A study on IDS system based on Hedge Algebras to detect DDOS attack in IoT systems

Trong Minh Hoang (hoangtrongminh@ptit.edu.vn)
Posts and Telecommunications Institute of Technology

Lan Nhu Vu
Thang Long University

Research Article

Keywords: Internet of Things, Intrusion Detection System, DDOS, Hedge Algebra, PSO algorithm.

Posted Date: November 4th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2125935/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
A study on IDS system based on Hedge Algebras to detect DDOS attack in IoT systems

Hoang Trong Minh\textsuperscript{1}* and Vu Nhu Lan\textsuperscript{2}

\textsuperscript{1}Telecommunication Faculty No1, Posts and Telecommunications Institute of Technology, Hanoi, Vietnam.
\textsuperscript{2}Informatics Faculty, Thang Long University, Hanoi, Vietnam.

\textsuperscript{*}Corresponding author(s). E-mail(s): hoangtrongminh@ptit.edu.vn; lanvu.cyber@gmail.com;

Contributing authors: hoangtrongminh@ptit.edu.vn; lanvu.cyber@gmail.com;

Abstract

In recent years, we have experienced the rapid and beneficial development of IoT solutions throughout all aspects of life. In addition to the obvious advantages, the increase in the number and variety of devices has resulted in more challenges on security issues. The DDOS attack, which originates from a broad range of sources and is a significant challenge for IoT systems, is one of the most prevalent but devastating attacks. IoT devices are typically simple and have few computing resources, putting them at risk of being infected devices and attackers. IDS intrusion detection systems are regarded to be the foremost line of protection against DDOS attacks. Therefore, the IDS system attracts many researchers and implements intelligent techniques such as machine learning and fuzzy logic to detect these DDOS attacks quickly and precisely. Along with the approach of intelligent computation, this study presents a novel technique for detecting DDOS attacks based on hedge algebra, which has never been implemented on IDS systems. We use the PSO swarm optimization algorithm to optimize the proposed model’s parameters for optimized performance. Our experiment on the IoT-23 dataset shows that the proposed model’s accuracy and performance metrics for DDOS attack detection are better than those proposed by other previous authors.

Keywords: Internet of Things, Intrusion Detection System, DDOS, Hedge Algebra, PSO algorithm.
A study on IDS system based on Hedge Algebras...

1 Introduction

IoT solutions have made significant contributions to human society with wide-ranging applications. In recent years, emerging technologies such as 5G have created more opportunities to deploy new IoT solutions and exponentially increase the number of IoT devices [1]. Due to the deep and broad impact of IoT applications on our life and the variety of IoT devices have created a series of new security challenges. Accompanied by innovative intelligent computing methods, threats and attacks are becoming increasingly sophisticated, dangerous, and devastating. Unfortunately, the majority of IoT devices that are geared towards simplicity and limited computing resources have become victims and sources of attacks. The above aspects have created new challenges in this field [2]. Distributed Distributed Services (DDoS) attacks considered as one of the classic attacks which are a major security threat against resource network availability. Nowadays, the large and diverse number of IoT devices equipped with operating systems working on distribute fashions are at risk of becoming a source of attacks. DDoS attacks use a huge number of geographically scattered compromised systems to launch their attacks, and wide variety of DDoS attacks have been identified for various domains of applications. Hence, DDOS in IoT applications causes massive unexpected danger and has been the topic of a number of recent studies. [3] [4].

IDS intrusion prevention systems have been widely regarded as the first barrier of defense against IoT cyber attacks, and are interesting tools that attempt to find and detect such cyber-attacks. Commonly, IDS systems based on signature-based and anomaly-based types and hybrid-type [5]. As a signature-based IDS observes the packets crossing the network, it compares these packets to a database of known attack signatures in order to identify any abnormal behavior. In contrast, anomaly-based intrusion detection systems can alert you to unknown suspicious behavior. Hence, anomaly detection solutions are widely used in diverse contexts and research fields[6]. In fact, today’s DDoS attack has become so dynamic and sophisticated because utilizing various of intelligent tools that that can bypass the detection system based on statistical probabilities or static thresholds solutions. Hence, the current IDS system uses variously intelligent computation approaches based on different intelligent computation methods to detect anomalies such as artificial neural networks, deep neural networks, Bayesian networks, genetic algorithms, fuzzy logic, Boltzmann Machine ... [7] [8] [9]. In which, data-driven approaches, such as IDS models utilizing deep learning neural networks, are frequently highly adaptive to the environment’s dynamics and offer great precision [10]. Unfortunately, these networks must optimize a large number of architecture characteristics; so that the DNN decisions are characterized by a high level of complexity due to a lack of experience and expert opinion. Otherwise, the heuristics and fuzzy approaches are commonly used to reduce the complexity of classification problems [11] [12]. However, to deliver great outcomes, it is essential to determine the appropriate rules. To tackle this challenge, we propose a novel, never-before-discussed model, namely the IDS based hedge algebra model to detect DDOS attacks...
in IoT environments. The proposed system’s parameters are automatically optimized via the particle swarm optimization (PSO) algorithm, thereby correcting the subjective errors of the expert on rule sets. Experimental results on the IoT-23 attack dataset using our proposed model demonstrate that DDOS attack detection results achieve DDOS attack detection rate is up to 99.99%.

The main contributions of this study include:

- Design a new model for IDS based on hedge algebra to detect DDOS attacks for IoT networks.
- Experimentally verify and compare the results with other proposals to highlight the proposal’s advantages.

The paper is organized as follows. The next section presents related work. The background of main theoretical issues in our proposal model are briefly described in the section 3. Our proposed model will be presented in Section 4. Experiment results on the IoT-23 dataset and related discussions are presented in Section 5. The conclusion and plan for our future investigations are presented in the final section.

2 Related work

The primary objective of IDS systems is to monitor traffic patterns for sudden and unexpected changes in traffic or other network traffic conditions [13]. Typically, these anomalies are the result of known or unknown security attacks on computer networks and their services. Methods for identifying anomalous network intrusions involve multiple processes, including preprocessing, feature selection/extraction, and classification. Specifically, classifications that rely on statistical methods can raise the error rate. In addition, the threshold between normal and abnormal behavior may not be well-defined, and even a slight change in monitored traffic might raise the rate of false positives and accordingly to a lowered attack’s detection accuracy.

Many popular data sets are available for IDS evaluation, such as BoT-IoT, IoT Network Intrusion, MQTT-IoT-IDS2020, IoT-23, and IoT-DS1 datasets [14]. In which, the dataset IoT-23 includes 20 malware captures executed in IoT devices and 3 captures for benign IoT device traffic [15]. This IoT network traffic was recorded in Stratosphere Laboratory, the AIC team, FEL, CTU University, and the Czech Republic. The IoT-23 dataset has been tested by several previous authors’ studies on the effectiveness of attack detection methods such as machine learning, deep learning, or fuzzy logic [16] [17] [18]. In [16], the authors evaluate several Artificial Neural Network class of algorithms such as Random Forest (RF), Naïve Bayes (NB), Multi Layer Perceptron (MLP) and get best results achieved by the Random Forest algorithm with a accuracy or 99.5%. A model for anomaly-based intrusion detection in IoT networks proposed in [17] uses a convolutional neural network (CNN) and gated recurrent unit (GRU) to classify IoT network data. the detection accuracy achieves as
99.5%. The authors in [18] have proposed a novel CNN-based anomaly detection model with better performance metrics than other previous IDS based CNN models while achieving the detection accuracy as 99.82%. We can see that, detection accuracy is dependent on a variety of parameters and how the data is processed, as is apparent from the given example proposals.

It is worth noting that several IDS using fuzzy logic offer several advantages to handling crisp boundary problems caused by unpredictable and uncertain conditions [19] [20]. However, we found that no fuzzy logic usage models have yet been evaluated on the IoT-23 dataset. In our previous works [21] [22], we proposed a Fuzzy Inference System-based IDS that enabled more efficient detection of DDOS attacks in wireless sensor networks than threshold approaches. However, the results are based solely on numerical simulations and not on specific data sets. Following the fuzzy logic approach, to formalize the order-based semantics of the words in the term-domain of linguistic variables, authors in [23] developed hedge algebras (HA), which can be used in a variety of application domains, such as information processing or intelligent control [24] [25]. Hedge algebras’ core element is that they capture the nature of fuzzy information by quantifying the qualitative semantics of linguistic concepts. Fuzziness measure, fuzziness interval of terms, and semantically quantifying mapping are the three quantitative features of HA (SQM). SQMs provide for the execution of a complete description and the logical and coherent demonstration of a rule set model and approximation inference process. Hedge algebra is used to model the domain of linguistic variables through quantitative semantics. Hence, the fuzzy rule system becomes the quantitative semantic rule system. Therefore, the inference problem becomes an interpolation problem with computational methods of low complexity. This paper will use the hedge algebra method applied to the IDS system to detect DDOS attacks in IoT networks. The experimental results on the IoT-23 dataset show the superior performance of the proposal compared with the previous suggestions of other authors.

3 Background

3.1 Hedge Algebra Overview

Following the definitions and properties of hedge algebras in [23], Hedge Algebra (HA) is a mathematical structure that has the order of collection of linguistic items. Several prominent features are listed below.

**Definition 1.** Hedge algebras of the linguistic variable X can be represented as an algebraic structure, the set of five components \( AX = (X, G, C, H, \leq) \). where \( X \) is a set of values of a linguistic domain regarded as a POSET (partially ordered set); \( G \) is a set of generators, which are designed as primary terms (semantic tendency expressions) denoted by \( c^- \) and \( c^+ \), \( G = \{c^-, c^+\}, c^- \leq c^+ \); \( C \) is a set of constants, \( C = \{0, W, 1\} \), (zero, neutral and unit elements, respectively); \( H \) is a set of urinary operations, \( H = H^- \cup H^+ \), \( H^- \) and \( H^+ \) is two
artificial hedges which are generated from $x$ by using operations in $H$; $\leq$ is a partial ordering relation on $X$.

**Definition 2.** Denote $fm$ is a fuzzy measurement of an element $x$ in $X$, and $\mu(h)$ is a fuzziness measure in the $H$ domain. [$X \to [0,1], \forall h \in H$] has the properties as bellow.

$$fm(c^+) + fm(c^-) = 1; \sum_{h \in H} fm(hx) = fm(x), \forall x \in X$$  

$$fm(0) = fm(W) = fm(1) = 0$$  

$$\mu(h) = \frac{fm(hx)}{fm(x)} = \frac{fm(hy)}{fm(y)}, \forall x, y \in X, h \in H$$  

$$\sum_{i=q,j \neq 0}^{p} fm(hx) = fm(x)$$  

$$fm(x) = fm(h_n h_{n-1} \ldots h_1 c) = \mu(h_n) \mu(h_{n-1}) \ldots \mu(h_1) c$$  

$$\sum_{i=-1}^{q} fm(h_i) = \alpha, \sum_{i=1}^{p} fm(h_i) = p \text{ with } \alpha, \beta > 0, \alpha + \beta = 1$$

**Definition 3.** The Semantically Quantifying Mapping (SQM) in HA is a function ($v$) which used to transform a linguistic variable into real linguistic values $v : X \to [0,1], \forall h \in H$.

$$v(W) = \theta = fm(c^-);$$

$$v(c^-) = q - afm(c^-) = bfm(c^-);$$

$$v(c^+) = q + afm(c^+) = 1 - bfm(c^+).$$

$$v(h_j x) = v(x) + \text{sgn}(h_j x) \left( \sum_{i = \text{sgn}(j)}^{j} fm(h_i x) - \varpi(h_j x) fm(h_j x) \right)$$

$$(h_j x) = \frac{1}{2} [1 + \text{sgn}(h_p, h_j)(\beta - \alpha)], -q \leq j \leq p, j \neq 0.$$  

### 3.2 Particle Swarm Optimization

Particle Swarm Optimization (PSO) models its behavior after animals’ swarming or flocking patterns [19]. It is very appealing because the simple conceptual framework and the analogy of birds flocking facilitated conceptual visualization of the search process. The basic PSO algorithm is shown in Figure 1. PSO has particles that make up its population, called a swarm. Each particle is moved from one location to another over mutation. This mutation is performed directly, in which each particle is moved from its previous location to a new, better location. The PSO algorithm has several advantages that make it an attractive optimization algorithm [26]:

- PSO is easy to set up and code.
PSO is controlled by only three parameters (inertia weight, cognitive ratio, and social ratio). A slight change in any of these three controlling parameters produces a difference in performance.

PSO is adaptable and can be combined with other optimization algorithms.

4 The proposed model

4.1 Preprocessing the dataset IoT23

The IoT-23 dataset classifies attacks into 16 categories: Part-Of-A-Horizontal-PortScan, Okiru, Okiru-Attack, DDoS, C&C-Heart Beat, C&C, Attack, C&C-, C&C- Heart Beat Attack, C&C-File download, C&C-Tori, C&C-Heart Beat File Download, C&C-Mirai. The data file has fifteen malicious layers and one secure layer. The preprocessed data is categorized into three sets: training, validation, and testing. The segmented technique guarantees equal training, validation, and testing samples from each processing layer.

Data processing is extracting network features from raw network traffic for training and evaluating models. These features are extracted from the conn.log.labeled file in the little IoT-23 dataset and exported to the file in CSV format. Because the model is designed for IoT networks, local network features such as stream ID, source IP address, destination IP address, and timestamp will be removed from the dataset. The index features of the dataset are coded and presented as binary data. The NaN values are substituted for 0. The featured columns are then normalized to a specified range [-1 1] to eliminate large values and speed up computation. Besides, we use feature selection
to improve accuracy and reduce noise for the model. The appropriate features are selected through the recursive feature elimination technique. For missing data values in a feature column, we deal with them by replacing them with the average value of each corresponding attack type in that feature. The principal components analysis (PCA) technique is applied to the data to help reduce complexity, overcome resource constraints, and increase the performance of anomaly detection models.

After being processed and reduced in dimension, we select the dataset’s secure (Benign) and DDoS patterns. This dataset is divided into 3 sets: Training-set, Validating-set, and Testing-set. The number of samples in the Training-set and the Testing-set are taken equally. The number of samples in the Training-set and Validating-set are in an 80/20 ratio. The training-set is a standard space for the following stages’ calculations. Figure 2 shows the data processing system.

![Fig. 2 Prepossessing the IoT23 data set](image)

### 4.2 The proposed model

The main components of the proposed model is illustrated on figure 3. To filter out training data for the distance matrix D-train, the input training dataset is calculated based on the neighbor data distance. The DtoN module represents the transformation between the distance matrix D and the density matrix N. D is the matrix of the distance of each point to the remaining points, and N is the two-column set containing the number of DDoS points and the number of benign points whose distance from each data point is less than or equal to r. where r is the optimal radius from any interesting data point in the data space to other data points. This yields the fuzzy data sets N-DDOS (x1) and N-Benign (x2). These data sets are denoted as L (Low) and H (High) density metrics. The script output (y) has presented the probability of DDOS attacks such as LA (Low attack), MA (Medium Attack), and HA (High Attack).

The Semantilization module is responsible for transforming linguistic variable domain $[a \ b]$ into their respective linguistic semantics value domain $[a_s \ b_s]$ over semantization process. If $a_s = 0$ and $b_s = 1$ we have a linear sematilization process. If it is not, we have a nonline sematilization process. Denote $x$ is a linguistic variable and $(x_h)$ semantic value, we describes linear sematilization process and non linear sematilization process as below.
A study on IDS system based on Hedge Algebras...

Fig. 3 The outline of the proposed model

Case 1: Linear sematization process.

\[
\text{LinearSematization}(x) = x_h = \frac{x-a}{b-a} \quad \text{(10)}
\]

\[
\text{LinearDeSemantization}(x_h) = a + (b - a)x_h
\]

Case 1: Non linear sematization process.

\[
\text{NonLinearSematization}(x) = x_h = f(x, sp) 
\]

\[
\text{NonLinearDeSemantization}(x_h) = x = g(x_h, dp) \quad \text{(11)}
\]

In which, \(sp\) is the nonlinear semantic parameter, \(sp \in [0, 1]\); \(dp\) is the nonlinear desemantization parameter, \(dp \in [0, 1]\). The function \(f(.)\) is continuous, co-variate and satisfies the condition below,

\[
f (x_s, sp) = x_s + sp \times x_s \, (1 - x_s) \\
0 \leq f (x_s, sp) \leq 1 \quad \text{(12)}
\]

The function \(f(.)\) is continuous, co-variate and satisfies the condition below,

\[
a \leq g (x, dp) \leq b; \\
g (x = a, dp) = a; \\
g (x = b, dp) = b. \quad \text{(13)}
\]

Note that semantic-based ordered structure is the key point of transform process as defined on definition 1. Assume \(x_i\) and \(x_k\) have ordered linguistic values as \(A(x_i)\) and \(A(x_i k)\), \(A(x_i) < A(x_k)\). Denote the qualitative linguistic value as \(A(x_*), \) we have \(A(x_i) \leq A(x_*) \leq A(x_k)\). Hence, we calculate the close degree
A study on IDS system based on Hedge Algebras... 9

\( \eta_i, \eta_k \) as the equation below.

\[
\eta_i = \frac{A(x_k) - A(x_*)}{A(x_k) - A(x_i)}; \quad \eta_k = \frac{A(x_*) - A(x_i)}{A(x_k) - A(x_i)} \tag{14}
\]

Where \( \eta_i + \eta_k = 1, 0 \leq \eta_i \leq 1, 0 \leq \eta_k \leq 1 \). The semantically Qualifying Measure (SQM) module includes the rules and algebras operators. SQM transforms linguistic semantic values into real values by algebras operators (6). The rules are illustrated as

1. \( \text{IF } x_1 = L \ \text{AND} \ x_2 = L \ \text{THEN} \ y = LA \)
2. \( \text{IF } x_1 = L \ \text{AND} \ x_2 = H \ \text{THEN} \ y = HA \)
3. \( \text{IF } x_1 = H \ \text{AND} \ x_2 = L \ \text{THEN} \ y = MA \)
4. \( \text{IF } x_1 = H \ \text{AND} \ x_2 = H \ \text{THEN} \ y = HA \) \tag{15}

The hedge algebra inference system uses piece-wise linear operators. The mathematics description of the model is expressed as.

\[
y_j = \frac{\sum_{j=1}^{L} \prod_{i=1}^{n} \eta_j(k_i)_j \times \phi_j(x_1...x_n)(p_{1j}...p_{1n})}{\sum_{j=1}^{L} \prod_{i=1}^{n} \eta_j(k_i)_j} \tag{16}
\]

Where, Input set is \( X = [x_1, x_2, ...x_n] \); a rule set is \( R = [1, 2, ...M] \); the parameters of rule \( j^{th} \) is \( (p_{0j}, p_{1j}...p_{nj}) \); \( A_{ij}(k_i) \) are the linguistic values in the rule \( j^{th} \) with its membership \( \mu_{k_i}(x_i) \); the fuzzy set is \( K_i = [k_1, k_2, ...k_K] ; \phi_j(.) \) is a linear function. From equations (10) (11) (15) (16), we construct a model includes the linguistic domains and variables and their semantic structure elements. Instead of performing fuzzification and defuzzification as in fuzzy logic, we adopted a simple method which termed as semantization and desemantization. We have a set of semantic transformation equations as follows:

1. \( L_{x_1} = \theta_1 \times (1 - \alpha_1) \)
2. \( H_{x_1} = \theta_1 + \alpha_1 \times (1 - \theta_1) \)
3. \( L_{x_2} = \theta_2 \times (1 - \alpha_2) \)
4. \( H_{x_2} = \theta_2 + \alpha_2 \times (1 - \theta_2) \) \tag{17}
5. \( LA = \theta \times (1 - \alpha) \)
6. \( MA = \theta \)
7. \( HA = \theta + \alpha \times (1 - \theta) \)

The PSO algorithm finds the optimal varies \( (\alpha, \alpha_1, \alpha_2, \theta, \theta_1, \theta_2) \) and \( r \) over Mean squared error (MSE) metric. The objective of the model is to use the PSO algorithm to find the membership function parameters of the two inputs and the optimal radius \( r \) so that the following cost function \( J \) is minimized:
A study on IDS system based on Hedge Algebras...

\[ J = \frac{1}{n} \sum_{i=0}^{n} (F(x^{(i)}) - y^{(i)}) \]  
(18)

\[ \text{fitness}^{(k)} = \frac{100}{J^{(k)} + 1} \]  
(19)

Where:

- \( n \) is the total number of samples of the test set.
- \( x^{(i)} \) is the \( i^{th} \) sample in the test set.
- \( y^{(i)} \) is the label of the \( i^{th} \) sample in the test set.
- \( F(x^{(i)}) \) is the predictive output of the model given the input is the \( i^{th} \) sample of the test set.
- The fitness value is used to feed into the PSO algorithm to update the position for each individual.

4.3 The core operations of the proposed model

The proposed model is operated by two phases such as initiated phased and optimized phase as below.

**Initiated phase**

The initial radius \( r \) is set at random between 0 and the set \( D \) train mean value that initiates the input member functions’ parameters. Four parameters are used to represent each membership function of a trapezoidal function with two inputs \((a, b, c, d)\) as in Figure 4. However, in order to make these 4 parameters compatible with the PSO algorithm, we encode them into a set of 3 parameters \((x, y, and z)\) as follows:

\[ a = x = y - z; \quad b = x - y; \quad c = x + y; \quad d = x + y + z. \]  
(20)

**Optimized phase**

This phase used to optimize the HA parameters of the proposed model, It...
illustrated by the proposed algorithm as below.

**Function**: hedge algebra parameter optimization

**Input**: IoT-23 dataset, encoded parameters

**Output**: $\alpha, \alpha_1, \alpha_2, \theta, \theta_1, \theta_2$

1. Calculate $D_{train}$ matrix from Training-set
2. For $i = 1$ to pop-size
3. Initialize position and velocity for individual $i$, $r := [0, \text{Avg}_{D_{train}}]$ $\theta := [0.25, 0.75]$, $\alpha := [0.2, 0.8]$ $\theta_1 := [0.25, 0.75]$, $\alpha_1 := [0.2, 0.8]$ $\theta_2 := [0.25, 0.75]$, $\alpha_2 := [0.2, 0.8]$
4. Calculation of Fitness score for individual $i$
5. Update the maximum values of the individual $i$ and the best values of the population.
6. End for
7. For $i = 1$ to $\text{max}$ iterative
8. For $j = 1$ to $\text{max}$ population size
9. Determine the individual $j$’s velocity value. $[v(t + 1) = w.v(t) + c_1.r_1.(pbest(t) - x(t)) + c_2.r_2.(gbest(t) - x(t))]$
10. Update new location for individual $j$ $x(t + 1) = x(t) + v(t + 1)$
11. Recalculate fitness points for an individual $j$
12. Update the maximum values of the individual $j$ and the best values of the population.
13. End for
14. End for
15. Choosing the best individual of the population is the optimal solution for the model.

**End function**

5 Experimental and Evaluation

5.1 Experimental parameters

All experiments of IDS using HA are conducted using Matlab. Training set and Validating set are applied to optimize the HA through the PSO optimization algorithm. The parameters of the PSO algorithm used in our simulation are shown in Table 1. The system is optimized and is evaluated by the IoT-

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>50</td>
</tr>
<tr>
<td>constriction factor</td>
<td>$c_1 = 2, c_2 = 2$</td>
</tr>
<tr>
<td>Inertia factor</td>
<td>1</td>
</tr>
<tr>
<td>Inertia factor reduction rate</td>
<td>0.99</td>
</tr>
<tr>
<td>Number of rounds</td>
<td>50</td>
</tr>
</tbody>
</table>
A study on IDS system based on Hedge Algebras...

23 dataset with varied processing methods. The system is evaluated through the parameters including accuracy, precision, recall, F1-score, and FPR (False Positive Rate). The accuracy is calculated as the fraction of correct predictions over the total number of predictions. The parameter is formulated as follows:

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{21}
\]

The fraction of correctly identified positives defines the precision. It is written as

\[
\text{Precision} = \frac{TP}{TP + FP} \tag{22}
\]

The recall is the fraction of actual positives that were correctly identified. The parameter could be expressed as

\[
\text{Recall} = \frac{TP}{TP + FN} \tag{23}
\]

The F1-score can be interpreted as a harmonic mean of the Precision and Recall. Its formula is

\[
F1 \cdot \text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{24}
\]

Finally, FPR is a measure of accuracy for a test. It is calculated as the ratio between the number of negative events wrongly categorized as positive (false positives) and the total number of actual negative events (regardless of classification).

\[
\text{FPR} = \frac{FP}{FP + TN} \tag{25}
\]

5.2 Experimental results

First, we evaluate the convergence speed of the PSO algorithm in the optimization problem of membership input parameters for the two boundaries acc and val. As shown in Figure 5 below, a very small number of loops (≤ 10 rounds) has resulted in convergence and stable. This shows that the time complexity of the algorithm is small and efficient. To examine the efficacy of the model with varied samples size and number of features, we apply the IoT-23 dataset to the attack detection model in the following manner. Use sampling at random to separate the test data sets (2500 samples and 5000 samples). We intend to conduct experiments with a modest sample size to enable IDS systems that can be deployed at the network’s edge, where computational resources are constrained. Using PCA method, we reduce the original IoT-23 dataset with varying numbers of features (10 features, 23 features and full 32 features). Due
to the minimal number of computational spatial dimensions, the small number of features reduces the computational complexity of a training model. We compare the efficiency of two linear and linear semantics schemes in Figure 6. The test results on different sample sets with different number of features show the adaptability of the nonlinear scheme. better online. The results show that the DDOS attack detection accuracy of the nonlinear semantic scheme is higher than that of the linear semantic scheme.

The results of the performance parameter evaluation of the model with our tests are presented in Table 2. We recognizes that the attack detection accuracy is not significantly reduced with small sample size and small number of features.

<table>
<thead>
<tr>
<th>Features</th>
<th>Samples</th>
<th>Acc</th>
<th>Pre</th>
<th>Recall</th>
<th>fpr</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 features</td>
<td>5000</td>
<td>0.996</td>
<td>0.997</td>
<td>0.995</td>
<td>0.0028</td>
<td>0.996</td>
</tr>
<tr>
<td>10 features</td>
<td>2500</td>
<td>0.994</td>
<td>0.993</td>
<td>0.995</td>
<td>0.0063</td>
<td>0.994</td>
</tr>
<tr>
<td>23 features</td>
<td>5000</td>
<td>0.996</td>
<td>0.996</td>
<td>0.996</td>
<td>0.0032</td>
<td>0.996</td>
</tr>
<tr>
<td>23 features</td>
<td>2500</td>
<td>0.995</td>
<td>0.912</td>
<td>1</td>
<td>0.0008</td>
<td>0.995</td>
</tr>
<tr>
<td>32 features</td>
<td>5000</td>
<td>0.997</td>
<td>0.996</td>
<td>0.999</td>
<td>0.0039</td>
<td>0.997</td>
</tr>
<tr>
<td>32 features</td>
<td>2500</td>
<td>0.993</td>
<td>0.997</td>
<td>1</td>
<td>0.0126</td>
<td>0.993</td>
</tr>
<tr>
<td>32 features full</td>
<td></td>
<td>0.998</td>
<td>1</td>
<td>0.995</td>
<td>0.0000</td>
<td>0.998</td>
</tr>
</tbody>
</table>

To evaluate the effectiveness of the proposal, we compare the results achieved with other proposals Table 3. The results showed that the proposed model gave reasonable results. Although the results of DDOS attack detection accuracy in our proposed model are not very outstanding, our model can work on small data sets with an accuracy of up to 99%. Hence, a significant contribution in our proposed model is that we have used a small number of
samples in the dataset and still have a higher attack detection rate then it can be deployed in limited resource devices.

6 Conclusion

In this paper, we have proposed a novel DDOS attack detection model for IDS system at edge computing. Specifically, the rule systems are processed by the hedging algebra method and the input variable member functions are optimized through the PSO algorithm. Our proposed model using fuzzy approach of input parameters for data points in IoT23 dataset has resulted in false positive avoidance and optimal classification by PSO algorithm. Hence, our proposed model can bring a good DDOS detection rate. Moreover, Our model combined with dimensionality reduction and data division techniques shows attack detection as high as 99 % with varied small data sets. Therefore, it will be a favorable condition to deploy the IDS system on the edge device and greatly reduce the harmful effects of attacks on IoT applications. We will continue to deploy our model on practical devices.
Declarations

Competing interests
The authors declare no conflict of interest.

Funding
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Availability of data and materials
The data that support the findings of this study are available from the corresponding author upon reasonable request.

Authors’ contributions
Trong-Minh Hoang: Conceived and designed the algorithms; Analyzed and interpreted the data; Wrote the paper. Nhu-Lan Vu: Performed the experiments.

References


A study on IDS system based on Hedge Algebras...


