Baghdad Science Journal

Volume 23 | Issue 2 Article 29

2-24-2026

Honey Badger Algorithm-Based Feature Selection for DDoS Attack Detection in IoT Networks

Mustafa Azeez al-Mayyahi
Software Department, College of Computer Science & Information Technology, Wasit University, Wasit, Iraq, mkhalaf@uowasit.edu.iq

Ahmed Raad Al-Sudani
Software Department, College of Computer Science & Information Technology, Wasit University, Wasit, Iraq, araad@uowasit.edu.iq

How to Cite this Article

al-Mayyahi, Mustafa Azeez and Al-Sudani, Ahmed Raad (2026) "Honey Badger Algorithm-Based Feature Selection for DDoS Attack Detection in IoT Networks," *Baghdad Science Journal*: Vol. 23: Iss. 2, Article 29.
DOI: https://doi.org/10.21123/2411-7986.5221

BAGHDAD SCIENCE JOURNAL 2026;23(2):747–761

RESEARCH ARTICLE

Honey Badger Algorithm-Based Feature Selection for DDoS Attack Detection in IoT Networks

Mustafa Azeez al-Mayyahi *, **Ahmed Raad Al-Sudani**

Software Department, College of Computer Science & Information Technology, Wasit University, Wasit, Iraq

ABSTRACT

By quick computer and communication technology improvement, Distributed Denial of Service (DDoS) attack harm is getting more important. DDoS attacks research is the important study domain; a number of methods exist which have been presented like the algorithm of evolutionary as well as artificial intelligence in literature to diagnose attacks of DDoS. Unfortunately, new popular models of DDoS diagnosis are deteriorating for validating DDoS attacks objective and prior identification. Because of DDoS attack modes diversity as well as various attack traffic amount, still there is not the technique of diagnosis with promising accuracy of diagnosis currently. By choosing the best subset of features, a feature selection (FS) approach helps to shorten computing times and increase computational complexity. For mitigating attacks of denial of service, this paper applies honey badger algorithm (HBA) with algorithm of machine learning known as HBIDS. Present strategy is given the making intrusion detection system (IDS) for meeting controlled area needs and could recognize between attack and normal traffics. Moreover, HBIDS chooses the most related from basic dataset of IDS which could aid recognizing normal low-speed DDoS attacks, then chosen features are conveyed to classifiers like decision tree, multilayer perceptron, naïve Bayes, and support vector machine for identifying attack kind. Generally accessible dataset as CIC-IDS 2017 and KDD Cup 99 are applied for our experimental research. From simulation outcomes, this is obvious that HBIDS with decision tree needs high diagnosis with the low false–positive rate (0.001) and accuracy (99.9).

Keywords: Distributed denial of service, Honey badger algorithm, Intrusion detection system, Machine learning, Security risk analysis

Introduction

The Internet of Things (IoT) lets various devices with the address of Internet Protocol (IP) like medical equipment, offices, wearable smart devices, home appliances, smart cars, and areas of monitoring tools be linked together through the Internet for collection, provision, exchange and storing data amongst themselves. Quick IoT rise has led to big data numbers creation and a huge raise in IP devices' numbers linked to the internet. IoT creates big data amounts which software of IoT applies for analyzing data.[1] Typically, devices of IoT apply wireless mediums to broadcast data which in turn presents attackers with a simpler objective to attack. Networks' improvement like programmable data planes for increasing supervising tasks of network and securing physical objects' networks joined to the area of internet/IoT. In a local network, the typical attack of communication could be limited to small local areas/ nodes. However, areas of IoT are largely influenced and had a devastating impact by attack development in the system of IoT. Attacks in a layer of the network are shared in 4 basic groups in a dataset of KDD99 and current version NSL-KDD, known as (a) DoS where the attacker attempted to access service/ machine occupied by smart customer; (b) Probe where the attacker attempted in achieving info on designated network applying scanning tasks of host and network; (c) U2R where the attacker has exploited some

Received 16 February 2024; revised 25 October 2024; accepted 27 October 2024.
Available online 24 February 2026

* Corresponding author.
E-mail addresses: mkhalaf@uowasit.edu.iq (M. A. A. Mayyahi), araad@uowasit.edu.iq (A. R. A. Sudani).

https://doi.org/10.21123/2411-7986.5221

conventional attack techniques such as malware infection and stolen credentials for escalating the chance from limited privilege to super/root user access, (d) R2L, where the attacker might perform local users simulation and obtain the right to designated machine. [2,3]

Attacks of DoS are the specific attacks prevalent level in online communication among networks using internet services for utility, storage, and processing. Such attacks develop network congestion defining zombie packets that are infected packets that contaminate good packets as they progress down layers of communication. Attacks of DDoS make security threats to the new Internet, particularly the area of IOT as the IoT devices have restricted memory, processing security, and power measures for avoiding the attacks of DoS. Attacks of DDoS are serious attacks in contrast to connected devices of IoT and make network performance worse. This decreases network and calculation sources like network bandwidth, CPU, and memory when conveying multimedia streaming data as streaming data are naturally ongoing, and time-sensitive. [4]

A class of algorithms known as metaheuristic search techniques just needs the objective function and the variables' domain, not the in-depth knowledge of the search space. Metaheuristic methods and machine learning performance are used in present years for diagnosing intruders and developing intrusion detection system (IDS) performance—methods of IDSs signature/ unusual diagnosis. An IDS based on signature is given the familiar attack signatures knowledge base that is made up-to-date on a regular base. The traffic of the network which matches attack signatures creates alarms. However, basic misuse-based IDS restriction is that this could not identify unfamiliar patterns of attack. [5]

Against, the unusual strategy of diagnosis could recognize the last unfamiliar patterns of attack that this method does not make an impact for matching specific signs of attack. Alternatively, this checks uneven runtime features which distinguish from normal network manner. Yet, there is a large number of false alarms number limitation as this has opened the broad study domain for investigators. Accessible data of network traffic in general includes extra/unrelated features that could importantly impact the performance of classification. Therefore, this could impact anomaly-based IDS performance and certainly delay the diagnosis engine in making an accurate decision. So, for achieving optimum features' subset, this is needed for employing the appropriate method of feature selection (FS) for dimensionality decrease. [6]

FS is the combinatorial issue of optimization which wants to extract related provided group features. The effective method of FS importantly accelerates the modeling classifier process, and simple classifier structure, at last, increases the performance of classification while this is strictly exaggerated by extra and unrelated features. For controlling attacks of DDoS, some methods of FS have been defined and combined in IDS which could distinguish between usual/ attack data. in general, there are two basic strategies to select important features, called wrapper and filter techniques. In methods of filter, FS is independently run with no classifiers usage, depending on features' intrinsic properties, while in methods of wrapper, the features' subset is evaluated given the performance of classification. Classifier works on selected groups of features when processing them through linear/nonlinear schemes.

Some in-depth algorithms of machine learning have been defined in literature for making efficient IDS, based on multilayer perceptions, artificial neural networks, fuzzy logic, principal element analysis, k-nearest neighbor, decision tree, and Bayesian network. Amongst such strategies, investigators have recognized the decision tree which is the encouraging engine in efficient IDS creation because of the good performance in terms of robustness transparency, performance, and particularly, use ease. This tries to achieve more suitable features for a predetermined model of learning which assures higher learning performances, although needs more computational time and has been recognized as economically more expensive in comparison with the scheme of the filter. Robust and effective IDS development is yet the present study issue for investigators because of dynamic area adaptation and unbalanced datasets with high amounts of traffic.

The increasing frequency of cyberattacks is the main obstacle in the context of network deployment. Numerous researchers are working in this sector to provide various methods for network security against cyberattacks in order to safeguard the network environment against unanticipated cyberattacks. Detection systems employ feature selection strategies to address a variety of problems, including big datasets, accuracy, asymmetric information circulation, and the ability to discern between normal and aberrant data. This makes it essential to talk about two concepts: feature selection (FS), which offers the principles of bioinspired algorithms meant to identify assaults, and intrusion detection systems (IDS), which present machine learning algorithms developed primarily to detect attacks. To address the above-mentioned issues, investigators have experimented with various algorithms' kinds to choose features for better-recognizing attacks on DDoS. Several works illustrate rising trends in metaheuristic

optimization methods' usage for coping with attacks of DDoS.[7] Metaheuristics is preferred as this has been identified that they get better solutions than others, like iterative techniques, optimization algorithms/simple heuristics with less computational impact. Some samples include artificial bee colony (ABC), chi-square (CHI), particle swarm optimization (PSO), colony optimization (ACO), and genetic algorithm (GA) employed as a wrapper; information gain (IG) and correlation feature selection (CFS), as filter methods. Feature decrease with metaheuristics illustrates better outcomes as they create the best optimal outcomes in less time, so this decreases computational IDS design cost and develops IDS classification performance. Most classifiers, application area irrespective, use a model of ranking for selecting the last features' subset, in an ad-hoc way and is widely performed in the last studies.

For increasing IDS performance, in this paper, the honey badger algorithm (HBA) with machine learning algorithms (HBIDS) is defined for coping with stagnation concern which is recognized in traditional methods of optimization. HBIDS is used for providing an effective and useful global search space for identifying related features and afterward uses support vector machine (SVM), naïve Bayes (NB), decision tree (C4.5), and multilayer perceptron (MLP) as classification algorithms to assess IDS performance on datasets of CIC-IDS 2017 and KDD Cup 99. This is recognized that the presented technique with C4.5 achieves the best performance compared with the other classifiers, obtaining a high rate of diagnosis, with a low false–positive rate (FPR) and accuracy. In addition, the presented method develops solution quality decreases training time, and develops classifier efficiency. The contributions of this paper are as follows.

- Adapt a scheme using honey badger algorithm to feature selection in diagnose DDoS attacks in network.
- Compare the result between using support vector machine (SVM), naïve Bayes (NB), decision tree (C4.5), and multilayer perceptron (MLP) for detecting DDoS attacks.
- Experiments were carried out to evaluate the performance of the proposed method, which is demonstrated by using datasets of KDD Cup 99 and CIC-IDS 2017.

Literature review

An infiltration effort to stop regular traffic to a website, server, or other network resource is known as a distributed denial-of-service (DDoS) attack. The result is a denial of service for authorized users. DDoS assaults are successful when they take use of several hacked computer systems as attack traffic sources. A denial-of-service attack (DDoS) might be conceptualized as a traffic congestion that obstructs a roadway, preventing fixed traffic from reaching its intended destination. Choosing the most pertinent aspect that may offer the best accuracy rate is the primary problem when it comes to DDoS Attacks. Numerous academics suggested various methods for choosing the finest characteristics in order to address this. Using metaheuristic algorithms, like the particle swarm optimization (PSO), is one of these methods. This paper proposes the use of HBA in conjunction with SVM to identify DDoS. In this section, discuss the related studies from perspective using feature selection in the diagnosis of DDoS attacks.

Alzahrani and Alzahrani,[8] they applied CICDDoS2019 which is the newest and full dataset available by the Canadian Institute for Cybersecurity. This has checked 6 different algorithms of ML: Support Vector Machine (SVM), Random Forest (RF), Naive Bayes (NB), K-Nearest Neighbors (KNN), Logistic Regression (LR), and Decision Tree (DT). Measurements such as accuracy, true-positive rate, precision, recall, and F-measure false-positive rate have been applied in the assessment. DT is greater than RF due to that this has less calculation time of 4.53 s and 84.2 s, in turn. Outcomes illustrate that schemes of ML are successful in attack traffic detection. The present paper targets contributing to the study performed here. The present article contributes that as illustrated in tests, feature selection techniques of random forest regressor (RFR) raise ML techniques' accuracy in diagnosing traffic of attack. Present paper performance could be employed in real-life systems in various IoT fields. At last, coming feasibilities and restrictions for network unusual systems of mitigation in IoT are explored. Outcomes illustrate that schemes of ML are successful in diagnosing traffic of attack. The present work aims to contribute study performed here. The experimental outcomes illustrated that applying feature selection techniques of random forest regressor (RFR) raises ML techniques' accuracy in diagnosing traffic of attack. For attacks like DDoS which require intervening with no time-wasting, detecting attack traffic is essential applying the resources of the system as effectively as feasible.

Karthik et al,[9] concentrates on Mirai and DDoS attack innovative diagnosis and mitigation in the area of IoT. The aim of the present study is securing IoT devices from DDoS and Mirai botnet attacks, improving the device of IoT, applying protocol of EDQP by Network simulator, modeling the Innovative algorithm of diagnosis as Hybrid Strawberry African

Buffalo Optimizer (HSABO) for diagnosing DDoS and Mirai botnet attacks. Accordingly, the encryption mechanism McEliece encryption with EDoS-Shield architecture is applied to prevent high-rate DDoS and Mirai botnet attacks. So, high crowed and bad plan is turned to EDoS-Shield architecture that shows in the mechanism of encryption. IoT devices' number is arranged with the new Hybrid Strawberry and African Buffalo Optimization (HSBABO) aid. Accordingly, DDoS attacks kinds are launched in improved IoT networks. In addition, strawberry and African Buffalo fitness existence is used for detecting and specifying kinds of attack. Accordingly, the new encryption of MCELIECE with the Cloud Shield model is improved for avoiding low and high-rate DDoS attacks in IOTs. At last, the presented scheme obtained 94% of attack diagnosis accuracy, 3% of the false negative rate, and 5.5% of the false positive rate.

The aim of Sambangi and Gondi[10] is to create a model of machine learning which is the feature selection ensemble applying info gain and regression analysis. For the experimental study, the dataset taken was generally known CICIDS 2017 dataset. Particularly, morning and afternoon Friday logfile are taken that have Benign, Bot, and DDoS levels. This has been seen via this ensemble model for the Friday morning dataset, the prediction accuracy of 97.86% is obtained. Similarly, for the Friday afternoon log file, the accuracy of prediction is achieved as 73.79% for 16 features via ML model based on regression analysis as well as feature selection based on information gain. The present work paved the way for illustrating regression analysis's importance in creating the model of ML and illustrates several significant visualizations like residual plots and fit charts that prove the model's significance and its suitability for taking predictions of the model. Here, the writers have restricted analysis for the one-day log files, and in the future, present work might be spread for taking whole traffic five-day log files and come out with a machine learning model based on consensus.

Dwivedi et al,[11] is conducted in two stages: at first, features are chosen via info gain and chosen features are conveyed to various classifiers such as SVM, C4.5, MLP, and NB for classifying attacks of DDoS. This has been recognized that the information gain method of the filter with the C4.5 classifier has better performance than other methods of prevailing according to false-positive rate, accuracy, and diagnosis for recognizing attacks of DDoS in a dataset of KDD Cup 99. Basic information gain is applied as the method of filter for choosing features, using MLP, C4.5, SVM, and NB as algorithms of classification for IDS performance assessment on CAIDA DDOS Attack 2007, CONFICKER worm, KDD Cup 99, UNINA traffic traces datasets. This is recognized that the presented algorithm (IGIDS) with C4.5 has the finest performance compared with the other classifiers, obtaining a high diagnosis rate and accuracy with a low false-positive rate. The presented method develops solution quality, decreases the time of training, and develops classifier efficiency. Based on the obtained outcomes, the presented method validates the efficacy in DDoS attack diagnosis given the metrics of performance like 3.01% FPR, 98.53% F-measure, 98.79% AUC, and 4.16% FPR, 97.43% F-measure, 97.57% AUC, using KDD Cup 99 and CONFICKER worm datasets in turn.

Asghari et al,[12] the developed Horse Optimization Algorithm (HOA) version known as BHOA is presented as the FS technique based on the wrapper. For converting ongoing to discrete search space, S-shaped and V-shaped functions of transfer are taken. In addition, for handling the pressure of choice, exploration, Power Distance Sums Scaling strategy, and abilities of exploitation are applied for selecting to scale the population fitness values. The presented technique of efficiency is computed on 17 standard benchmark datasets. Outcomes of implementation prove the presented technique's efficiency given the V-shaped transfer functions group in comparison with other transfer functions and other FS algorithms based on the wrapper.

Aljebreen et al,[13] models the novel attack diagnosis of DDoS applying snake optimizer with ensemble learning (DDAD-SOEL) method on the platform of IoT platform. DDAD-SOEL strategy aim depends on the DDoS attack's effectual and automated recognition. To obtain this, the DDAD-SOEL method uses SO algorithm for feature subset choice. Moreover, 3 DL strategies' ensemble is known as bidirectional long short-term memory (BiLSTM), long short-term memory (LSTM), and deep belief network (DBN) strategy. At last, the Adadelta optimizer could be used for DL algorithms parameter tuning. DDAD-SOEL method simulation value was tested on a database of benchmarks and the result shows DDAD-SOEL method developments over other recent schemes in terms of distinct measures.

AIghuraibawi et al,[14] presents a technique to diagnose ICMPv6 DDoS attacks applying an edited Flower Pollination Algorithm (MFPA). The main aim of the presented technique, at first MFPA is to choose the most related features from a dataset of ICMPv6 for detecting attacks of ICMPv6 DDoS. Secondly, the increased diagnosing DDoS flooding attacks scheme given the multi-objective FPA has presented. Outcomes illustrate that the first presented technique (MFPA) obtained the best accuracy equal to 97.96%

with 10 features and the second presented technique Multi-Objective FPA obtained the best accuracy of 97.01% with five features.

Dwivedi et al,[15] for mitigating denial of service attacks, apply the grasshopper optimization algorithm (GOA) with the algorithm of ML known as GOIDS. The present strategy is given to make the IDS meet controlled area needs and differentiate between attack and normal traffic. Additionally, GOIDS chooses the most related features from a basic dataset of IDS which could help to differentiate normal low-speed DDoS attacks, and selected features are transferred to classifiers, like SVM, multilayer perceptron, decision tree, naïve Bayes for recognizing attack kind. Generally, accessible datasets such as CIC-IDS 2017 and KDD Cup 99 are applied for our experimental research. From simulation outcomes, this is obvious that GOIDS with a decision tree needs high diagnosis and with a low false–positive rate and accuracy.

Pandithurai et al,[16] demonstrated a DDoS attack prediction utilizing a honey badger optimization method based on feature selection and Bi-LSTM. The method starts with gathering input characteristics from the DDoS attack dataset. After that, input features are sent to Z-Score and Bayesian normalization, among other preprocessing stages. The feature selection step, which uses Honey Badger Optimization (HBO), receives preprocessed data. To select the best feature in this instance, the features are selected by lowering their MSE. The Bidirectional Long short-term Memory (Bi-LSTM) classifier is then supplied with the best characteristics to anticipate DDoS assaults.

Mahmood et al,[17] four machine-learning based classifiers have been trained in this work using two distinct sets of chosen characteristics. The Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) approaches, respectively, serve as the foundation for the two sets of chosen characteristics. It is well known that these evolutionary-based methods work well for resolving optimization issues. The NSL-KDD dataset was used to train and evaluate the classifiers in this work, which include Naïve Bayes, k-Nearest Neighbor, Decision Tree, and Support Vector Machine.

Khairnar et al,[18] presents an Aquila Optimization Algorithm-based Feature Selection with Optimal ML (AOAFS-OML) method for detecting DDoS attacks in an IoT environment. The presented technique employs a multifaceted technology, beginning with feature selection using the Aquila optimizer (AO), which aims to improve the relevance of input data for greater precision. For robust detection, this model employs the dominating XGBoost model, which is well-known for its ability to handle tough datasets. An innovative Artificial Rabbit Optimization (ARO) model is used to tune parameters and improve performance.

Chen et al,[19] presents a more thorough metaheuristic algorithm known as GQBWSSA, which is an improved version of the Salp Swarm Algorithm with various strategy enhancements. Using this technique, a threshold voting-based feature selection framework is created to produce an optimum collection of features. This approach easily reduces the number of dimensions in the data, preventing the negative impacts of having a large number of dimensions while effectively extracting the most relevant and critical information. The retrieved feature data is then merged with the LightGBM algorithm to create a lightweight and efficient ensemble learning strategy for IoT attack detection.

The proposed method

This paper applies the honey badger algorithm (HBA) with the algorithm of machine learning known as HBIDS. The present strategy is given the make an intrusion detection system (IDS) for meeting controlled area needs and could recognize between normal and attack traffic. Moreover, HBIDS chooses the most related features from a basic dataset of IDS which could aid in recognizing normal low-speed DDoS attacks, the chosen features are conveyed to classifiers like decision trees, naïve Bayes, multilayer perceptron, and support vector machines for identifying attack kind, Fig. 1 shows the modules of this method.

Datasets

This paper applied two sets of data CIC-IDS 2017 as the latest set of data and KDD Cup 99 as the old dataset for the presented strategy fair comparative assessment. The basic reason behind such dataset choice is to prove the presented strategy's effectiveness in diagnosing attacks of DDoS. Among them, a dataset of CIC-IDS 2017 is the most prevalent and inclusive set of data for attack trends modern type recognition and is widely fair in IDSs assessment while the KDD Cup 99 set of data includes old attack trends however this is yet applied frequently by an investigator.

Gathering data

Gathering data refers to the preliminary and perilous stage based on intrusion detection. This is the info-collection technique in a systematic path. For success and in the IDSs model, two assigning

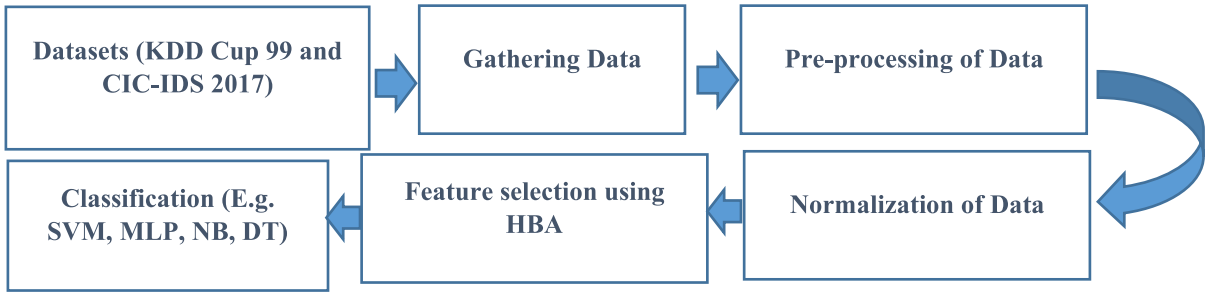

Fig. 1. Block diagram of the proposed method.

agents known as location and source of data play an important role in where data are collected. For proposing the most appropriate security for under network/attack host, this paper aims for an IDS based on the network for the presented model analysis.

Presented IDS runs on a nearby victim router and supervises the incoming network traffic. In the step of training, gathered data records are grouped concerning internet protocols/transport layer and are compared with area info. Conversely, data collected in the step of testing are grouped just conferring to protocol kind.

Pre-processing of data

Preprocessing of data is a time-consuming and needed step in the area of data mining. Data is collected from different platforms and has noisy, unstable, not completed, and extra features. So, this is important to transfer raw data in a format suitable for investigation. Here, the stage of preprocessing contains outliers and extra examples of removal with data conveying. Sets of data are used here, including binary, numeric, and nominal values. Since different classifiers proceed just with numerical values, the process of data conveying is deliberated essential and has a noteworthy effect on IDS accuracy. The nominal technique of binary is used here for conveying whole nominal features in binary numeric features.

It helps avoid hidden bias in consecutive data which impacts the usual profile and attack identification creation via our strategy.

Normalization of data

Here, for example, a dataset of KDD Cup 99 includes 5 dis-similar DoS attack types (Pod, tear, Smurf, Neptune attacks, ground) and there are 3 valid traffic types (UDP traffic, TCP, ICMP) that exist for investigations in tests. This paper just concentrates on the attack of DDoS which contains nominal, float, binary, and integer records. Before the tests, a requisite exists for controlling divided features computing their values frequency and altering them in numerical features, and also transferring whole features to normalized shape to eliminate bias in favor of higher-valued features from a set of data. every feature in every instance is normalized by a huge amount and recognized in similar intervals as $[-1, +1]$. A similar process is used in data testing.

Feature selection using Honey Badger algorithm

The Honey Badger method (HBA) is a nature-inspired optimization method that mimics honey badger foraging behavior. These animals are well-known for their persistence, adaptability, and ability to use various techniques to get food, making them an effective metaphor for addressing optimization problems. The HBA method's flexibility to flip between exploration and exploitation aids in avoiding local optima, a typical problem in feature selection in which the algorithm may settle on a poor subset of features. By picking the most relevant features, HBA improves the performance of the IDS classifiers, resulting in higher detection rates and fewer false positives.

The Honey Badger algorithm (HBA)[20] mimics the honey badger foraging manner. To place the resources of food, honey badger makes a hole/smell/pursue honeyguide birds. The first term is known as the mode of digging mode, and the second one is known as the mode of honey. In the previous mode, that applies the capability of smelling to becoming close to the situation of prey; during arrival, that goes across prey to choose a suitable location to make a hole and get the prey. In recent mode, honey badgers take the advice of honeyguide birds for straightly placing beehives.

Each honey badger in the population represents a possible subset of features. The position of a honey badger in the search space is represented by a binary vector, with each element indicating whether a

certain feature is selected (1) or not (0). The quality of each solution (feature subset) is assessed using an objective function, which often includes a classifier (such as SVM, Decision Trees, etc.) trained on the chosen features. The goal function may combine classification accuracy, True Positive Rate, and the amount of features to provide a fitness score.

The honey badgers explore the search space by "digging," which involves updating their positions based on the intensity (quality) of the solutions and their distance from the best-known solution (prey). This helps in discovering diverse feature subsets that might be overlooked if the search is too focused (i.e., exploitation-heavy). The digging behavior is mathematically simulated using a cardioid function that incorporates the distance between the badger and the prey, random factors, and the intensity of the solution, allowing the algorithm to explore new regions of the feature space. Honey badgers refine their search by utilizing the most well-known options. They proceed closer to the most promising feature subsets detected during the digging phase, led by the prey's position (the best answer found thus far). Randomization variables and the density agent impact the movement towards better solutions, which lowers with time to reduce unpredictability and concentrate the search on the most promising locations.

Throughout the iterations, the algorithm updates the optimal solution as it identifies feature subsets that produce improved classification results. The algorithm avoids becoming stuck in local optima by periodically exploring new regions of the search space (by the digging phase) and refining the best answers (during the honey phase).

Mathematical model

According to the previous discussion, HBA is shared in 2 steps which are the "digging" and the "honey", stated in detail:

Algorithmic steps

The mentioned part defines the mathematical presented HBA algorithm formulation. The theoretical order, HBA is equipped with two steps of exploration and exploitation and, therefore could be referred to as the universal mechanism of optimization. HBA algorithm contains crowd initialization, crowd assessment, and updating parameters. Mathematically, presented HBA stages are detailed below. Now, voluntary solutions in the HBA[16] crowd are shown as in Eq. (1):

$$\begin{matrix} X_{11} & X_{12}\, X_{13} & \cdots & X_{1D} \\ \vdots & \ddots & & \vdots \\ X_{n1} & X_{n2}\, X_{n3} & \cdots & X_{nD} \end{matrix} \quad (1)$$

Voluntary solutions population $= x_i = [x_i^1 . x_i^1 . \cdots x_i^D]$

Step 1: A step of initialization. Initialize honey badgers value (size N crowd), respective locations of them given the Eq. (2):

$$\mathrm{x_i} = \mathrm{lb_s} + \mathrm{r_1} \times \left(\mathrm{Ub_i} - \mathrm{lb_i}\right), \quad (2)$$

r_1 refers to a random amount among 0 and 1 that x_i is the ith situation of honey badger referring to the voluntary solution in N crowd when lb_i and ub_i are respectively lower and upper search field limits.

Step 2: Describing intensity (I). Intensity is related to with the strength of prey concentration, the distance among the honey badger, and that. That represents the intensity of prey smell; while the smell is high, the movement would be quick and vs. that is provided through Inverse Square Law and is described in Eq. (3):

$$I_i = r_2 \times \frac{S}{4\pi d_i^2}. r_2 \text{ is random number between 0 and 1}$$

$$S = (x_i - x_{i+1})^2 \; d_i = x_{prey} - x_i \quad (3)$$

That S stands for strength of focus and resource. In Eq. (3), d_i indicated the distance between the prey and ith badger.

Step 3: Update the agent of density. Time-varying randomization is handled by an agent of density (α) to guarantee a smooth transition from exploring to exploiting.

Apply Eq. (4) to update reduction agent α, which decreases with iterations to reduce randomness over time:

$$\alpha = C \times \exp\left(\frac{-t}{t_{max}}\right). t_{max}$$

$$= \text{maximum number of iterations} \quad (4)$$

That C is the constant ≥ 1 (default $= 2$).

Step 4: Evading the local optimum. The mentioned stage and the two next stages are applied for escaping from local optima regions. Here, this strategy utilizes flag F that notices the direction of surfing to avail high chances for agents scanning space of search rigorously.

Step 5: Agents' situation updating. As previously mentioned, the update process of the HBA situation is shared in two sections that are "phase of digging" and "honey". Below is a better explanation:

Step 5-1: Step of digging. During the step of digging, honey badger works like a cardioid. Eq. (5) could be used to simulate the movement of the cardioid:

$$x_{new} = x_{prey} + F \times \beta \times I \times x_{prey} + F \times r_3 \times \alpha \times d_i \times |\cos(2\pi r_4) \times [1 - \cos(2\pi r_5)]| \quad (5)$$

That x_{prey} is the prey location that is the best situation recognized up to now– in other words, the universe's best situation. $\beta \geq 1$ (default = 6) refers to the honey badger's capability for catching food. According to Eq. (2), d_i shows the distance between the honey badger and its prey. The three random numbers, r_3, r_4, and r_5, range from 0 to 1. F acts as a flag that notices direct surfing which is assigned by Eq. (6):

$$F = \begin{cases} 1 & if\ r_6 \leq 0.5 \quad r_6\ is\ a\ random\ number \\ & between\ 0\ and\ 1 \\ -1 & else \end{cases} \quad (6)$$

In the digging step, the honey badger depends on the distance between the badger and prey d_i, prey x smell intensity I, and time-differing surfing impact agent α. Furthermore, during the task of digging, the badger may acquire each disturbance F that lets it recognize better positions of prey.

Step 5-2: Honey step. A term while chases a honeyguide bird in an attempt to obtain a beehive could be simulated as Eq. (7):

$$x_{new} = x_{prey} + F \times r_7 \times \alpha \times d_i.\ r_7\ is\ a\ random\ number\ between\ 0\ and\ 1 \quad (7)$$

That recent one refers to a novel honey badger situation, while F and α are assigned by respectively applying Eq. (5) and Eq. (6), and x prey is the situation of prey. From Eq. (7), it could be noticed that, the honey badger carries out the surfing approximately to the situation of prey x recognized up to now, given the info of distance d_i. here, surfing is impacted through the manner of surfing differing by time (α). In addition, honey badgers may be able to identify disturbance F.

In IDS contexts, where datasets may contain a huge number of features, HBA's ability to efficiently search and choose features is critical. By focusing on the most relevant features, HBA minimizes data dimensionality, which not only speeds up classifier training and evaluation but also improves performance by reducing irrelevant or duplicated information. The performance of HBA can be affected by the parameters selected, such as the population size, density agent constants, and randomization variables. If these parameters are not carefully set, the algorithm may not function optimally, resulting in inferior feature selection or slower convergence times.

Techniques of classification

Support vector machine (SVM)

SVM is one of the classical methods of ML that could yet aid in solving issues of large data classification. Particularly, this could aid multidomain applications in large big data areas. Although, mathematically SVM is complicated and computationally expensive.[21] Here, the hyperplane is shaped to distinguish between negative and positive examples taking structural risk minimization law. The predicted SVM feature is that instead of the entire dataset, this performs classification using support vectors, so it is widely robust for outliers and efficiently predicts. M shows points of training data $\{(x_1, y_1), (x_2, y_2), \ldots, (x_M, y_M)\}$ that $x_i \in Rd$ and $y_i \in \{+1, -1\}$. Each point of data includes the allied Lagrangian multiplier βi assigning a comparative weight. Class prediction for data point x is shown below in Eq. (8) and the hyperplane is assigned as (w, c).

$$F(x) = sgn(w.k(x.x_i) - c) = \text{sgn}\left\{\sum_{i=1}^{M} \beta_i y_i e^{\left(x - \frac{x_i^2}{2\sigma^2}\right)} - c\right\} \quad (8)$$

That sgn shows the function of sign, $k(\bullet, \bullet)$ illustrates the function of radial basis kernel, x illustrates the point of data in space of input, w shows weight, c illustrates bias, and σ shows the standard deviation. While the hyperplane is well-defined, whole points located nearby include $\beta i > 0$ known as support vectors. Moreover, residual points show $\beta i = 0$. The SVM has efficiently carried out to resolve problems in real scenarios including bioinformatics and image classification.

Multi-layer perceptron

The Multilayer Perceptron Neural Network refers to the feed-forward kind of Neural Network. This applies the Backpropagation method to learning. This has neurons' input layers which act as receivers, one/more neurons' hidden layers which calculate data and undergo iterations, and an output layer

which predicts output.[22] This technique's main strategy refers to mapping a lot of real-valued inputs in results altering weight amongst the internal nodes. The MLP needs the function $f(x)$: $R_i \rightarrow R_t$ in training the dataset, using the learning strategy of back-propagation that $i,t \in Q+$ individually shows input and output dimensions. They are listed in Eq. (9):

$$y = \delta\{\sum_{i=1}^{m}(w_i X + b)\} = \delta\left(W^T X + b\right) \quad (9)$$

That δ illustrates the function of activation, w shows vectors of weight, X shows vectors of input, and b signifies bias. Specifically, neural classifiers have been widely conducted in some domains such as prediction, pattern classification, and recognition.

Naïve bayes

The classifier of Naïve Bayes (NB) has played a prominent role due to its efficiency, simplicity, and tractability. Implicit independent features' assumption located in class eases NB implementation considerably, as this lets sample likelihood decomposition in univariate marginal product.[23] Additionally, normally NB computes fewer parameters than other renowned classifiers, therefore this is less prone to overfitting. The main strategy of NB classification relies on the law theorem of Bayes to look for the maximum feasibility hypothesis that identifies the class label. This is employed for predicting points of data using the max back function computed by the Eq. (10):

$$P((L|O) = arg_{w\in(1.2\ldots M)} \max P\left(L_w\right) \prod_{j=1}^{M} P\left(\left(O_j|L_w\right)\right) \quad (10)$$

That L shows the label of the class, O shows each level observation, w shows several classes, P(L|O) shows presented class probability and $\prod_{j=1}^{M} P((O_j|L_w))$ shows the whole samples' probabilities multiplication aimed in their levels for achieving max result.

Decision tree

A decision tree is a method based on the tree where each way starting from a root is defined by data separating order till the result of Boolean at the node of the leaf is obtained. This is hierarchical knowledge relationships exemplification which includes nodes and links. While relations are applied for grouping, nodes show aims.[24] Sheets specify laws of classification, while branches show features that lead to data classification. Laws of decision are assigned by if–then laws for data entry classification. Such laws are particularly used for solving different issues of feature choice (like, strings, and actual numbers). This relies on dividing and conquering models to create a decision tree. As illustrated in Eq. (11) and Eq. (12), the Gain Ratio is a decision tree measure for computing performance. Consider, a training dataset exists as D, the expected info needed for appropriately grouping the attack in IDS, $x_j \in D$ that is listed as:

$$Entropy = -\sum_{j=1}^{m} p_j log\left(p_j\right) \quad (11)$$

$$GainRatio = \frac{Gain\,(p)}{SplitInfo\,(p)} \quad (12)$$

The function of $SplitInfo$ is listed in Eq. (13) below:

$$SplitInfo\,(p.test) = -\sum_{j=1}^{m} p\left(\frac{j}{p}\right) log\left(p\left(\frac{j}{p}\right)\right) \quad (13)$$

That p shows data probability share while log applies base as 2 because of estimate info as bit.

Results and discussion

Experimental setup

Whole tests here are performed applying the area of MATLAB R2020b under the Windows 10 operating system with 2.4 GHz Pentium Core i7, and 16 GB of RAM. Whole optimization strategies' parameters are considered as population size is 50, max iterations number of is 100.

Datasets description

This paper applied two sets of data CIC-IDS 2017 as the latest set of data and KDD Cup 99.

The dataset of KDD Cup 99 was obtained from the dataset of DARPA 98 created from the 1998 DARPA intrusion diagnosis assessment plan. This includes 5 various levels that are usual and 4 attack kinds however this paper just concentrates on DoS service attacks. This includes 4898,430 records of links having 41 features. Eliminating duplicate records, left traffic includes 812,813 usual records of link and 247,267 DoS records of link. Among these, 300,000 records including both normal and DoS attack traffic were applied to train the presented strategy and new methods.[25]

The dataset of CIC-IDS 2017 created by the Canadian Institute for Cyber Security contains benign and

Table 1. Proposed settings of the parameter for the honey badger algorithm.

Parameter	Values
Number of Iterations	100
Training dataset	80%
Test dataset	20%
Population size	50
The lower bound of the search interval	0
The upper bound of the search interval	1

the newest typical attacks. Present sets of data that are accessible generally have the traffic diversity absence, amounts, limitations on different attacks, anonymized packet info payload, metadata absence, and set of features. Here, 6 attack types exist however this paper is just concentrating on DDoS Attacks. CIC-IDS 2017 includes 128,027 records for DDoS attacks where 2700 records are to train and 3300 records to test including 79 features.[26]

Both datasets of CIC-IDS 2017 and KDD Cup 99 are selected for the unbiased presented method assessment. Applied datasets merit over the other IDS sets of data are first, this does not contain extra records in a set of the train, thus classifiers could not be biased to more often records. Secondly, the chosen record number from every hard class set is inversely proportional to the records' percent in basic datasets of IDS. so, distinct ML techniques of classification rates differ in a broader range making this more effective to have the appropriate various learning methods' assessment. Thirdly, record numbers in groups of tests and training are fair making this affordable to execute tests on full groups with no requirement to randomly choose a small portion. Accordingly, various study work assessment outcomes could be stable and comparable.

Settings of the parameter

The list of parameters, as shown in Table 1 was used for the proposed method algorithm.

Measures of performance

This paper computes strategy performance by applying seven metrics of performance, like specificity, accuracy, f-measure, sensitivity, AUC, precision, and FPR (domain under the curve). Such metrics of performance are assigned as in Eqs. (14) to (20):

$$Accuracy = \frac{TN + TP}{TP + TN + FP + FN} \tag{14}$$

$$Recall\ (Re) = Sensitivity = Detection\ Rate\ (DR) = \frac{TP}{TP + FN} \tag{15}$$

$$Precision\ \ (\mathrm{Pr}) = \frac{TP}{TP + FP} \tag{16}$$

$$F - measure = \frac{2 * Pr * Re}{Pr + Re} \tag{17}$$

$$False\ positive\ rate = \frac{FP}{FP + TN} \tag{18}$$

$$Specificity = 1 - FPR \tag{19}$$

$$AUC = \frac{Sensitivity + Specificity}{2} \tag{20}$$

In this case, FN, FP, TN, and TP are false–negative, false–positive, true-negative, and true-positive, in independent sets of data. whole classification and performance assessment tests are carried out by applying the Matlab tool.

Simulation outcomes and discussion

The presented strategy's classification on a dataset of KDD Cup 99 is assessed by applying 10 fold-CV strategies. The dataset of KDD Cup 99 has been widely employed in the IDS area and yet used in some recent experiments. This has been adopted for computing anomalies-based IDS performance in novel intrusions' diagnosis. HBA-based IDS technique (HBIDS) outcomes of Simulation are analyzed for assigning the effectiveness in DDoS attacks diagnosis; outcomes are achieved compared to other methods based on wrapper outcomes like HBA, GA, PSO with various classifiers like SVM, MLP, C4.5, NB for comparing IDS performance based on HBA in DDoS attack recognition.

This paper applies repeated 10 times to ensure that the classifier generalizes well for the unpredicted data and increases IDS performance. As it is obvious from Table 2 and Table 3, this paper observed that HBA with C4.5 outperforms the other method obtaining various FS methods based on wrapper comparative analysis with various classifiers with 99.38% accuracy, 99.16% detection rate, 99.65% precision in the KDD Cup 99 dataset while 99.9% accuracy, 99.99% detection rate, 99.78% precision in CIC-IDS 2017 in DDoS attack diagnosis.

Table 2. Presented method comparison with other methods on a dataset of KDD Cup 99.

Measures	Algorithm optimization	MLP	NB	SVM	DT
Accuracy	GA	72.82	48.63	76.31	99.02
Detection rate		71.75	89.23	77.26	99.15
Precision		71.6	48.23	75.54	98.25
Accuracy	PSO	90.52	88.63	96.15	98.15
Detection rate		89.15	87.13	96.32	97.61
Precision		90.29	86.61	97.32	96.84
Accuracy	HBA	78.47	52.52	77.30	99.38
Detection rate		80.82	1	80.52	99.16
Precision		78.37	52.22	76.85	99.65

Table 3. Presented method comparison with other methods on a dataset of CIC-IDS 2017.

Measures	Algorithm optimization	MLP	NB	SVM	DT
Accuracy	GA	85.12	70.23	75.61	96.52
Detection rate		64.45	91.41	96.27	96.35
Precision		94.68	61.13	66.49	95.89
Accuracy	PSO	95.58	93.37	96.53	97.46
Detection rate		91.63	92.74	95.56	97.29
Precision		92.74	92.28	95.64	97.32
Accuracy	HBA	86.84	78.57	78.36	99.9
Detection rate		69.74	98.27	97.73	99.99
Precision		99.80	67.28	67.19	99.78

Table 4 and Table 5, the presented method performance given the chosen features assigned considering FPR, AUC, and F -measure with various other methods based on the wrapper. Given the obtained results, the presented method validates the efficacy in DDoS attack diagnosis given the metrics of performance respectively 0.015% FPR, 98.85% F-measure, 98.79% AUC, 0.001% FPR, 99.89% F-measure, 99.91% AUC in KDD Cup 99 dataset and CIC-IDS 2017.

In IDS, selecting the most relevant features is critical for accurately distinguishing between normal and malicious traffic. The HBA's exploration-exploitation mechanisms help in identifying the most informative features, reducing dimensionality without losing

Table 4. Techniques' assessment performance on a dataset of KDD Cup 99.

Measures	Algorithm optimization	MLP	NB	SVM	DT
FPR	GA	10.03	13.4	12.2	13.02
F-measure		82.45	63.52	78.5	98.25
AUC		81.37	58.9	77.43	98.11
FPR	PSO	11.41	18.54	18.41	13.2
F-measure		74.21	65.41	78.72	98.55
AUC		73.62	59.79	77.42	98.14
FPR	HBA	0.1827	0.9869	0.1827	0.015
F-measure		82.91	68.45	82.91	98.85
AUC		82.27	50.48	82.27	98.79

Table 5. Techniques' assessment performance on a dataset of CIC-IDS 2017.

Measures	Algorithm optimization	MLP	NB	SVM	DT
FPR	GA	9.27	4.23	3.46	2.42
F-measure		81.45	75.46	74.21	94.47
AUC		82.24	79.48	75.34	96.23
FPR	PSO	10.21	4.24	3.32	1.53
F-measure		79.17	77.42	70.87	95.78
AUC		72.84	80.44	75.32	96.42
FPR	HBA	0.001	0.3648	0.3642	0.001
F-measure		82.1	79.88	79.64	99.89
AUC		84.82	80.9	80.66	99.91

Table 6. Presented method comparison with new methods on both sets of data.

Dataset	Method	DR	Accuracy
KDD Cup 99	SVM, MLP, DT, NB[11]	98.80	99.25
	Proposed method	99.16	99.38
CIC-IDS 2017	SVM, K-NN, DT, NB, RF and LR[8]	99	99
	Bi-LSTM[16]	–	97
	Proposed method	99.99	99.9

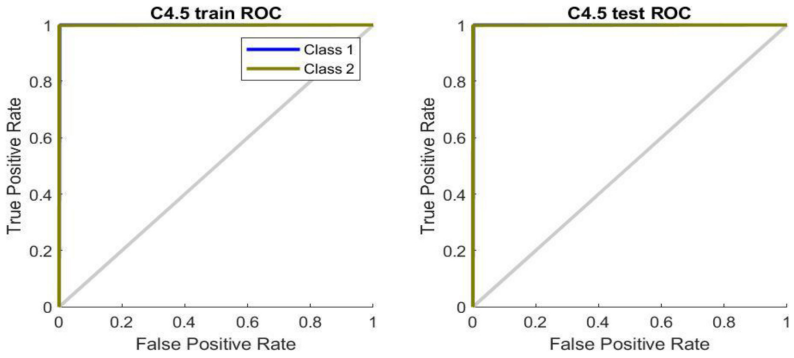

Fig. 2. Curve of ROC for presented technique.

important information. This leads to better classification performance by the subsequent machine learning classifiers (e.g., SVM, MLP, Naïve Bayes, Decision Trees), improving the overall accuracy and efficiency of the IDS. As illustrated in Table 6, outcomes of comparison illustrate that the presented technique has obtained important outcomes in comparison with prevailing new ones such as IDS wrapper and filter methods discussed in the last works, with high accuracy and confirmed its effectiveness on the applied sets of data.

The curve of receiver operating characteristics (ROC) for the presented algorithm is illustrated in Fig. 2. ROC shows the presented algorithm performance including true and false positive rates, the presented algorithm provided an ROC of 0.99. The chart provides important information about the trade-off between the TPR and FPR rates over different numbers of epochs and categorization criteria. It displays how well the suggested model predicts things when it comes to classifying different classes. The Honey Badger Algorithm indicates in Fig. 2, combined with the machine learning classifiers, is effective in detecting intrusions with minimal false alarms.

Results of classification for a scheme of performance are illustrated on matrices of confusion in Fig. 3. Model predictions are illustrated on the X-axis and pathology detections are illustrated on the Y-axis.

When assessing the detection accuracy rate, the objective function's benefits are contingent upon the feature that is chosen. Therefore, if the ideal solution

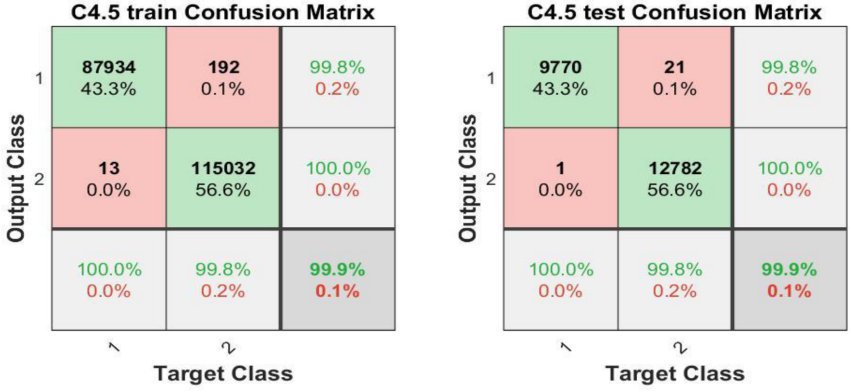

Fig. 3. Matrices of confusion.

for a big subset includes some features, then more work and resources are needed. The ideal solution that aids in optimal subset selection is chosen by the accuracy rate of the present process.

Conclusion

The attack of Distributed Denial of Service (DDoS) happens when big traffic numbers from hundreds, thousands/millions of other computers are routed to the server/ network to crash the system and disrupt the function. Diagnosing attacks of DDoS is a hard function that should be obtained before each mitigation approach can be applied. DDoS attack recognition has been successfully performed by applying machine learning/deep learning (ML/DL). However, because of the ML/DL framework's inherent restriction, full accomplishment is likewise out of access. It is the term where an ML/DL-based system does not generate satisfactory outcomes to recognize attacks of DDoS. At this time, present work on the prediction of DDoS attacks has shown the unexpected prediction diversity using classifiers of ML and conventional strategies to encode a feature and in this paper, presented IDS based on HBA known as HBIDS that evolves programs of computers for attack detection. Honey badger optimization is utilized for optimal features. The present strategy is carried out in two phases: first, basically, features are chosen in HBA and chosen features are conveyed to various classifiers SVM, NB, MLP, and C4.5 for grouping the attack of DDoS. This is recognized that the method of HBA with the classifier of C4.5 proposes better performance than other prevailing methods regarding FPR, accuracy, and detection for recognizing attacks of DDoS. The suggested model is also investigated with respect to several of the current methodologies. The proposed model attained 99.9% accuracy, 99.99% detection rate, and so on when the performance was assessed using the current methodology. The suggested methodology works well for locating DDoS in network. In future, the suggested method may be used to other common datasets, such as UNSW-NB15 databases, and the results obtained from these datasets can be contrasted with those obtained from CIADA datasets.

Authors' declaration

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are ours. Furthermore, any Figures and images that are not ours have been included with the necessary permission for republication, which is attached to the manuscript.
- No animal studies are present in the manuscript.
- No human studies are present in the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee at University of Wasit.

Authors' contribution statement

M. A. Al. designed the study, organized field sampling, and simulation work, and wrote the manuscript. A. R. Al. contributed to field sampling, simulation work, result analysis, and manuscript preparation.

References

1. Ojajuni OP, Ismail Y, Lawson A. Distributed denial-of-service attack detection and mitigation for the Internet of Things. IJTD. 2020;11(2):18–32. https://doi.org/10.4018/IJTD.2020040102.
2. Sarma SK. Hybrid optimised deep learning belief network for attack detection in the Internet of things. J Exp heor Artif Intell. 2022;34(4):695–724. doi:https://doi.org/10.1080/0952813X.2021.1924868.
3. Zolfagharipour L, Kadhim MH, Mandeel TH. Enhance the Security of Access to IoT-based Equipment in Fog. AICCIT. 2023;142–146. https://doi.org/10.1109/AICCIT57614.2023.10218280.
4. Gopi R, Sathiyamoorthi V, Selvakumar S, Manikandan R, Chatterjee P, Jhanjhi NZ, *et al.* Enhanced method of ANN based model for detection of DDoS attacks on multimedia internet of things. Multimed Tools Appl. 2021;1–9. https://doi.org/10.1007/s11042-021-10640-6.
5. He K, Kim DD, Asghar MR. Adversarial machine learning for network intrusion detection systems: a comprehensive survey. IEEE Commun Surv Tutor. 2023. https://doi.org/10.1109/COMST.2022.3233793.
6. Nayak S, Pradhan S, Mishra P, Biswal P. An upgraded machine learning approach for glaucoma detection in retinal images using optimized ensemble feature selection and Weighted-Sum Salp Swarm Algorithm. Biomedical Signal Processing and Control. 2026 May 15;117:109567. https://doi.org/10.1016/j.bspc.2026.109567.
7. Gül MF, Bakır H. A novel metaheuristic-enhanced quantum-classical neural network for attack detection in agriculture IoT systems. The Journal of Supercomputing. 2026 Jan;82(2):44. https://doi.org/10.1007/s11227-025-08118-5.
8. Alzahrani RJ, Alzahrani A.Security analysis of ddos attacks using machine learning algorithms in networks traffic. ELEC. 2021;10(23):1–15. https://doi.org/10.3390/electronics10232919.
9. Karthik MG, Krishnan MM. Securing an internet of things from distributed denial of service and Mirai botnet attacks using a novel hybrid detection and mitigation mechanism. Int J Intell Eng Syst. 2021;14:113–123. http://dx.doi.org/10.22266/ijies2021.0228.12.
10. Sambangi S, Gondi L. A machine learning approach for ddos (distributed denial of service) attack detection using multiple linear regression. Proceedings. 2020;63(1):1–12. https://doi.org/10.3390/proceedings2020063051.

11. Dwivedi S, Vardhan M, Tripathi S. Distributed denial-of-service prediction on IoT framework by learning techniques. Open Comput Sci. 2020;10(1):220–230. https://doi.org/10.1515/comp-2020-0009/html.
12. Asghari Varzaneh Z, Hosseini S, Javidi MM. A novel binary horse herd optimization algorithm for feature selection problem. Multimed. Tools Appl. 2023;82(26):1–35. https://doi.org/10.1007/s11042-023-15023-7.
13. Aljebreen M, Mengash HA, Arasi MA, Aljameel SS, Salama AS, Hamza MA. Enhancing DDoS attack detection using snake optimizer with ensemble learning on internet of things environment. IEEE Access. 2023;11:104745–104753. https://doi.org/10.1109/ACCESS.2023.3318316.
14. AIghuraibawi AH, Manickam S, Abdullah R, Alyasseri ZA, Jasim HM, Sani NS. Modified Flower Pollination Algorithm for ICMPv6-Based DDoS Attacks Anomaly Detection. Procedia Comput Sci. 2023;220:776–781. https://doi.org/10.1016/j.procs.2023.03.103.
15. Dwivedi S, Vardhan M, Tripathi S. Defense against distributed DoS attack detection by using intelligent evolutionary algorithm. Int J Comput Appl. 2022;44(3):219–29. https://doi.org/10.1080/1206212X.2020.1720951.
16. Pandithurai O, Venkataiah C, Tiwari S, Ramanjaneyulu N. DDoS attack prediction using a honey badger optimization algorithm based feature selection and Bi-LSTM in cloud environment. Expert Syst Appl. 2024;241:122544. https://doi.org/10.1016/j.eswa.2023.122544.
17. Mahmood RA, Abdi A, Hussin M. Performance evaluation of intrusion detection system using selected features and machine learning classifiers. Baghdad Sci J. 2021;18(2):884–898. http://dx.doi.org/10.21123/bsj.2021.18.2(Suppl.).0884.
18. Khairnar PN, Vidhya SS, Keerthana D, Mohana M, Srimathi S, Vigneshwaran P. Aquila Optimization Algorithm based Feature Selection with Optimal Machine Learning for Security Internet of Things Environment. ICKECS. 2024;18(1):1–6. IEEE. https://doi.org/10.1109/ICKECS61492.2024.10616777.
19. Chen W, Yang H, Yin L, Luo X. Large-scale IoT attack detection scheme based on LightGBM and feature selection using an improved salp swarm algorithm. Sci Rep. 2024;14(1):19165. https://doi.org/10.1038/s41598-024-69968-2.
20. Hashim FA, Houssein EH, Hussain K, Mabrouk MS, Al-Atabany W. Honey Badger Algorithm: New metaheuristic algorithm for solving optimization problems. Math Comput Simul. 2022;192:84–110. https://doi.org/10.1016/j.matcom.2021.08.013.
21. Suthaharan S, Suthaharan S. Support vector machine. Machine learning models and algorithms for big data classification: thinking with examples for effective learning. Springer. 2016:207–235. https://doi.org/10.1007/978-1-4899-7641-3.
22. Desai M, Shah M. An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN). Clin eHealth. 2021;4:1–11. https://doi.org/10.1016/j.ceh.2020.11.002.
23. Blanquero R, Carrizosa E, Ramírez-Cobo P, Sillero-Denamiel MR. Variable selection for Naïve Bayes classification. Comput. Oper Res. 2021;135:1–11. https://doi.org/10.1016/j.cor.2021.105456.
24. Charbuty B, Abdulazeez A. Classification based on decision tree algorithm for machine learning. JASTT. 2021;2(01):8–20. https://doi.org/10.38094/jastt20165.
25. Anuradha D, Gupta G. A self-explanatory review of decision tree classifiers. ICRAIE. 2014;1–7. https://doi.org10.1109/ICRAIE.2014.6909245.
26. Gavankar SS, Sawarkar SD. Eager decision tree. I2CT. 2017;837–840. https://doi.org/10.1109/I2CT.2017.8226246.

خوارزمية غرير العسل لاختيار الميزات للكشف عن هجمات رفض الخدمة في شبكات إنترنت الأشياء

مصطفى عزيز المياحي ، احمد رعد السوداني

القسم البرامجيات، كلية علوم الحاسوب وتكنولوجيا المعلومات، جامعة واسط، واسط، العراق.

الخلاصة

مع التطور السريع لتكنولوجيا الحاسوب والاتصالات، تزداد أهمية أضرار هجمات رفض الخدمة الموزعة (DDoS). يعتبر بحث هجمات DDoS مجالًا دراسيًا مهمًا؛ حيث تم تقديم عدد من الطرق في الأدبيات لتشخيص هجمات DDoS، مثل الخوارزمية التطورية والذكاء الاصطناعي. لسوء الحظ، فإن النماذج الشائعة الجديدة لتشخيص DDoS تتدهور في التحقق من هدف هجمات DDoS والتعرف عليها مسبقًا. نظرًا لتنوع طرق هجمات DDoS ومقدار حركة المرور المختلفة للهجمات، لا توجد حاليًا تقنية تشخيص ذات دقة واعدة في التشخيص. من خلال اختيار أفضل مجموعة فرعية من الميزات، يساعد نهج اختيار الميزات (FS) على تقليل أوقات الحوسبة وزيادة تعقيد الحوسبة. لتخفيف هجمات رفض الخدمة، تستخدم هذه الورقة خوارزمية النمل البري (HBA) مع خوارزمية التعلم الآلي المعروفة باسم HBIDS. يتم تقديم الاستراتيجية الحالية لإنشاء نظام اكتشاف التسلل (IDS) لتلبية احتياجات المنطقة المحكومة والتمييز بين حركة المرور العادية والهجومية. علاوة على ذلك، يختار HBIDS الأكثر ارتباطًا من مجموعة البيانات الأساسية لـ IDS والتي يمكن أن تساعد في التعرف على هجمات DDoS منخفضة السرعة العادية، ثم يتم نقل الميزات المختارة إلى التصنيفات مثل شجرة القرار، والشبكة العصبية متعددة الطبقات، وبايز الساذج، وآلة الدعم المتجه لتحديد نوع الهجوم. يتم تطبيق مجموعة البيانات المتاحة عمومًا مثل CIC-IDS 2017 و KDD Cup 99 للبحث التجريبي لدينا. من نتائج المحاكاة، من الواضح أن HBIDS مع شجرة القرار يحتاج إلى تشخيص عالي مع معدل خطأ إيجابي منخفض (0.001) ودقة (99.9).

الكلمات المفتاحية: رفض الخدمة الموزعة، خوارزمية غرير العسل، نظام الكشف عن التسلل، التعلم الآلي، تحليل المخاطر الأمنية.