw Comput Appl 130:1–13

[4] Chen W Xiao S Liu L Jiang X Tang Z (2020) A ddos attacks traceback scheme for sdn-based smart city. Comput & Electr Eng 81:106503

[5] Gaikwad D. P. & Thool R. C. (2015). Intrusion detection system using bagging with partial decision treebase classifier. *Procedia Computer Science 49* 92-98

[6] Aljawarneh S. Aldwairi M. & Yassein M. B. (2018). Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model. *Journal of Computational Science 25* 152-160.

[7] Krishna E. P. & Thangavelu A. (2021). Attack detection in IoT devices using hybrid metaheuristic lion optimization algorithm and firefly optimization algorithm. *International Journal of System Assurance Engineering and Management* 1-14

[8] MiarNaeimi F. Azizyan G. & Rashki M. (2021). Horse herd optimization algorithm: A nature-inspired algorithm for high-dimensional optimization problems. *Knowledge-Based Systems 213* 106711.

[9] Meraihi Y. Ramdane-Cherif A. Acheli D. & Mahseur M. (2020). Dragonfly algorithm: a comprehensive review and applications. *Neural Computing and Applications 32* 16625-16646

[10] Sarvari Samira et al. An efficient anomaly intrusion detection method with feature selection and evolutionary neural network. *IEEE Access* 8 (2020): 70651-70663

[11] Saied A. Overill R. E. & Radzik T. (2016). Detection of known and unknown DDoS attacks using Artificial Neural Networks. *Neurocomputing 172* 385-393

[12] Battiti R. Brunato M. & Mascia F. (2008). Reactive search and intelligent optimization (Vol. 45). Springer Science & Business Media.

[13] Karlik B. & Olgac A. V. (2011). Performance analysis of various activation functions in generalized MLP architectures of neural networks. International Journal of Artificial Intelligence and Expert Systems 1(4) 111-122.

[14] K. Han Y. Wang C. Zhang C. Li C. Xu Autoencoder inspired unsupervised feature selection in: 2018 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP) IEEE 2018 pp. 2941–2945.

[15] Mohammadi S. & Amiri F. (2019). An efficient hybrid self-learning intrusion detection system based on neural networks. *International Journal of Computational Intelligence and Applications 18*(01) 1950001

[16] Kadhim Y. & Mishra A. (2019 November). Radial Basis Function (RBF) Based on Multistage Autoencoders for Intrusion Detection system (IDS). In *2019 1st International Informatics and Software Engineering Conference (UBMYK)* (pp. 1-4). IEEE

[17] https://arzdigital.com/what-is-ddos-attack/

[18] L. Yang S.-H. Yang and L. Plotnick How the internet of things technology enhances emergency response operations *Technological Forecasting and Social Change* Vol. 80 No. 9 pp. 1854-1867 Nov. 2013.

[19] Mahdi Ozbak Zaei Internet of things challenges https://www.researchgate.net/ publication/ 320170323 2017

[20] Bettini C. & Riboni D. (2015). Privacy protection in pervasive systems: State of the art and technical challenges..