electronics

Article

Comparing Metaheuristic Search Techniques in Addressing the Effectiveness of Clustering-Based DDoS Attack Detection Methods

Alireza Zeinalpour * and Charles P. McElroy

Department of Information Systems, Monte Ahuja College of Business, Cleveland State University, Cleveland, OH 44115, USA; c.p.mcelroy@csuohio.edu
* Correspondence: a.zeinalpour@csuohio.edu

Abstract: Distributed Denial of Service (DDoS) attacks have increased in frequency and sophistication over the last ten years. Part of the challenge of defending against such attacks requires the analysis of very large volumes of data. Metaheuristic algorithms can assist in selecting relevant features from the network traffic data for use in DDoS detection models. By efficiently exploring different combinations of features, these methods can identify subsets that are informative for distinguishing between normal and attack traffic. However, identifying an optimized solution in this area is an open research question. Tuning the parameters of metaheuristic search techniques in the optimization process is critical. In this study, a switching approximation is used in a variety of metaheuristic search techniques. This approximation is used to find the best solution for the analysis of the network traffic features in either lower or upper values between 0 and 1. We compare the fine-tuning of this parameter against standard approaches and find that it is not substantially better than the BestFirst algorithm (a standard default approach for feature selection). This study contributes to the literature by testing and eliminating various fine-tuning strategies for the metaheuristic approach.

Keywords: DDoS attack; wrapper method; hybrid approach; metaheuristics; clustering method

Citation: Zeinalpour, A.; McElroy, C.P. Comparing Metaheuristic Search Techniques in Addressing the Effectiveness of Clustering-Based DDoS Attack Detection Methods. *Electronics* **2024**, *13*, 899. https://doi.org/10.3390/electronics13050899

Academic Editor: Martin Reisslein

Received: 9 January 2024
Revised: 19 February 2024
Accepted: 24 February 2024
Published: 27 February 2024

1. Introduction

Distributed Denial of Service (DDoS) attacks exist as one of the greatest threats to the Internet [1]. These attacks consume network resources by preventing the provision of regular services [2]. In the case of a DDoS attack, an attacker transmits a large number of network traffic packets through several computing devices. These attacks are intended to make the target server inaccessible by overwhelming its resources. These resources include the CPU, memory, and network bandwidth connections [1]. DDoS attacks have become commonplace and effective due to their efficiency and concealment [2]. There has been an increasing prevalence of DDoS attacks recently as hackers become more proficient in disabling whole systems and platforms [3,4].

DDoS attack detection methods can be effective in recognizing and neutralizing DDoS events [5]. Intrusion detection systems (IDS) play a major role in distinguishing between normal and abnormal network traffic activity [6]. One approach is based on anomaly detection [1]. An anomaly-based IDS is able to provide a novel capability by combining known attack profiles compared with typical network activity [7]. According to Idhammad et al. [8], anomaly detection methods are used to identify DDoS attacks, which are based on machine learning algorithms, e.g., supervised and unsupervised learning approaches, where the supervised methods depend on labeled data and unsupervised algorithms analyze unlabeled data, respectively.

The clustering-based DDoS attack detection method is an unsupervised approach that suffers from the curse of dimensionality [5]. Here machine learning algorithms analyzing network traffic data are overwhelmed by the computational load due to the astronomically large number of features present in the data [9]. Various strategies to enhance feature selection have been proven to improve the performance of machine learning models through the elimination of irrelevant and duplicate data [10]. The filter and wrapper methods are commonly used to perform feature selection [7]. The filter method uses statistical measures while the wrapper method applies learning models to assess the relevancy of the features [10].

Zeinalpour and Ahmed [9] found that the addition of the filter method prior to applying the clustering algorithms did not lead to better detection of DDoS attacks, but that the addition of the wrapper method alone was shown to be a superior approach. In previous work, the addition of the wrapper method formed a hybrid approach through its incorporation into the process after the filter method. In this study, we consider the same effectiveness of the hybrid approach in contrast to relying solely on the wrapper method. Meta-heuristic techniques are inspired by nature as optimization methods that are applied to complex search scenarios. They are used in realizing a suitable solution [11].

Cuckoo Search Algorithm (CSA), Flower Pollination Algorithm (FPA), and Firefly Search Algorithm (FSA) are meta-heuristic techniques that are examples of this type of approach. CSA, FPA, and FSA apply an approximation in finding the best solution in either lower or upper given values by a user that is between 0 and 1. This approximation is referred to as a switching probability in metaheuristic techniques. In a study by Palaniswamy and Kandhasamy [12], they found that the upper approximation may exist between 0.7 and 0.9, and the lower approximations may exist between 0.1 and 0.3. The switching probability or approximation in CSA is the discovery rate, in FPA, it is the pollination approximation, and in the FSA, it is the distance attractiveness approximation of a firefly. The discovery approximation in CSA is represented by the '*Pa*' parameter. The pollination approximation in FPA is represented by the 'pollination' parameter. The distance attractiveness approximation of the firefly is represented by the 'betaMin' parameter of FSA.

The need for reliable network operations has been increasing due to our reliance on using network services [11]. Meta-heuristic techniques have been used to ameliorate feature selection problems in cybersecurity [10]. Up to this point, no study in the literature has been found to assess the impact of the switching approximation of metaheuristic search techniques on the effectiveness of clustering-based DDoS attack detection methods. This study investigated whether this type of fine-tuning could improve the performance of the techniques as opposed to a standard approach such as 'BestFirst'.

2. Materials and Methods

According to Saw and Oo [10], the optimal strategy to integrate the filter and wrapper methods is an open research question. To this purpose, we applied two experiments of an ex post facto design of an A-B single group. According to Tanious and Onghena [13], an A-B single group is a phase design in having baseline measurements followed by measuring independent variables through their manipulations in which 'A' reflects a baseline measure and 'B' reflects an experimental measure. In this regard, the baseline measures correspond to a control method, while the experimental measures correspond to the intervention. A phase design includes assessing various levels of independent variables in successive and uninterrupted phases [13]. In the first experiment, we compared the results obtained from the default search method of the wrapper method, known as 'BestFirst', against the use of metaheuristic search techniques. In the second experiment, we evaluated the results from using the hybrid approach of the filter–wrapper method considering 'BestFirst' against metaheuristic search techniques in the wrapper method. We considered the 'BestFirst' for our baseline measure and metaheuristic search techniques for our experimental intervention.

We used the available CICIDS2017 network traffic dataset. This dataset contains information from benign and DDoS attack network data [5]. The independent variables were the wrapper method and the hybrid approach. The dependent variable was the false positive rate for the clustering-based DDoS attack detection methods. We designed the hybrid approach to have the filter method preceded by the wrapper method. The wrapper and hybrid approach were adjusted using the switching approximations of the metaheuristic search techniques versus the 'BestFirst' approach. CSA, FPA, and FSA were applied to fine-tune the switching approximation process in order to find the best solution in either the lower or upper bounds. We used the Weka workbench as the corresponding data mining tool in our experimentation. This tool enables modeling and assessing of DDoS attack detection methods [9]. It comprises various machine learning algorithms. Likewise, the Weka workbench includes data preprocessing techniques [5]. These techniques ensure that network traffic data are clean and usable by this tool for accurate prediction by the machine learning algorithms.

Obtaining high performance in intrusion detection methods is essential for the field. Large volumes of data lead to lower performance of clustering-based DDoS attack detection methods in terms of their false positive rates [5]. Feature selection is an important process to address high-dimensional data and enable better detection [10]. Feature selection impacts the performance of the attack detection models by providing more accurate identification of the attacks. Consequently, we formulated the following research questions in this study. With respect to the first experiment, our research question was as follows: "Does incorporating the switching approximations in the meta-heuristic search techniques in the wrapper method differ in effectiveness as opposed to using the 'BestFirst' method?" The null hypothesis was that incorporating the switching approximations in the meta-heuristic search techniques in the wrapper method does not differ in effectiveness for clustering-based DDoS attack detection methods as opposed to using the 'BestFirst' method. With respect to the second experiment, our research question was as follows: "Does incorporating the switching approximations in the meta-heuristic search techniques in the hybrid approach differ in effectiveness from the 'BestFirst' method?" The null hypothesis is that incorporating the switching approximations into the meta-heuristic search techniques in the hybrid approach does not differ in effectiveness as opposed to using the 'BestFirst' method.

3. Literature Review

This section provides a literature review of DDoS attacks and consideration of meta-heuristic search methods. Section 3.2 provides an overview of the application clustering algorithms in DDoS attack detection. Subsequently, this section explains the application of CRISP-DM framework in realizing whether incorporating meta-heuristic search techniques increases the effectiveness of clustering-based DDoS attack detection methods.

3.1. DDoS Attacks and the Consideration of Metaheuristic Search Methods

The Cisco corporation estimated that in 2023 there were more than 5.3 billion Internet users, which is approximately 2/3 of the global population [14]. With this enormous number of Internet users, the resulting cybersecurity attack surface increases significantly. The Zayo consultancy, for instance, reports that in the first two quarters of 2023, DDoS attacks have risen by 200% [15]. AI and machine learning tools have been an important part of cybersecurity methodologies for the last thirty years. However, the feature selection process in machine learning has been a particular challenge since its inception. The challenge arises due to the problem of selecting the least number of features that will increase the accuracy but decrease the computational cost of the overall data classification process. In addition, due to the presence of high dimensional datasets, the overfitting of the classification models has become a particularly trenchant problem. In particular, problems of instability and lengthy convergence times have increased in prevalence. Therefore, accurate and efficient feature selection methodologies are urgently needed.

An important subfield in the machine learning landscape is the metaheuristic approach. A metaheuristic methodology is a biologically inspired artificial intelligence strategy that is utilized to address NP-hard problems. In the field of cybersecurity, it is used to achieve satisfactory but not fully optimized solutions that achieve the design parameters of a solution for the problem. For example, in a Distributed Denial of Service attack (DDoS), hackers attempt to exhaust critical network resources by issuing a flood of TCP or UDP requests from a broad spectrum of sources in a short period of time. The metaheuristic approach can be used by defenders to identify the attacks as they are propagated. An illustration of this approach is a study by Chen et al. [16], who utilize this methodology to address low-rate distributed denial of service problems. This is a DDoS attack where the hacker generates the connection and attempts to leave the sessions open for as long as possible. Considering DDoS attack detection methods in analyzing large volumes of network traffic data, proper identification of attacks becomes challenging. Subsequently, the filter and wrapper methods are used to perform network feature selection or optimization to address the analysis of large volumes of data by attack detection methods. Both of these methods apply a search method to enable the realization of a suitable solution. The filter method uses the 'Ranker' search method to determine a solution [5]. On the other hand, metaheuristic search methods are used by the wrapper method. The wrapper method is able to address the feature selection problem with generally good results [10].

Choosing suitable metaheuristic methodologies for this type of optimization problem is challenging [17]. Current methodologies use a switching approximation factor that, according to Palaniswamy and Kandhasamy [12], identifies the upper and lower approximations that range from 0 to 1, and the upper approximation could be between 0.7 and 0.9 and lower approximations could be between 0.1 and 0.3. Adjusting the parameters of the metaheuristic search algorithms has a major impact on their effectiveness [18]. Analyzing the values of some parameters for the enhancement of these algorithms in the optimization process of network traffic features is critical. Determining proper values can lead to a more accurate identification of this form of attack.

DDoS attacks make an online site or a server inactive with malicious network traffic [19]. Therefore, this is a big challenge, as attack detection systems analyze high-dimensional network traffic data. Metaheuristic algorithms have had a significant role in resolving optimization problems [20]. They are able to determine approximations for a solution to a challenging problem [21]. These algorithms are capable of identifying a set of possible solutions [22]. The study by Zafar et al. [21] illustrated that metaheuristic search techniques achieved higher performance than when all features were considered. According to Zafar et al. [21], the experimentation was conducted using data based on motor imagery and mental arithmetic tasks for the evaluation and enhancement of brain-computer interface applications. These algorithms can be considered a black box, giving the most suitable variable values for the most suitable solution [23].

Arivudainambi et al. [20] investigated the lion optimization algorithm as a proposed metaheuristic optimization algorithm in IDSs for enhancing their performance in DDoS attack detection. They were able to achieve a high accuracy of 96%. In the study by Palaniswamy and Kandhasamy [12], when the incorporation of CSA was investigated, it was able to gain superior performance in their experimentation with the average correlation values '0.9999' and '1.0000'. These values were obtained using datasets from the Gene Expression Omnibus (GEO) repository. According to Palaniswamy and Kandhasamy [12], the datasets represented real-world time series of human gene information collected from different scenarios. Demirci et al. [23] proposed a metaheuristic technique referred to as the Electrical Search Algorithm, based on electricity movement, in being able to realize exact solutions and being efficient. In another study by Saw and Oo [10], the incorporation of the particle swarm optimization technique was able to achieve relatively high performance in contrast to the genetic algorithm, ant colony optimization, and evolutionary algorithm. These algorithms are metaheuristic techniques that are inspired by nature in determining suitable features.

In summary, we can conclude that feature selection is an important process for DDoS attack detection methods through various means in choosing proper network traffic features. Feature selection leads to an accuracy improvement [24]. Incorporating feature selection is an essential step for every attack detection method [25]. Applying metaheuristic algorithms in feature selection to address optimization problems of DDoS attack detection methods is advantageous. This approach to feature selection could lead to the better identification of attacks. A good classification model is constructed through robust and representative features [26]. Therefore, a good attack detection method should include a vigorous feature selection process to facilitate detection improvement.

3.2. Application of the CRISP-DM to Applied IT Problem

Dimensionality reduction is an important process in clustering analysis [27]. Clustering analysis is an unsupervised approach to categorizing data with the same characteristics in distinct groups [28]. Nevertheless, DDoS attack detection methods that incorporate clustering algorithms are not effective in attack recognition because of the curse of dimensionality [5]. The mean differences of Tukey and Dunnett's C in ANOVA statistical analyses conducted by Zeinalpour and Ahmed [9] illustrated that the addition of the filter method did not result in better effectiveness in detecting DDoS attacks vs. the addition of the wrapper and clustering methods. However, the addition of the wrapper method did improve the results overall. To further investigate the effectiveness of clustering-based DDoS attack detection methods, we considered two clustering algorithms of k-means and EM. The k-means algorithm uses distance-based analysis that enables the maximization of the intra-cluster distances and the minimization of inter-cluster distances [5]. Appiah et al. [29] define the k-means algorithm in the following way: where N is the number of feature vectors to be categorized, K is the number of sets for observations, X_i is the corresponding observation, μ_k is the center of the cluster, and z_{ik} is the indicator of either belonging to a cluster or not.

$$J(k) = \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \mid\mid X_i - \mu_k \mid\mid^2 \tag{1}$$

The EM algorithm uses similarity-based analysis. According to Sun et al. [30], this algorithm performs the distribution assessment of data, and it does this in two steps of expectation and maximization iteratively. The EM algorithm tries to achieve the maximum likelihood estimations by creating lower boundaries and performing optimization of these estimations [29]. Sun et al. [30] present the algorithm below, where the E-step, presented in the second formula, achieves the expectation of the Z variable based on the distribution parameter of θ, and the M-step in the third formula finds a new distribution based on the maximization of the likelihood of the *L(θ)* function.

$$Q_i(z_i) = P(z_i|x_i;\theta) \tag{2}$$

$$\theta \ := \text{arg}\ max_{\theta} \sum_{i=1}^{n} \sum_{zi} Q_i(z_i) \ln\left({}^{P(x_i, z_i;\theta)}/_{Q_i(z_i)}\right) \tag{3}$$

In this study, we applied a cross-industry standard process for data mining (CRISP-DM) framework to facilitate our research. This framework is a data mining process model that is open-sourced and well-grounded [31]. CRISP-DM provides industrial independence [32]. It can help IT and information security practitioners build a proper model using machine learning algorithms from existing network traffic data and deploying it properly to facilitate the identification of unknown attacks. This framework allows businesses to safeguard their systems from service interruptions instigated by a DDoS attack [5]. CRISP-DM contains six phases of business understanding, data assessment, data preparation, modeling, evaluation, and deployment [31–33]. These phases are explained below.

Business Understanding: This phase attempts to determine the goals of a project from the business point of view [31]. This phase encompasses a domain issue that companies are facing [5]. The knowledge that is gained is then transformed into a data mining problem description to reach the strategic goals [32]. DDoS attack detection methods that apply clustering algorithms for recognizing attacks do not have a good performance due to the curse of dimensionality [5]. The aim of this research was to examine whether including meta-heuristic search techniques enhances the performance of clustering-based DDoS attack detection methods, given their switching approximations.

Data Assessment: This phase develops the data assembly and becoming familiarized with data [25]. We used the CICIDS2017 network dataset. This dataset represents a realistic and recent dataset for this problem domain [34]. The CICIDS2017 network traffic dataset has data that represent normal network traffic and DDoS attacks [5].

Data Preparation: This phase permits the transformation of data through its cleaning and formatting [31]. The CICIDS2017 network traffic dataset has redundant attributes, as well as features that are not relevant, null values, and unidentified/infinity values that require data processing [35]. The data preparation of the CICIDS2017 dataset was followed based on previous work [5] that initially considered applying manual attribute removal, numeric cleaner technique, min–max maximization, EM imputation, spread subsample procedure, and randomization. Applying these techniques and procedures for data preparation enabled us to remove any unwanted network attributes, set the values in the proper range of 0 and 1, and balance the data instances of benign and DDoS attack events to ensure the results are not biased. These data preparation techniques and procedures were also considered in previous work [9], where the effectiveness of DDoS attack detection methods based on the clustering method used a vote classifier.

Modeling: This phase involves constructing various data mining models [31,33]. Usually, these models are for the identified data mining problem [32]. The purpose of this study was to realize whether incorporating meta-heuristic search techniques increases the effectiveness of clustering-based DDoS attack detection methods, given their switching approximations. CSA, FPA, and FSA were applied for tuning the switching approximation by using the values of 0.25, 0.50, and 0.75.

According to Maddaiah et al. [36], CSA applies three rules that are based on the breeding behavior of some cuckoo birds. The first rule is that every cuckoo lays one egg and places it in a random nest; the second rule is that the nest with the highest fitness proceeds to the next generation [37]. The third rule is that the probability of finding an egg is in the range of 0 and 1 using the 'pa' parameter [36,37]. As brought by Salgotra [37], incorporating these rules is achieved by using the Levy flight to generate a new solution as presented below. Based on Sicuaio et al. [38], Lévy (λ) is the Levy distribution in performing random walks. X_i^t is considered the previous solution, ⊕ is regarded as entry-wise multiplication, and α is the step size for which in most cases the value of 1 is used [37].

$$X_i^{t+1}= X_i^t \ + \ \alpha \oplus \text{Lévy } (\lambda) \tag{4}$$

FPA follows four rules in which the first rule is performing global pollination, and the second is conducting local pollination [39]. The equation that is numbered 5 is the global pollination formula. The equation that is numbered 6 is the local pollination formula. Global pollination is achieved through Levy distribution [39,40]. According to Salgotra et al. [37], the third rule states that flower consistency is proportionate to the likeness of two flowers, which is represented as the reproduction probability. The fourth rule refers to the use of pollination approximation. This switching approximation balances the global and local pollinations [37]. Emary et al. [40] present the global and local pollinations as follows:

$$X_i^{t+1} = X_i^t + L\left(X_i^t - g_*\right) \tag{5}$$

$$X_i^{t+1} = X_i^t + \in\left(X_j^t - g_k^t\right) \tag{6}$$

Global pollination is the cross-pollination, while local pollination is the self-pollination [37]. With respect to Equation (5) above, X_i^t is the solution vector in a given iteration of t, g_* is the current greatest solution, and L is the pollination strength calculated through performing the Levy distribution [40]. In the local pollination formula, ∈ represents the uniform distribution between 0 and 1 [37,40]. X_j^t and g_k^t represent random solution vectors that are selected [40]. FSA follows three rules for the optimization of features. The first rule is that a firefly is attracted to other fireflies irrespective of their sex, and the second rule is that they have attractiveness proportional to their brightness, with the third rule being that having an objective function determines the brightness [41]. Zhang and Wang [42] present FSA as shown below, where I is the objective function in which I_0 denotes the brightest firefly, γ represents the light absorption coefficient, r_{ij} is the distance between fireflies, and β is the attraction, with β_0 denoting the highest degree for attraction. The α is the randomization parameter, and '*rand*' is a randomly generated number that is between 0 and 1 [43].

$$I = I_0\ .\ e^{-\gamma.r_{\mathrm{ij}}} \tag{7}$$

$$X_i^{t+1} = X_i^t + \beta\ .\ \left(X_j^t - X_i^t\right) + \alpha\ .\ \left(rand - \frac{1}{2}\right) \tag{8}$$

$$\beta(r) = \beta_0\ .e^{-\gamma.r_{ij}^2} \tag{9}$$

We compared the chosen metaheuristic search techniques against the 'BestFirst' search technique of the wrapper method. The 'BestFirst' is the default search technique in Weka when the wrapper method is incorporated for feature selection. This technique applies a heuristic approach to find reliable solutions [44]. In all cases of using the wrapper method, we applied Naïve Bayes and k-means algorithms as the feature evaluators. The wrapper method incorporates a learning algorithm in assessing network traffic features [5]. Naïve Bayes is a learning model [45]. The k-means algorithm is able to categorize a dataset into classes [46], and in this case, it will categorize the network traffic according to clusters based on the closest averages [5]. This can enhance the capability of the wrapper method in selecting features. We chose the Chi-squared and information gain techniques as the feature evaluators in the filter method, as it was chosen in previous work [5]. Having considered the filter method in the hybrid approach as well, we used the threshold of '0.5' in its 'Ranker' search method. An evaluation score above 0.5 indicates the relevancy of a feature to a category of data that can provide information that is useful [5].

Evaluation: This phase allows for the assessment of the models that were constructed during the modeling phase [31]. To evaluate clustering-based DDoS attack detection models, we applied the false positive rate metric. This metric computes the ratio that is obtained from the number of incorrect identifications of normal data instances as attacks divided by all normal instances [47].

Deployment: This phase allows users to apply data mining models for decision-making purposes [31]. We recommend organizations incorporate DDoS attack detection models that we find effective outside of demilitarized zones (DMZs) to facilitate the detection of attacks directly from the Internet. Placing DDoS attack detection models outside of a DMZ helps in recognizing attacks directly from networks that are external to organizations, and this promotes a cleaner network traffic path [5].

4. Experiments and Evaluation

We conducted this study using an ex post facto design of an A-B single group. We assessed whether incorporating meta-heuristic search techniques differs in the effectiveness of clustering-based DDoS attack detection methods, given their switching approximations. We considered 0.25, 0.50, and 0.75 switching approximation values when incorporating the wrapper method and the hybrid approach in evaluating clustering-based DDoS attack detection methods. In this study, we ensured internal, predictive, conclusion, and external validities. To guarantee internal validity, we used the Weka (Waikato Environment for Knowledge Analysis) workbench. This tool is licensed under GNU-GPL, which was developed by the University of Waikato in New Zealand by comprising a collection of Java libraries for mining knowledge from databases [48]. In ensuring predictive and conclusion validities, we applied a ten-fold cross-validation method. This method enables accurate performance of classifiers [49]. The reliability of results is guaranteed by the ten-fold cross-validation method [5]. To guarantee external validity, we used the entire CICIDS2017 network traffic dataset. This dataset includes up-to-date attack methods that contain related network features in facilitating the modeling of patterns and recognizing attacks within the context of benign network traffic [48]. Figure 1 below presents the construction of the clustering-based DDoS attack detection models.

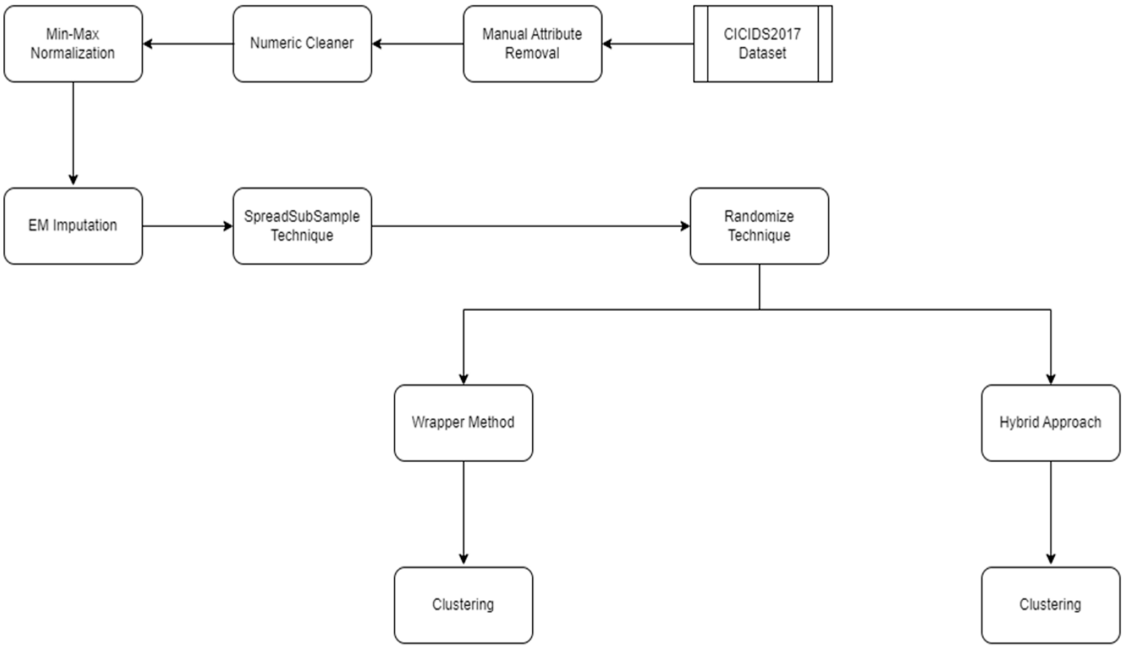

Figure 1. DDoS attack detection method construction.

4.1. Statistical Analysis Using One-Way ANOVA

ANOVA presents a set of statistical models for identifying any significance of differences among the means of two or more independent groups [50]. We performed one-way ANOVA analyses due to considering one factor variable and one dependent variable. The factor variable categorizes cases into two or more levels, and the dependent variable differentiates cases on a quantitative measure [51]. In this study, we conducted two one-way ANOVA analyses. The factor variable was the "Method", and the dependent variable was the false positive rates of clustering-based DDoS attack detection methods. The "Method" variable, in the first one-way ANOVA analysis, included the wrapper method when the 'BestFirst' was included against the wrapper method using metaheuristic search techniques. In the second one-way ANOVA analysis, the hybrid approach of the filter–wrapper method was included when the 'BestFirst' was incorporated against the filter–wrapper method using metaheuristic search techniques.

In this study, we considered the one-way ANOVA F-test. The results of the first one-way ANOVA analysis for the first experiment are presented in Table 1. The outcomes show that the test was not significant with $F(1, 38) = 0.17$ and $p = 0.7$. The "Sig" column below represents the p-value. The p-value was greater than 0.05, which led us not to reject the null hypothesis. The η^2, named "Partial Eta Squared", in the table below with the value of 0.004 shows that there were no strong relationships among DDoS attack detection methods when incorporating the wrapper using the 'BestFirst' against metaheuristic search techniques.

The results of the second one-way ANOVA analysis for the second experiment are presented in Table 2. The outcomes presented below show that the test was not significant with $F(1, 78) = 0.7$ and $p = 0.4$. The "Sig" column below represents the p-value. The p-value was greater than 0.05, which led us not to reject the null hypothesis in that incorporating meta-heuristic search techniques in the hybrid approach does not differ in the effectiveness of clustering-based DDoS attack detection methods as opposed to using the 'BestFirst', due to switching approximations of metaheuristic search techniques. The η^2, named "Partial Eta Squared", in the table below with the value of 0.009 shows that there were no strong relationships among DDoS attack detection methods when the hybrid approach of the filter–wrapper method was applied using the 'BestFirst' in contrast to metaheuristic search techniques given their switching approximation values.

Table 1. Tests of between-subjects effects for the wrapper method.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	0.001 [a]	1	0.001	0.167	0.685	0.004
Intercept	0.413	1	0.413	52.508	<0.001	0.580
Method	0.001	1	0.001	0.167	0.685	0.004
Error	0.299	38	0.008			
Total	1.552	40				
Corrected Total	0.300	39				

[a] R Squared = 0.004 (adjusted R Squared = −0.022).

Table 2. Tests of between-subjects effects for the hybrid approach.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	0.008 [a]	1	0.008	0.691	0.408	0.009
Intercept	0.383	1	0.383	32.638	<0.001	0.295
Method	0.008	1	0.008	0.691	0.408	0.009
Error	0.915	78	0.012			
Total	2.248	80				
Corrected Total	0.923	79				

[a] R Squared = 0.009 (adjusted R Squared = −0.004).

4.2. Comparison of the Wrapper Method in the First Experimentation

The mean results of descriptive statistics presented in Table 3 show that the performance of clustering-based DDoS attack detection methods was slightly better when the 'BestFirst' was applied in the wrapper method. This was not in the case of metaheuristic search techniques considering switching approximations of 0.25, 0.50, and 0.75. The mean result with the value of 0.159, when using the 'BestFirst', reflects a slightly higher performance than metaheuristics with a mean value of 0.178.

Table 3. Descriptive statistics for the wrapper method.

Method	Mean	Std. Deviation	N
BestFirst-Wrapper	0.15975	0.143486	4
Metaheuristic-Wrapper	0.17883	0.082272	36
Total	0.17693	0.087703	40

4.3. Comparison of the Hybrid Approach in the Second Experimentation

The mean results of the descriptive statistics of Table 4 show that the performance of the clustering-based DDoS attack detection methods was somewhat superior when the hybrid approach was considered using the 'BestFirst'. The hybrid approach did not perform better with the metaheuristic search techniques by applying switching approximations of 0.25, 0.50, and 0.75. The mean result with the value of 0.098 for when the 'BestFirst' was applied in the hybrid approach reflects on slightly higher performance than metaheuristics with the mean value of 0.132.

Table 4. Descriptive statistics for the hybrid approach.

Method	Mean	Std. Deviation	N
BestFirst-Hybrid	0.09850	0.118605	8
Metaheuristic-Hybrid	0.13206	0.107217	72
Total	0.12870	0.108076	80

5. Discussion

The sophistication of cyber intrusion techniques has grown significantly in the last ten years. This poses ever greater challenges for traditional IDS detection methods [52]. Intrusion detection methods are beneficial in guaranteeing that network administrators are correctly alerted to dangerous events [7]. Anomaly-based intrusion detection methods analyze the characteristics of network events in identifying attacks through network features or data. Feature selection processes in dealing with high-dimensional datasets are necessary [10]. When constructing machine learning models, quality is impacted by training data in that one way of achieving good quality training data is through effective feature selection [7]. Dimensionality reduction is considered a serious issue while incorporating high-dimensional data and metaheuristic search methods are utilized to address this issue [10]. Dimensionality reduction or feature selection represents a key optimization problem in this process. Shukla [53] examined the false positive rates of genetic algorithms (GA), particle swarm optimization (PSO), differential evolution (DE), grasshopper optimization algorithm (GOA), and artificial bee colony (ABC) in detecting intrusions using the CICIDS2017 dataset. According to Boveiri and Khayami [17], these optimization algorithms are metaheuristic. The false positive rates for identifying intrusions using the CICIDS2017 dataset presented in the study by Shukla [53] were 3.025, 2.364, 2.842, 1.005, and 1.578, respectively. In the study by Sarvari et al. [54], the application of CSA in the optimization problem using the NSL-KDD network traffic dataset enabled the achievement of a false positive rate of 0.03 in attack detection.

Feature selection is useful in selecting relevant features and leads to a decrease in the false positive rates [55]. The previous study [9], illustrated that the hybrid approach of the filter–wrapper method was more effective than the application of the filter method selecting proper features for detecting attacks. Considering the hybrid approach of the filter–wrapper method, the previous work [9] showed that the best performance was 0.012 in the false positive rate. This was achieved using Chi-squared for the filter method and J48 for the wrapper method. We were able to achieve the best performance of 0.000 in the false positive rate. This was when k-means and BestFirst were applied in the hybrid approach. We obtained the second best performance of 0.005 in two cases as presented in Tables A2 and A4 of Appendix B. In general, the 'BestFirst' performance was slightly better, given the mean differences from Tables 3 and 4 realized from the descriptive statistics above.

6. Limitations and Implications

Limitations are issues that cannot be controlled and pose problems to internal validity [56]. The limitation of this study was the curse of dimensionality. DDoS attack detection methods that apply machine learning algorithms are susceptible to the curse of dimensionality due to the existence of a large volume of traffic data [9]. To address this limitation in our study, we considered metaheuristic search techniques for the wrapper method and the hybrid approach. Likewise, implications are assumptions and delimitations [9]. Researchers regard assumptions to be factual without investigation, and delimitations are considerations of variables that are purposefully disregarded [56]. In other words, assumptions are those that are regarded to be true in the case of machine learning algorithms in analyzing data, and delimitations determine the boundaries of studies in evaluating these algorithms. In this study, we assumed that the results of this study would be reflective of the real world as we used the CICIDS2017 dataset. This dataset is one of the newest iterations for intrusion detection modeling in containing up-to-date attacks [57]. The delimitation of this study was investigating signature-based intrusion detection systems. These systems match patterns to identify known attacks [52].

7. Conclusions

Applying vigorous feature selection processes in intrusion detection methods is essential. Parameter-tuning of metaheuristic search techniques is one critical element in improving the field [18]. In this study, we applied three metaheuristic techniques of CSA, FPA, and FSA for both the wrapper method and the hybrid approach. Approximations in CSA, FPA, and FSA were used in realizing the best solution in either lower or upper given values ranging from 0 to 1. We used three values of 0.25, 0.50, and 0.75. The results of this study showed that applying these approximation values has a positive impact on the performance of clustering-based DDoS attack detection methods in contrast to when clustering was only applied with no consideration of metaheuristic search techniques. We were able to achieve the best performance of 0.000 in the false positive rate using k-means and BestFirst in the hybrid approach. We achieved the second best performance of 0.005 in two cases. The first case was in Table A2 under Appendix B, when the k-means algorithm was used to cluster DDoS attacks by incorporating Chi-squared and FSA in the hybrid approach. The switching approximation that was used for FSA in this case was 0.50. The second case was in Table A4 under Appendix B, when the k-means algorithm was used to cluster DDoS attacks by applying the wrapper method. In this case, the wrapper method incorporated the k-means algorithm and the 'BestFirst'. One future research direction would be to examine the performance of DDoS attack detection methods by applying the 'BestFirst' and metaheuristic search techniques on various network traffic datasets. This is achieved by considering the wrapper and hybrid approach of filter–wrapper method. The results of future studies could reflect our findings on other datasets than the CICIDS2017 network traffic dataset used in this study. In addition, other future research efforts can explore a broader spectrum of parameters to tune and the impact that these interventions have on the holistic performance of those algorithms with respect to fundamental issues such as the false positive rate, the speed to convergence, etc.

Author Contributions: The authors of this research study worked in collaboration to complete the research study. All authors have read and agreed to the published version of the manuscript.

Funding: This research had no external funding.

Data Availability Statement: The authors of this study used the CICIDS2017 dataset. This dataset is publicly available at https://www.unb.ca/cic/datasets/ids-2017.html (accessed on 8 April 2023).

Acknowledgments: The authors conducted this study using data preprocessing methodologies considered by Zeinalpour [5].

Conflicts of Interest: The authors of this study declare no conflicts of interest. The authors of this research guided the study with no sponsorship.

Appendix A. Independent Variables Table

Table A1. Independent variables table.

Independent Variables	Procedures
Wrapper Method	(SimpleKMeans and CuckooSearch) (SimpleKMeans and FireFlySearch) (SimpleKMeans and FlowerSearch) (NaïveBayes and CuckooSearch) (NaïveBayes and FireFlySearch) (NaïveBayes and FlowerSearch) (SimpleKMeans and BestFirst) (NaïveBayes and BestFirst)

Table A1. *Cont.*

Independent Variables	Procedures
Hybrid Approach	(InfoGainAttributeEval) + (SimpleKMeans and CuckooSearch) (InfoGainAttributeEval) + (SimpleKMeans and FireFlySearch) (InfoGainAttributeEval) + (SimpleKMeans and FlowerSearch) (ChiSquaredAttributeEval) + (SimpleKMeans and CuckooSearch) (ChiSquaredAttributeEval) + (SimpleKMeans and FireFlySearch) (ChiSquaredAttributeEval) + (SimpleKMeans and FlowerSearch) (InfoGainAttributeEval) + (NaïveBayes and CuckooSearch) (InfoGainAttributeEval) + (NaïveBayes and FireFlySearch) (InfoGainAttributeEval) + (NaïveBayes and FlowerSearch) (ChiSquaredAttributeEval) + (NaïveBayes and CuckooSearch) (ChiSquaredAttributeEval) + (NaïveBayes and FireFlySearch) (ChiSquaredAttributeEval) + (NaïveBayes and FlowerSearch) (InfoGainAttributeEval) + (SimpleKMeans and BestFirst) (ChiSquaredAttributeEval) + (SimpleKMeans and BestFirst) (InfoGainAttributeEval) + (NaïveBayes and BestFirst) (ChiSquaredAttributeEval) + (NaïveBayes and BestFirst)

Appendix B. Experimental Results

Table A2. FPR table using simple k-means and metaheuristics.

DDoS Attacks Detection Methods Applied Procedures	False Positive Rates in DDoS Attack Identification
Cuckoo_0.25 → EM	0.152
Cuckoo_0.50 → EM	0.129
Cuckoo_0.75 → EM	0.059
FireFly_0.25 → EM	0.071
FireFly_0.50 → EM	0.117
FireFly_0.75 → EM	0.108
Flower_0.25 → EM	0.054
Flower_0.50 → EM	0.129
Flower_0.75 → EM	0.129
ChiSquared_Cuckoo_0.25 → EM	0.045
InfoGain_Cuckoo_0.25 → EM	0.079
ChiSquared_Cuckoo_0.50 → EM	0.021
InfoGain_Cuckoo_0.50 → EM	0.016
ChiSquared_Cuckoo_0.75 → EM	0.407
InfoGain_Cuckoo_0.75 → EM	0.088
ChiSquared_FireFly_0.25 → EM	0.118
InfoGain_FireFly_0.25 → EM	0.142
ChiSquared_FireFly_0.50 → EM	0.053
InfoGain_FireFly_0.50 → EM	0.163
ChiSquared_FireFly_0.75 → EM	0.226
InfoGain_FireFly_0.75 → EM	0.010
ChiSquared_Flower_0.25 → EM	0.105
InfoGain_Flower_0.25 → EM	0.144
ChiSquared_Flower_0.50 → EM	0.284
InfoGain_Flower_0.50 → EM	0.161
ChiSquared_Flower_0.75 → EM	0.284
InfoGain_Flower_0.75 → EM	0.161
Cuckoo_0.25 → SimpleKMeans	0.146
Cuckoo_0.50 → SimpleKmeans	0.094
Cuckoo_0.75 → SimpleKMeans	0.156
FireFly_0.25 → SimpleKMeans	0.112
FireFly_0.50 → SimpleKMeans	0.107

Table A2. *Cont.*

DDoS Attacks Detection Methods Applied Procedures	False Positive Rates in DDoS Attack Identification
FireFly_0.75 → SimpleKMeans	0.112
Flower_0.25 → SimpleKMeans	0.093
Flower_0.50 → SimpleKMeans	0.088
Flower_0.75 → SimpleKMeans	0.088
ChiSquared_Cuckoo_0.25 → SimpleKMeans	0.092
InfoGain_Cuckoo_0.25 → SimpleKMeans	0.006
ChiSquared_Cuckoo_0.50 → SimpleKMeans	0.176
InfoGain_Cuckoo_0.50 → SimpleKMeans	0.007
ChiSquared_Cuckoo_0.75 → SimpleKMeans	0.091
InfoGain_Cuckoo_0.75 → SimpleKMeans	0.009
ChiSquared_FireFly_0.25 → SimpleKMeans	0.098
InfoGain_FireFly_0.25 → SimpleKMeans	0.006
ChiSquared_FireFly_0.50 → SimpleKMeans	0.005
InfoGain_FireFly_0.50 → SimpleKMeans	0.006
ChiSquared_FireFly_0.75 → SimpleKMeans	0.008
InfoGain_FireFly_0.75 → SimpleKMeans	0.094
ChiSquared_Flower_0.25 → SimpleKMeans	0.086
InfoGain_Flower_0.25 → SimpleKMeans	0.007
ChiSquared_Flower_0.50 → SimpleKMeans	0.027
InfoGain_Flower_0.50 → SimpleKMeans	0.008
ChiSquared_Flower_0.75 → SimpleKMeans	0.027
InfoGain_Flower_0.75 → SimpleKMeans	0.008

Table A3. FPR table using NaïveBayes and metaheuristics.

DDoS Attacks Detection Methods Applied Procedures	False Positive Rates in DDoS Attack Identification
Cuckoo_0.25 → EM	0.229
Cuckoo_0.50 → EM	0.288
Cuckoo_0.75 → EM	0.184
FireFly_0.25 → EM	0.329
FireFly_0.50 → EM	0.282
FireFly_0.75 → EM	0.244
Flower_0.25 → EM	0.294
Flower_0.50 → EM	0.252
Flower_0.75 → EM	0.252
ChiSquared_Cuckoo_0.25 → EM	0.329
InfoGain_Cuckoo_0.25 → EM	0.154
ChiSquared_Cuckoo_0.50 → EM	0.272
InfoGain_Cuckoo_0.50 → EM	0.162
ChiSquared_Cuckoo_0.75 → EM	0.216
InfoGain_Cuckoo_0.75 → EM	0.109
ChiSquared_FireFly_0.25 → EM	0.349
InfoGain_FireFly_0.25 → EM	0.160
ChiSquared_FireFly_0.50 → EM	0.318
InfoGain_FireFly_0.50 → EM	0.154
ChiSquared_FireFly_0.75 → EM	0.292
InfoGain_FireFly_0.75 → EM	0.162
ChiSquared_Flower_0.25 → EM	0.269
InfoGain_Flower_0.25 → EM	0.160
ChiSquared_Flower_0.50 → EM	0.208
InfoGain_Flower_0.50 → EM	0.157
ChiSquared_Flower_0.75 → EM	0.208
InfoGain_Flower_0.75 → EM	0.157

Table A3. *Cont.*

DDoS Attacks Detection Methods Applied Procedures	False Positive Rates in DDoS Attack Identification
Cuckoo_0.25 → SimpleKMeans	0.302
Cuckoo_0.50 → SimpleKmeans	0.127
Cuckoo_0.75 → SimpleKMeans	0.218
FireFly_0.25 → SimpleKMeans	0.248
FireFly_0.50 → SimpleKMeans	0.187
FireFly_0.75 → SimpleKMeans	0.293
Flower_0.25 → SimpleKMeans	0.229
Flower_0.50 → SimpleKMeans	0.268
Flower_0.75 → SimpleKMeans	0.268
ChiSquared_Cuckoo_0.25 → SimpleKMeans	0.155
InfoGain_Cuckoo_0.25 → SimpleKMeans	0.010
ChiSquared_Cuckoo_0.50 → SimpleKMeans	0.247
InfoGain_Cuckoo_0.50 → SimpleKMeans	0.060
ChiSquared_Cuckoo_0.75 → SimpleKMeans	0.282
InfoGain_Cuckoo_0.75 → SimpleKMeans	0.014
ChiSquared_FireFly_0.25 → SimpleKMeans	0.298
InfoGain_FireFly_0.25 → SimpleKMeans	0.014
ChiSquared_FireFly_0.50 → SimpleKMeans	0.224
InfoGain_FireFly_0.50 → SimpleKMeans	0.042
ChiSquared_FireFly_0.75 → SimpleKMeans	0.192
InfoGain_FireFly_0.75 → SimpleKMeans	0.043
ChiSquared_Flower_0.25 → SimpleKMeans	0.213
InfoGain_Flower_0.25 → SimpleKMeans	0.010
ChiSquared_Flower_0.50 → SimpleKMeans	0.275
InfoGain_Flower_0.50 → SimpleKMeans	0.025
ChiSquared_Flower_0.75 → SimpleKMeans	0.275
InfoGain_Flower_0.75 → SimpleKMeans	0.025

Table A4. FPR table using simple k-means and BestFirst.

DDoS Attacks Detection Methods Applied Procedures	False Positive Rates in DDoS Attack Identification
SimpleKMeans_BestFirst → EM	0.086
SimpleKMeans_BestFirst → SimpleKMeans	0.005
ChiSquared_ SimpleKMeans_BestFirst → EM	0.006
ChiSquared_ SimpleKMeans_BestFirst → SimpleKMeans	0.102
InfoGain_ SimpleKMeans_BestFirst → EM	0.000
InfoGain_ SimpleKMeans_BestFirst → SimpleKMeans	0.006

Table A5. FPR table using NaïveBayes and BestFirst.

DDoS Attacks Detection Methods Applied Procedures	False Positive Rates in DDoS Attack Identification
NaïveBayes_BestFirst → EM	0.330
NaïveBayes_BestFirst → SimpleKMeans	0.218
ChiSquared_ NaïveBayes_BestFirst → EM	0.330
ChiSquared_ NaïveBayes_BestFirst → SimpleKMeans	0.218
InfoGain_ NaïveBayes_BestFirst → EM	0.088
InfoGain_ NaïveBayes_BestFirst → SimpleKMeans	0.038

References

1. Zhou, L.; Zhu, Y.; Xiang, Y.; Zong, T. A novel feature-based framework enabling multi-type DDoS attacks detection. *World Wide Web* **2023**, *26*, 163–185. [CrossRef]
2. Xu, J.; Li, X.; Wang, P.; Jin, X.; Yao, S. Multi-modal noise-robust DDoS attack detection architecture in large-scale networks based on tensor SVD. *IEEE Trans. Netw. Sci. Eng.* **2023**, *10*, 152–165. [CrossRef]
3. Prasad, A.; Chandra, S. VMFCVD: An optimized framework to combat volumetric DDoS attacks using machine learning. *Arab. J. Sci. Eng.* **2022**, *47*, 9965–9983. [CrossRef]

4. Mishra, A.; Gupta, N.; Gupta, B.B. Defensive mechanism against DDoS attack based on feature selection and multi-classifier algorithms. *Telecommun. Syst.* **2023**, *82*, 229–244. [CrossRef]
5. Zeinalpour, A. Addressing High False Positive Rates of DDoS Attack Detection Methods. D.I.T. Thesis, Walden University, Minneapolis, MN, USA, 2021.
6. Li, Y.; Ghoreishi, S.M.; Issakhov, A. Improving the accuracy of network intrusion detection system in medical IoT systems through butterfly optimization algorithm. *Wirel. Pers. Commun.* **2022**, *126*, 1999–2017. [CrossRef]
7. Megantara, A.A.; Ahmad, T. A hybrid machine learning method for increasing the performance of network intrusion detection systems. *J. Big Data* **2021**, *8*, 142. [CrossRef]
8. Idhammad, M.; Afdel, K.; Belouch, M. Semi-supervised machine learning approach for DDoS detection. *Appl. Intell.* **2018**, *48*, 3193–3208. [CrossRef]
9. Zeinalpour, A.; Ahmed, H.A. Addressing the effectiveness of DDoS-attack detection methods based on the clustering method using an ensemble method. *Electronics* **2022**, *11*, 2736. [CrossRef]
10. Saw, T.; Oo, W.M. Ranking-based feature selection with wrapper PSO search in high-dimensional data classification. *Int. J. Comput. Sci.* **2023**, *50*, 1–16.
11. Thakur, K.; Kumar, G. Nature inspired techniques and applications in intrusion detection systems: Recent progress and updated perspective. *Arch. Comput. Methods Eng.* **2021**, *28*, 2897–2919. [CrossRef]
12. Palaniswamy, S.; Kandhasamy, P. Rough fuzzy cuckoo search for triclustering microarray gene expression data. *Turk. J. Electr. Eng. Comput. Sci.* **2019**, *27*, 4328–4339. [CrossRef]
13. Tanious, R.; Onghena, P. Randomized single-case experimental designs in healthcare research: What, why, and how. *Healthcare* **2019**, *7*, 143. [CrossRef] [PubMed]
14. Cisco.com. *Cisco Annual Internet Report (2018–2023) White Paper*; Cisco: San Jose, CA, USA, 2020. Available online: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html (accessed on 12 November 2023).
15. Zayo. Protecting Your Business from Cyber Attacks: The State of DDoS Attacks DDoS Insights from Q1 & Q2, 2023. 2023. Available online: https://go.zayo.com/zayo-ddos-protection-ebook/ (accessed on 12 November 2023).
16. Chen, H.H.; Lee, C.H.; Huang, S.K. A Unified Ant Agent Framework for Solving DoS and QoS Problems. *J. Inf. Sci. Eng.* **2016**, *32*, 1397–1434. Available online: https://jise.iis.sinica.edu.tw (accessed on 13 November 2023).
17. Boveiri, H.R.; Khayami, R. On the performance of metaheuristics: A different perspective. *arXiv* **2020**, arXiv:2001.08928. [CrossRef]
18. Khalfi, S.; Carafni, F.; Iacca, G. Metaheuristics in the balance: A survey on memory-saving approaches for platforms with seriously limited resources. *Int. J. Intell. Syst.* **2023**, *2023*, 1–32. [CrossRef]
19. Sumathi, S.; Rajesh, R.; Lim, S. Recurrent and deep learning neural network models for DDoS attack detection. *J. Sens.* **2022**, *2022*, 1–21. [CrossRef]
20. Arivudainambi, D.; Kumar, V.; Sibi Chakkaravarthy, S. LION IDS: A meta-heuristics approach to detect DDoS attacks against Software-Defined Networks. *Neural Comput. Appl.* **2019**, *31*, 1491–1501. [CrossRef]
21. Zafar, A.; Hussain, S.J.; Ali, M.U.; Lee, S.W. Metaheuristic optimization-based feature selection for imagery and arithmetic tasks: An fNIRS study. *Sensors* **2023**, *23*, 3714. [CrossRef]
22. Mirhosseini, M.; Nezamabadi-pour, H. Metaheuristic search algorithms in solving the n-similarity problem. *Fundam. Informaticae* **2017**, *152*, 145–166. [CrossRef]
23. Demirci, H.; Yurtay, N.; Yurtay, Y.; Zaimoğlu, E.A. Electrical search algorithm: A new metaheuristic algorithm for clustering problem. *Arab. J. Sci. Eng.* **2023**, *48*, 10153–10172. [CrossRef]
24. Kim, Y.E.; Kim, Y.S.; Kim, H. Effective feature selection methods to detect IoT DDoS attack in 5g core network. *Sensors* **2022**, *22*, 3819. [CrossRef]
25. Chaudhary, P.; Gupta, B.; Singh, A.K. Implementing attack detection system using filter-based feature selection methods for fog-enabled IoT networks. *Telecommun. Syst.* **2022**, *81*, 23–39. [CrossRef]
26. Dabas, N.; Ahlawat, P.; Sharma, P. An effective malware detection method using hybrid feature selection and machine learning algorithms. *Arab. J. Sci. Eng.* **2023**, *48*, 9749–9767. [CrossRef]
27. Mohamed, A.A. An effective dimension reduction algorithm for clustering Arabic text. *Egypt. Inform. J.* **2019**, *21*, 1–5. [CrossRef]
28. Melnykov, V.; Michael, S. Clustering large datasets by merging k-means solutions. *J. Classif.* **2019**, *37*, 97–123. [CrossRef]
29. Appiah, S.K.; Wirekoh, K.; Aidoo, E.N.; Oduro, S.D.; Arthur, Y.D. A model-based clustering of expectation–maximization and k-means algorithms in crime hotspot analysis. *Res. Math.* **2022**, *9*, 2073662. [CrossRef]
30. Sun, Y.; Zhao, Y.; Hao, L.; Zhao, X.; Lu, J.; Shi, Y.; Ma, C. Role of the EM clustering method in determining the geochemical background of As and Cr in soils: A case study in the north of Changchun, China. *Environ. Geochem. Health* **2023**, *45*, 6675–6692. [CrossRef]
31. Pivk, A.; Vasilecas, O.; Kalibatiene, D.; Rupnik, R. On approach for the implementation of data mining to business process optimisation in commercial companies. *Technol. Econ. Dev. Econ.* **2013**, *19*, 237–256.
32. Brzozowska, J.; Pizoń, J.; Baytikenova, G.; Gola, A.; Zakimova, A.; Piotrowska, K. Data engineering in CRISP-DM process production data—Case study. *Appl. Comput. Sci.* **2023**, *19*, 83–95. [CrossRef]
33. Jaggia, S.; Kelly, A.; Lertwachara, K.; Chen, L. Applying the CRISP-DM framework for teaching business analytics. *Decis. Sci. J. Innov. Educ.* **2020**, *18*, 612–634. [CrossRef]

34. Szczepański, M.; Pawlicki, M.; Kozik, R.; Choraś, M. The application of deep learning imputation and other advanced methods for handling missing values in network intrusion detection. *Vietnam. J. Comput. Sci.* **2023**, *10*, 1–23. [CrossRef]
35. Azzaoui, H.; Boukhamla, A.Z.E.; Arroyo, D.; Bensayah, A. Developing new deep-learning model to enhance network intrusion classification. *Evol. Syst.* **2022**, *13*, 17–25. [CrossRef]
36. Maddaiah, P.N.; Narayanan, P.P. An improved cuckoo search algorithm for optimization of artificial neural network training. *Neural Process. Lett.* **2023**, *55*, 12093–12120. [CrossRef]
37. Salgotra, R.; Mittal, N.; Mittal, V. A new parallel cuckoo flower search algorithm for training multi-layer perceptron. *Mathematics* **2023**, *11*, 3080. [CrossRef]
38. Sicuaio, T.; Niyomubyeyi, O.; Shyndyapin, A.; Pilesjö, P.; Mansourian, A. Multi-objective optimization using evolutionary cuckoo search algorithm for evacuation planning. *Geomatics* **2022**, *2*, 53–75. [CrossRef]
39. Yang, X.S.; Karamanoglu, M.; Xingshi, H. Flower pollination algorithm: A novel approach for multiobjective optimization. *Eng. Optim.* **2013**, *46*, 1222–1237. [CrossRef]
40. Emary, E.; Zawbaa, H.M.; Hassanien, A.E.; Parv, B. Multi-objective retinal vessel localization using flower pollination search algorithm with pattern search. *Adv. Data Anal. Classif.* **2017**, *11*, 611–627. [CrossRef]
41. Yang, X.S. Firefly algorithms for multimodal optimization. *arXiv* **2010**, arXiv:1003.1466. [CrossRef]
42. Zhang, X.; Wang, S. Firefly search algorithm based on leader strategy. *Eng. Appl. Artif. Intell.* **2023**, *123*, 106328. [CrossRef]
43. Alomoush, W.; Omar, K.; Alrosan, A.; Alomari, Y.M.; Albashish, D.; Almomani, A. Firefly photinus search algorithm. *J. King Saud Univ.–Comput. Inf. Sci.* **2018**, *32*, 599–607. [CrossRef]
44. Samal, A.; Saxena, A.; Ray, D. Comparative study of algorithms in artificial intelligence: Best first search, greedy best first search and iterative deepening. *Int. J. Softw. Hardw. Res. Eng.* **2018**, *6*, 6–11.
45. Haviluddin; Puspitasari, N.; Burhandeny, A.E.; Nurulita, A.D.A.; Trahutomo, D. Naïve Bayes and K-nearest neighbor algorithms performance comparison in diabetes mellitus early diagnosis. *Int. J. Online Biomed. Eng.* **2022**, *18*, 202–215. [CrossRef]
46. Arora, N.; Singh, A.; Al-Dabagh, M.Z.N.; Maitra, S.K. A Novel architecture for diabetes patients' prediction using K-Means clustering and SVM. *Math. Probl. Eng.* **2022**, *2020*, 1–9. [CrossRef]
47. Khraisat, A.; Gondal, I.; Vamplew, P.; Kamruzzaman, J. Survey of intrusion detection systems: Techniques, datasets and challenges. *Cybersecurity* **2019**, *2*, 20. [CrossRef]
48. Rodríguez, M.; Alesanco, Á.; Mehavilla, L.; Garcva, J. Evaluation of machine learning techniques for traffic flow-based intrusion detection. *Sensors* **2022**, *22*, 9326. [CrossRef]
49. Karim, A.; Salleh, R.; Khan, M.K. SMARTbot: A behavioral analysis framework augmented with machine learning to identify mobile botnet applications. *PLoS ONE* **2016**, *11*, e0150077. [CrossRef]
50. Iskandar1, K.; Noprianto; Abbas, B.S.; Soewito, B.; Kosala, R. Two-way ANOVA with interaction approach to compare content creation speed performance in knowledge management system. In Proceedings of the International Conference on Knowledge, Information and Creativity Support Systems (KICSS), Yogyakarta, Indonesia, 10–12 November 2016. [CrossRef]
51. Green, S.B.; Salkind, N.J. *Using SPSS for Windows and Macintosh: Analyzing and Understanding the Data*, 8th ed.; Pearson: Upper Saddle River, NJ, USA, 2017; p. 131.
52. Panagiotou, P.; Mengidis, N.; Tsikrika, T.; Vrochidis, S.; Kompatsiaris, l. Host-based intrusion detection using signature-based and AI-driven anomaly detection methods. *Inf. Secur.* **2021**, *50*, 37–48. [CrossRef]
53. Shukla, A.K. Detection of anomaly intrusion utilizing self-adaptive grasshopper optimization algorithm. *Neural Comput. Appl.* **2021**, *33*, 7541–7561. [CrossRef]
54. Sarvari, S.; Sani, N.F.M.; Hanapi, Z.M.; Abdullah, M.T. An efficient anomaly intrusion detection method with feature selection and evolutionary neural network. *IEEE Access* **2020**, *8*, 70651–70663. [CrossRef]
55. Almaghthawi, Y.; Ahmed, I.; Alsaadi, F.E. Performance analysis of feature subset selection techniques for intrusion detection. *Mathematics* **2022**, *10*, 4745. [CrossRef]
56. Ellis, T.J.; Levy, Y. Towards a guide for novice researchers on research methodology: Review and proposed methods. *J. Issues Inf. Sci. Inf. Technol.* **2009**, *6*, 323–337. [CrossRef]
57. Zhou, Y.; Cheng, G.; Jiang, S.; Dai, M. Building an efficient intrusion detection system based on feature selection and ensemble classifier. *Comput. Netw.* **2020**, *174*, 107247. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.