A Lightweight Anomaly Detection Model using SVM for WSNs in IoT through a Hybrid Feature Selection Algorithm based on GA and GWO

Azam Davahi\textsuperscript{a}, Mahboubeh Shamsi\textsuperscript{b,∗}, Golnoush Abaei\textsuperscript{c}

\textsuperscript{a}Department of Computer Engineering, Qom Branch, Islamic Azad University, Qom, Iran.
\textsuperscript{b}Faculty of Electrical and Computer Engineering, Qom University of Technology, Qom, Iran.
\textsuperscript{c}Faculty of Electrical, Computer, and Biomedical Engineering, Shahabdanesh University, Qom, Iran.

\textbf{A R T I C L E  I N F O.}

Article history:
Received: 22 October 2019
Revised: 10 February 2020
Accepted: 24 April 2020
Published Online: 10 July 2020

Keywords:

\textbf{A B S T R A C T}

As a result of an incredibly fast growth of the number and diversity of smart devices connectable to the internet, commonly through open wireless sensor networks (WSNs) in internet of things (IoT), the access of attackers to the network traffic in the form of intercepting, eavesdropping and rebroadcasting has become much easier. Anomaly or intrusion detection system (IDS) is an efficient security mechanism, however despite the maturity of anomaly detection technologies for wired networks, current technologies with high computational complexity are improper for resource-limited WSNs in IoT and they also fail to detect new WSN attacks. Furthermore, dealing with the huge amount of intrusion wireless traffic collected by sensors, causing slow detecting process, higher resource usage and inaccurate detection. Hence, considering WSN limitations for developing an IDS in IoT, establishes a significant challenge for security researchers. This paper proposes a new model to develop a support vector machine (SVM)-based lightweight IDS (LIDS) using combination concepts of genetic algorithm (GA) and mathematical equations of grey wolf optimizer (GWO) which is called GABGWO. The GABGWO through applying two new crossover and mutation operators tries to find the most relevant traffic features and eliminate worthless ones, in order to increase the performance of the LIDS. The performance of LIDS is evaluated using AWID real-world wireless dataset under two scenarios with and without using GABGWO. The results showed a promising behavior of the proposed GABGWO algorithm in choosing optimal traffics, decreasing the computational costs and providing high accuracies for LIDS. The hybrid algorithm is also compared to pure GA and GWO and other recent methods and it is found that its performance is better than them.

\textcopyright 2020 JComSec. All rights reserved.
1 Introduction

A network of things that is equipped with sensors and try to exchange data commonly through open wireless sensor networks (WSNs) is called internet of things (IoT) \[1\]. Beside of many advantages which has been provided by IoT, these networks have some security challenges \[2,3\] and face some new attacks \[4,5\]. Anomaly detection system or intrusion detection system (IDS) is an efficient security technique that by gathering traffic data and analyzing them identifies the attacks \[6,8\]. The numerous existing IDS techniques in wired networks are not able to handle the zero-day wireless intrusions in IoT \[9,10\]. Additionally, most traditional IDSs have high computational costs and don’t consider resource limitations in wireless networks. Therefore, the advancement of a suitable and lightweight IDS for WSNs in IoT is necessary \[12,13\] such that it decreases computational costs and increases detection accuracies \[14,15\].

Based on the analysis and detection technique there are two main IDSs methods including misuse and anomaly detection. Anomaly detection methods are the most popular intrusion detection techniques in which malicious behaviors are distinguished from normal behaviors. In anomaly detection methods, there is no need to database for saving malicious traffics and they are able to recognize new attacks unlike misuse-based methods \[12,16,17\]. These methods usually are developed by machine learning algorithms or classifiers such as support vector machine (SVM), Artificial neural network (ANN), k-nearest neighbor (KNN), and so on, which among them SVM is the most successful classifier in this area \[15,18,19\]. Nevertheless, pervasiveness IoT devices that connected to a Wi-Fi network result in generate a huge number of wireless traffic data, which some of them are worthless and redundant that not only increase the computational costs, but also decrease the performance of the IDS \[9,20\].

A solution that can effectively solve the high dimensionality problem is feature selection (FS) mechanism. A FS mechanism through the optimal features and eliminating the extra and useless features leads to dimensionality reduction \[21,22\]. FS methods generally are divided to three categories, which are called filter, wrapper, and hybrid \[23,24\]. The filter techniques evaluate the chosen features based on the data characteristics while the wrapper and hybrid techniques exploit a classifier for assessing the chosen features unlike the filter \[22,24\]. The wrapper mechanisms have shown more successful accordingly are the most popular \[22,23,25\]. Investigating all possible feature subsets in large search space such as WSNs space makes FS to a NP-hard problem \[26,27\], thus an effective global search technique is required to solve the FS problem optimally \[26\]. Nowadays metaheuristic optimization algorithms have shown satisfactory capabilities to handle FS problem \[28,29,30\]. These algorithms’ efficiency is chiefly affected by their exploration and exploitation abilities \[29,31\].

Exploration is the process of looking for good solutions in the whole search space whereas exploitation is probing a limited region of the search space with the hope of improving the achieved solutions through exploration. Therefore, a metaheuristic algorithm to achieve high-quality solutions needs to make a balance between these two processes \[29,30\]. There are various metaheuristic algorithms that applied to FS that among them genetic algorithm (GA) as an old algorithm and grey wolf optimizer (GWO) as a new algorithm have shown excellent abilities in this domain \[29,32,33\]. Despite the many successfulness each of the two algorithms face weaknesses that can be solved by hybridization them with each other \[34,35\]. In this study is presented a lightweight anomaly detection model using SVM for WSNs in IoT (termed as LIDS) through a hybrid feature selection algorithm based on GA and GWO that named GABGWO. The main contributions of the proposed model include the following:

1. Applying SVM classifier to distinguish between anomalous wireless traffic from normal wireless traffic with goal of development of the LIDS.
2. GABGWO improves the efficiency of the SVM prediction model by means of recognition of the most relevant and informative wireless traffic intelligently.
3. AWID as a real Wi-Fi traffic dataset is used for experimentation and validation purposes. Data preprocessing technique was carried out to accelerate the speed of searching by the GABGWO and to transform the AWID dataset into a compatible format supported by the SVM classifier and as a result obtain the accurate result of the proposed LIDS.
4. The proficiency of the GABGWO hybrid algorithm in advancing LIDS is judged under two scenarios. The different performance measures used for evaluations consisting of the number of selected features (SF), accuracy (ACC), f-score (F1), recall (R), precision (P), false alarm rate (FAR), and computational times (CTs) that includes the time used by LIDS to detect intrusions (LIDS_Time) and the time used by GABGWO...
The remainder of this paper is structured as follows: Section 2 presents the related works. In Section 3, a preliminary background of the GA and GWO is given. Section 4 explains the proposed model, which consists of data preprocessing, combining GA and GWO to FS, and anomaly classification major stages. The experiments and their setup and results are presented in Section 5 and Section 6, respectively. Finally, Section 7 concludes the paper with future works.

2 Related Works

The literature review shows that there are different researches in which used the various metaheuristic algorithms to advance and improve the performance of the classifier-based anomaly detection and other application fields. In this section, firstly the existing researches in other applications are overviewed. Secondly, the existing researches in anomaly detection application are summarized in Table 1.

2.1 Overview of Existing Researches for Other Applications

The authors in [27] proposed a new wrapper FS approach based on a new artificial bee colony (ABC) optimizer and KNN classifier that integrated with a non-dominated sorting process and genetic operators. They performed two different implementations that include binary ABC and continuous ABC. Their method was examined on 12 various datasets. The results in compared with other methods showed that their binary approach outperformed the other methods in terms of both ACC and CTs.

Two binary versions of WOA with KNN was proposed to do the FS for classification purposes [59]. In the first version, the authors aimed to study the efficiency of using the tournament and roulette wheel selection techniques instead of using a random operator for searching process. Then, they used crossover and mutation operators to improve the exploitation of the WOA. Their evaluations are performed on 20 UCI datasets and then compared to ant lion optimizer (ALO), GA and particle swarm optimization (PSO), and five filter FS techniques. The evaluation results with ACC, SF and CTs metrics showed the efficiency of this approach.

The authors have developed two binary versions of GWO with KNN for FS in paper [60], and named the binary versions BGWO1 and BGWO2 respectively. In the BGWO1, while other agents moved toward the first three best agents also binaries, then a probabilistic crossover was performed between them (the three basic moved) to find the updated binary grey wolf position. In BGWO2, a sigmoidal function to binaries has been used to update the grey wolf position vector only and these values randomly threshold to find the updated binary grey wolf position. The performed evaluations over 18 various datasets from the UCI repository were indicative of a good ability of the proposed BGWOs in terms of ACC and SF measures.

Two different approaches for presenting different binary versions of ALO to FS problem were developed in [61]. The first approach only is based on ALO operators while the continuous stages were threshold via a threshold function in the second one. A set of evaluations based on 21 UCI datasets and with SF, ER and CTs metrics, proved efficiency of the proposed method in compared to GA, PSO, and bat algorithms.

A new FS method through multi-objective GWO was proposed in [62]. In this method to find a subset of features with minor redundancy, at the first steps was used filter technique. Then at the later steps was employed wrapper technique with KNN machine learning algorithm. The performed assessments of this method over different 8 UCI datasets and against GA and PSO and other single-objective methods showed that the proposed multi-objective method achieved better ACC.

Another wrapper FS method based on GWO and KNN was presented in [63]. In this work, GWO tries to find the optimal features of the complex features space via the interaction of agents in the population. The result of the implementation of the proposed method in this work against GA and PSO over 8 UCI datasets proved that the presented approach provides better performance in both SF and ACC.

In [64] the researches have been used firefly algorithm (FFA) for presenting an efficient FS method. Their method was called return-cost-based binary FFA (Rc-BBFA). Their mechanism was conducted in three steps. First, they measured a firefly’s attractiveness from other ones by an indicator based on the return-cost. Second, in order to search the attractive one for every FFA, they proposed a Pareto dominance-based method. Third, to update the position of a FFA they developed a binary operator based on the return-cost attractiveness and the adaptive jump. They showed that their method with providing the better ACC, F1 and CTs outperformed other FS methods that are based on GA, PSO, and traditional FFA through a set of experiments over 10 UCI datasets.

By using a multi-objective PSO algorithm and KNN, a FS method to choose the unreliable data was advanced in [65]. In this study, an efficient multi-objective FS algorithm is developed by providing two
<table>
<thead>
<tr>
<th>S. No.</th>
<th>Author &amp; Year</th>
<th>Metaheuristic Algorithm</th>
<th>Machine Learning Algorithm</th>
<th>Traffic Type</th>
<th>Dataset</th>
<th>Performance Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Tao et al. 2018 [20]</td>
<td>GA</td>
<td>SVM</td>
<td>Wired</td>
<td>KDDcup99</td>
<td>R, ER, FAR, CTs</td>
</tr>
<tr>
<td>3</td>
<td>Senthilnayaki et al. 2015 [37]</td>
<td>GA</td>
<td>SVM</td>
<td>Wired</td>
<td>KDDcup99</td>
<td>ACC, SF</td>
</tr>
<tr>
<td>4</td>
<td>Ahmad et al. 2014 [35]</td>
<td>GA</td>
<td>SVM</td>
<td>Wired</td>
<td>KDDcup99</td>
<td>R, FAR, SF, CTs</td>
</tr>
<tr>
<td>8</td>
<td>Desale et al. 2015 [42]</td>
<td>GA</td>
<td>NB, J48</td>
<td>Wired</td>
<td>NSL-KDD</td>
<td>ACC, CTs, SF</td>
</tr>
<tr>
<td>9</td>
<td>Senthilnayaki et al. 2013 [33]</td>
<td>GA</td>
<td>J48</td>
<td>Wired</td>
<td>KDDcup99</td>
<td>R, ER, CTs</td>
</tr>
<tr>
<td>11</td>
<td>Alzubi et al. 2019 [45]</td>
<td>GWO</td>
<td>SVM</td>
<td>Wired</td>
<td>NSL-KDD</td>
<td>ACC, R, FAR, SF</td>
</tr>
<tr>
<td>12</td>
<td>Sathish et al. 2017 [46]</td>
<td>GWO</td>
<td>SVM</td>
<td>Wired</td>
<td>KDDcup99</td>
<td>ACC, FAR</td>
</tr>
<tr>
<td>14</td>
<td>Roopa Devi and Suganthe 2017 [48]</td>
<td>GWO</td>
<td>SVM, NB</td>
<td>Wired</td>
<td>NSL-KDD</td>
<td>ACC, R, P, F1, SF, CTs</td>
</tr>
<tr>
<td>15</td>
<td>Seth and Chandra 2016 [49]</td>
<td>GWO</td>
<td>ANN</td>
<td>Wired</td>
<td>NSL-KDD</td>
<td>ACC, SF</td>
</tr>
<tr>
<td>16</td>
<td>Davahli et al. 2020 [50]</td>
<td>GA, GWO</td>
<td>SVM</td>
<td>Wireless</td>
<td>AWID</td>
<td>ACC, R, FAR, SF, P, F1, CTs</td>
</tr>
<tr>
<td>17</td>
<td>Roopa Devi et al. 2018 [51]</td>
<td>GWO, Cuckoo Search Optimization (CuSO)</td>
<td>SVM</td>
<td>Wired</td>
<td>NSL-KDD</td>
<td>ACC, R, TNR, P</td>
</tr>
<tr>
<td>18</td>
<td>Mazini et al. 2019 [52]</td>
<td>ABC</td>
<td>AdaBoost</td>
<td>Wired</td>
<td>NSL-KDD, ISCXIDS2012</td>
<td>R, ACC, FAR, SF, CTs</td>
</tr>
<tr>
<td>21</td>
<td>Li et al. 2018 [54]</td>
<td>Bat</td>
<td>RF</td>
<td>Wireless</td>
<td>KDDcup99</td>
<td>ACC, R, P, F1, FAR, CTs</td>
</tr>
<tr>
<td>22</td>
<td>Usha and Kavitha 2017 [19]</td>
<td>PSO</td>
<td>SVM</td>
<td>Wireless</td>
<td>AWID</td>
<td>ACC, R, P, F1, FAR, CTs</td>
</tr>
<tr>
<td>23</td>
<td>Bostani et al. 2017 [55]</td>
<td>Gravitational Search Algorithm (GSA)</td>
<td>SVM</td>
<td>Wired</td>
<td>NSL-KDD</td>
<td>ACC, R, FAR, SF, CTs</td>
</tr>
<tr>
<td>25</td>
<td>Bamakan et al. 2016 [57]</td>
<td>PSO</td>
<td>SVM, Multiple Criteria Linear Programming (MCLP)</td>
<td>Wired</td>
<td>KDDcup99</td>
<td>ACC, R, FAR, SF</td>
</tr>
<tr>
<td>26</td>
<td>Eesa et al. 2015 [58]</td>
<td>Cuttle Fish Algorithm (CFA)</td>
<td>Decision Tree (DT)</td>
<td>Wired</td>
<td>KDDcup99</td>
<td>ACC, R, FAR</td>
</tr>
</tbody>
</table>
new operators for PSO. The first operator was designed to overwhelm the decline phenomenon of particles. Hybrid mutation as another operator was designed to enhance the suggested algorithm’s search ability. Experimental and comparison results of the proposed method over 6 UCI datasets proved that this algorithm is highly competitive in terms of ACC and SF with unreliable data.

Through combining a “chaotic” version of ALO with KNN was proposed another metaheuristic-based FS approach. This method tried to control the exploration and exploitation rate by means the parameter I. The approach was examined using 18 UCI datasets with different number of features. The results of examination which were compared with the GA and PSO using ACC, F1 and SF quality metrics, proved that this method provides better performance [66].

The researchers in work [67] aimed to advance a PSO-KDE model in which PSO optimizer was used to simultaneously determine the kernel bandwidth and select the optimal features for kernel density estimation (KDE) classifier. They also used the classification performance and number of selected features to advance the fitness function for PSO-KDE. They evaluated the performance of their method by employing two datasets including wisconsin breast cancer dataset (WBCD) and wisconsin diagnosis breast cancer database (WDBC) using ACC, R and TNR. Evaluating results has demonstrated that the PSO-KDE method performs better than GA-KDE method in average, in diagnosis of breast cancer.

It is noteworthy that in addition to above-mentioned single metaheuristic algorithm-based FS, there are several researches in which a hybrid metaheuristic algorithm-based was presented to FS for general purposes such as [34, 35]. In paper [34], have presented two hybridization models to design different FS mechanisms through whale optimization algorithm (WOA) and simulated annealing (SA) algorithm with KNN. In their first method, SA is embedded in WOA, whereas in their second method, WOA is embedded in SA whereas it is applied to enhance the best solution found after each run of WOA in their second method. Their main purpose of applying WOA was to improve the exploitation by probing the most promising regions that founded by WOA. The evaluations on 18 UCI datasets and in compared with other techniques confirmed the efficiency of the proposed method in terms of ACC, SF and CTs.

To solve FS problem a new hybrid algorithms based on binary bat and PSO algorithms and KNN classifier have been proposed in [35]. The hybrid algorithm that was called HBBEPSO, combined the bat optimizer with its capacity for helping explore the search space and the PSO with its ability to converge to the best global optimal in the features space. The performance of the HBBEPSO was investigated in compared with the bat, PSO and other algorithms which have been applied to FS. This method was evaluated over 20 UCI datasets and in terms of SF and F1 evaluation metrics. The assessment results proved the capability of HBBEPSO algorithm to search the search space for optimal feature subsets.

2.2 Summarization of Existing Researches for Anomaly Detection

Each work that presents a metaheuristic-based FS method in anomaly detection field is categorized regarding the following characteristics: the authors name and publishing year, the applied metaheuristic algorithm, the utilized machine-learning algorithm, the type of network traffic, the evaluation network dataset and the validation metrics. It should be noted that the order of the studies is as follows: first, they are ordered based on the applied metaheuristic algorithm, second the utilized classifier and third the year of publication. As it can be seen in Table 1, it is tried to place researches which are recently presented and based on GA and GWO with SVM classifier at the top of the table.

3 Background

In this section, an overview of the main processes and characteristics of GA and GWO is given.

3.1 Genetic Algorithm

A genetic algorithm is one of the oldest methods for solving optimization problems. It was proposed by Holland in [68–70]. This algorithm tries to mimic animals and human genes behavior and the new generation production process. The algorithm has three main steps:

- Selection: In this step, it is tried to select some good and high-fitness solutions as parents. The details of selecting parents are proposed in different works in various ways. But for producing new solutions, we need to choose good solutions as parents in GA.
- Crossover: After choosing (usually two) parents in the selection, two new offsprings (new solutions) is produced in this step. A cross point (or two) is selected randomly, and two new solutions are produced by choosing the value of problem parameters from each of the parents on each side of the cross point. This step implemented by the probability of $P_c$.
- Mutation: To increase the chance of reaching new
and unseen solutions, in this step we choose a parameter (or more) randomly, and change its value to other possible search space values. Again this step is carried out by probability of $P_m$.

### 3.2 Grey Wolf Optimizer

The GWO algorithm was proposed by Mirjalili recently [71], and during these few years, it has gained more attention by computer science researchers [72, 73]. The main idea of this algorithm is obtained from kinds of wolves and their behavior for praying. As an algorithmic method, it tries to show some mathematical behavior like wolves. There exist four kinds of wolves which are called alpha ($\alpha$), beta ($\beta$), delta ($\delta$), and omega ($\omega$). In GWO, every one of candidate solutions (omega wolves) tries to be closer in a direction at a position that is produced by current positions of alpha, beta, and delta wolves. More mathematically, these steps are occurred for every solution to produce a new solution:

$$X[t + 1] = X[T] - A \times D$$  \hspace{1cm} (1)

$X[t]$ denotes current position and $X[t+1]$ denotes the next position of each wolf, $D$ as a difference vector is calculated based on Eq. (2).

$$D = CX_p[t] - X[t]$$  \hspace{1cm} (2)

$A$ and $C$ are coefficient vectors that are estimated based on Eq. (3) and Eq. (4) respectively.

$$A = 2ar_1 - a$$  \hspace{1cm} (3)

$$C = 2r_2$$  \hspace{1cm} (4)

The value of $r_1$, $r_2$ vectors are generated between 0 and 1 randomly and a vector’s value is decreased from 2 to 0 through Eq. (5) in each iteration.

$$a = 2 - t\left(\frac{2}{\text{MaxCycle}}\right)$$  \hspace{1cm} (5)

$t$ is the number of the current round and MaxCycle is the total number of rounds.

The following equations (6) to (8) are used for updating the position of omega wolves based on the positions of alpha, beta, and delta wolves:

$$X[t + 1][i] = \frac{X_1[t][i] + X_2[t][i] + X_2[t][i]}{3}$$  \hspace{1cm} (6)

$t$ represents the current round, $i$ denotes the index of variables and the $X_1$, $X_2$ and $X_3$ as ith omegas’ positions at the ith round, are initialized by Eq. (7).
Figure 2. Pseudo-Code of Hybrid GABGWO Algorithm.

\[ X_{\alpha}[i] = X_{\alpha}[i] - A_1[i]D_{\alpha}[i], \]
\[ X_{\beta}[i] = X_{\beta}[i] - A_2[i]D_{\beta}[i], \]
\[ X_{\delta}[i] = X_{\delta}[i] - A_3[i]D_{\delta}[i], \]  
(7)

\[ Z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \]  
(9)

Where \( Z_i \) is considered as the normalized value, \( X_i \) represents the corresponding feature value in which \( i \) can initialize from 1 to 154, and \( \min(x) \) and \( \max(x) \) are the minimum and maximum values of the feature \( x \), respectively. In the fourth step, to avoid outnumbering the attack instances by normal instances, the dataset with the ratio 1-to-1 between normal and anomaly records is balanced.

Finally, because of the huge amount of the AWID dataset in the fifth step, 10000 instances as the training dataset are randomly sampled that consist of various intrusions, afterward, 5000 instances as testing dataset are sampled.

4.2 The Wrapper Feature Selection Stage

Because the nature of intrusion wireless traffic is nonlinear the LIDS faces the huge amount of network data that some amount of them are redundant and irrelevant features causing slow training and testing procedure, higher resource usage, and inaccurate detection. Hence, selecting the informative and optimal features for LIDS to achieve detection with low computational costs is crucial. Hence, an effective wrapper FS hybrid algorithm (GABGWO) with two novel operators namely crossover and mutation based on concepts of GA and mathematical equations of GWO is presented in this stage. GABGWO’s pseudo-code is given in Figure 2. The steps related to combination

4.1 The Data Preprocessing Stage

Generally, the preprocessing stage is conducted in five main steps: In the first step, the question marks are replaced with zero value. In the second step, the features that have nominal value e.g. source address or initialization vector (IV) with hexadecimal value are changed into an integer value. In the third step, the continuous values i.e. timestamps are normalized between zero and one, according to Eq. (9).

4 The Suggested Model

To advance a lightweight anomaly detection model by considering the resource limitation problem in wireless networks, as Figure 4 illustrates three main stages including (1) data preprocessing stage, (2) wrapper feature selection stage, and (3) the anomaly classification stage are integrated.

According to the Figure 4 each stage itself has some steps which are described in detail as below:
GA and GWO are presented in detail in the following steps:

Step 1- Initialization the parameters the parameter setting is one of the processes that affect the performance of the new hybrid GABGWO algorithm, like any optimization algorithm. In general, process of parameters setting of the random meta-heuristic algorithms is performed by considering the algorithm’s application or optimization problem. In this step, the initialization of all GABGWO’s parameters is empirically carried out in a way that a trade-off between the number of selected features and accuracy is established. Table 2 shows the initialization of all GABGWO’s parameters.

Step 2- Generating a primary search space This process creates a primary population of candidate solutions (feature subsets) by sets of ‘1’ and ‘0’ bits as seen in Figure 1. Each candidate solution will present a possible feature subset. The bit (gene) with value ‘1’ represents the feature is selected, and ‘0’ represents the feature is not selected. The \( N \) denotes the total number of candidate solutions, and \( D \) denotes the dimension of the candidate solutions that initialization of them is given in Table 2. Afterward, the following processes are applied over the solutions until a features subset deemed optimal is found or \( MaxCycle \) is satisfied. \( MaxCycle \) is the maximum number of cycles to find the best solutions among of population that is considered as termination criterion and initialized in Table 2.

Step 3- Evaluating the quality of the solutions GABGWO via its fitness function assesses the quality of each feature subset in this step. Since one of the main goals of the proposed hybrid FS algorithm is to increase the LIDS’s accuracy, yield accuracy of the SVM that integrated with GABGWO is considered as the arithmetic value of GABGWO’s fitness that calculated by Eq. (10).

Step 4- Selecting the best solutions as parents GABGWO chooses the three best feature subsets with based on their fitness value and names them \( x_\alpha \), \( x_\beta \), \( x_\delta \), respectively. \( x_\alpha \), \( x_\beta \), \( x_\delta \) are considered as parents which will be incorporated in producing the new population in the next step. The main goal of this process is to update the other feature subsets according to the values of these best feature subsets by crossover operators which can enhance the GABGWO’s exploitation ability.

Step 5- Producing new solutions To avoid locally optimal solutions GABGWO tries to update each feature subset of the current population through applying its two new operators, namely crossover and mutation over the best-chosen feature subsets as follows:

---

### Table 2. Parameters Setting.

<table>
<thead>
<tr>
<th>S. No</th>
<th>Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No. of Candidate Solutions (N)</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>MaxCycle</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>Number of Runs (M)</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>Candidate Solution Dimension (D)</td>
<td>Total number of features =154</td>
</tr>
<tr>
<td>5</td>
<td>Search Space Domain</td>
<td>0 and 1</td>
</tr>
<tr>
<td>6</td>
<td>GABGWO_Probability_Crossover (Pc)</td>
<td>0.8</td>
</tr>
<tr>
<td>7</td>
<td>GABGWO_Probability_Mutation (Pm)</td>
<td>0.03</td>
</tr>
<tr>
<td>8</td>
<td>( \alpha )</td>
<td>90</td>
</tr>
<tr>
<td>9</td>
<td>( \beta )</td>
<td>134</td>
</tr>
<tr>
<td>10</td>
<td>( \delta )</td>
<td>154</td>
</tr>
<tr>
<td>11</td>
<td>GA_Probability_Crossover (Pc)</td>
<td>0.8</td>
</tr>
<tr>
<td>12</td>
<td>GA_Probability_Mutation (Pm)</td>
<td>0.03</td>
</tr>
<tr>
<td>13</td>
<td>FWP-SVM-GA_Probability_Crossover (Pc)</td>
<td>0.75</td>
</tr>
<tr>
<td>14</td>
<td>FWP-SVM-GA_Probability_Mutation (Pm)</td>
<td>0.09</td>
</tr>
<tr>
<td>15</td>
<td>SVM_kernel</td>
<td>RBF</td>
</tr>
<tr>
<td>16</td>
<td>SVM_gamma</td>
<td>1/k which k is number of features</td>
</tr>
<tr>
<td>17</td>
<td>K, in K-fold_Validation_Cross</td>
<td>10</td>
</tr>
</tbody>
</table>
• **Crossover operator** First a random value in 0,1 is generated and as Figure 3 illustrates the crossover operator is only performed if the generated random value is lesser than the control parameter $P_c$ (the probability that GABGWO must do the crossover and is called GABGWO_Probability_Crossover that initialized in Table 2), then another integer random number is produced between [0, D], with regard to that the random number’s value whether is $\in \{0, \ldots, \alpha\}$ or $\in \{\alpha+1, \ldots, \beta\}$ or $\in \{\beta+1, \ldots, \delta\}$, then the bits which are located in these ranges are updated using Eqs. (1) to (8). The $\alpha$, $\beta$, and $\delta$ are integer numbers between 0 to D which are initialized in Table 2. These processes that are called crossover in our new GABGWO algorithm, try to update the other feature subsets in accordance with the values of the three best feature subsets via Eqs. (1) to (8). This not only increases the diversity of the GABGWO but also accelerates its convergence.

• **Mutation operator** This operator selects the worst solution resulted from the crossover and modifies its values in a way that ‘0’ is changed to 1 or vice versa. The worst solution produced by crossover is selected and converted by mutation operator in a way that its ‘0’ values are changed to 1 and its ‘1’ values are changed to 0. After that, the fitness value of the mutated solution is calculated and if its fitness value is better than the worst solution, it is replaced with the worst solution. Note that this operator is done with the probability of $P_m$ (the probability that GABGWO must do the mutation and is called GABGWO_Probability_Mutation that initialized in Table 2) and its value is initialized less than $P_c$.

### 4.3 Anomaly Classification Stage

Exploiting classifiers to recognize the anomaly traffic from normal ones is one of the most popular methods to advance anomaly detection [74, 75]. Since the SVM classifier has shown an excellent performance in the intrusion detection domain and due to its high ability to deal with huge traffic [19, 20, 74, 75], it is utilized to build the anomaly detection model in this stage. To enhance the performance of SVM and to develop a lightweight anomaly detection system, only the informative and optimal traffic features of the AWID wireless dataset chosen by GABGWO are given to SVM. To high accuracy obtain, the SVM performs k-fold cross-validation (k=10) over the selected traffics by GABGWO where in the wireless traffics are randomly split into 10 equal sections. For learning the anomaly detection system, nine sections are considered and the one remaining section is considered for anomaly identifying. These operations are repeated until all the sections are utilized for identifying. Indeed, these processes are done until all the traffic data are examined. Finally, the mean of 10 examination results is considered as an accurate and lightweight anomaly detection.

### 5 Experiments

This section describes the used evaluation intrusion dataset, experimental setup, parameter setting, and evaluation criteria.

#### 5.1 The Intrusion Dataset

KDDcup99 and its derived versions are the most commonly used datasets utilized for evaluating the anomaly detection models. However, on the one hand, the KDDcup99 and its derivations are old and do not contain the new attacks, and on the other hand, because they are based on the DARPA 1998 TCP/IP data, they do not have the basic features of wireless networks [76–78]. AWID is a Wi-Fi network benchmark dataset that has recently been created [79] and can be used for the network anomaly detection evaluation, especially for Wi-Fi wireless networks [80, 81]. Thus, AWID is used for evaluating the LIDS. This dataset is freely available from http://icsweb.aegean.gr/awid/. The CLS with 4 target classes and ATK with 16 target classes are two types of AWID and both types have 155 features. Notably, the 16 classes of ATK version are classified into the 4 classes in the CLS version that used in this work. CLS includes one normal class and three intrusion classes. It also contains 1,795,575 wireless traffic in-
stances in which 1,633,190 samples are normal traffics, and 162,385 samples are various types of attacks. This benchmark intrusion dataset contains different types of data such as nominal, discrete, and continuous with various ranges. This results in difficulties for LIDS. Besides, unknown data values in AWID have been shown by ‘?’ and this dataset is unbalanced. The normal samples bias SVM and subsequently, affect the performance of the anomaly detection method. In addition to these, the network traffic datasets have a huge number of instances that not only decrease the searching speed of the feature selection algorithm but also complicate the detection process \[79–81\]. Therefore, the data preprocessing should be conducted to overcome the limitations in advance using the AWID dataset. Besides, because of ubiquitous Wi-Fi devices, the AWID Wi-Fi intrusion dataset has a very large number of records, some of which are redundant and reduce the speed of the GABGWO and LIDS. Therefore, for the mentioned considerations, before using AWID, its necessary to preprocess it.

### 5.2 Experimental Setup

Experiments are performed by a system with Intel(R) Core(TM) i7-4720HQ processor @ 2.60 GHz, 8 GB RAM, and Windows 10 operating system. Python software is used for dataset preprocessing. The GABGWO and other FS algorithms are implemented by java programming language. SVM is called from the machine learning library of WEKA \[82\] \[83\] that attached to java.

### 5.3 Parameters Setting

The metaheuristic algorithms’ parameters initialization is one of the procedures which significantly affects the efficiency of metaheuristic algorithms \[30\] \[34\]. This task is usually carried out by considering the algorithm applications or optimization problems \[30\] \[39\] \[84\]. In this work, the initialization of all parameters is empirically performed. Table 2 shows the initialization of all parameters used in this work.

### 5.4 Evaluation Measures

As seen in Table 1 the existing related works have used different measures for their evaluations. Among them, ACC, R, FAR, SF, and CTs (note that CTs consists of LIDS_Time that is the time consumed for anomaly detection by SVM and FS_Time that is the time consumed for choosing optimal traffics by GABGWO) are the most commonly used metrics. However, in this paper in addition to these metrics, we adopt some other evaluation metrics include P, F1. These metrics are computed according to four main criteria of the confusion matrix in Table 3 and Eqs. (10) to (14) \[24\] \[75\] \[85\].

- **True positive**: denotes the number of anomaly traffics that are identified truly as an anomaly.
- **False Negative**: denotes the number of anomaly traffics that are falsely identified as normal.
- **False Positive**: denotes the number of normal traffics that are falsely identified as an anomaly.
- **True Negative**: denotes the number of normal traffics that are identified truly as normal.

\[
ACC = \frac{(TP + TN)}{(TP+TN+FP+FN)} \tag{10}
\]

\[
R = \frac{TP}{(TP+FN)} \tag{11}
\]

\[
P = \frac{TP}{(TP+FP)} \tag{12}
\]

\[
F1 = \frac{(2 \times TP)}{(2 \times TP) + FP+FN} \tag{13}
\]

\[
FAR = \frac{FP}{(FP+TN)} \tag{14}
\]
Table 5. The Average of the Obtained Results From Experiments Under Scenario 2.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>SF</th>
<th>ACC (%)</th>
<th>R (%)</th>
<th>P (%)</th>
<th>F1 (%)</th>
<th>FAR (%)</th>
<th>LIDS_Time (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA</td>
<td>76</td>
<td>98.87</td>
<td>99.25</td>
<td>95.39</td>
<td>97.28</td>
<td>0.81</td>
<td>11.124</td>
</tr>
<tr>
<td>GWO</td>
<td>64</td>
<td>97.22</td>
<td>99.01</td>
<td>93.13</td>
<td>95.97</td>
<td>1.28</td>
<td>12.161</td>
</tr>
<tr>
<td>FWP-SVM-GA</td>
<td>81</td>
<td>98.87</td>
<td>99.20</td>
<td>95.48</td>
<td>97.30</td>
<td>0.82</td>
<td>11.390</td>
</tr>
<tr>
<td>BGWO</td>
<td>104</td>
<td>98.61</td>
<td>96.35</td>
<td>96.99</td>
<td>96.67</td>
<td>0.75</td>
<td>12.091</td>
</tr>
<tr>
<td>GABGWO</td>
<td>94</td>
<td>99.09</td>
<td>99.30</td>
<td>96.31</td>
<td>97.84</td>
<td>0.68</td>
<td>10.906</td>
</tr>
</tbody>
</table>

Figure 4. The Reduction Time for LIDS by GABGWO.

6 Experimental Scenarios, Results, and Analysis

In this section, two scenarios are defined for evaluating the performance of the proposed anomaly detection method and also investigating the effectiveness of the GABGWO on it. It is important to note that, to achieve reliable results from random and nondeterministic GABGWO algorithm, it is performed for M (that is set in Table 4) times, then the average of M results is considered as the final result.

Scenario 1 The performance of the LIDS with all traffic features is compared to its performance with traffic features subset chosen from GABGWO and the average of the achieved results is given in Table 4.

The experiment results in Table 4 show the GABGWO can increase anomaly detection’s speed with reducing the dimensionality of the huge network traffic, as seen in Figure 4. Besides, as Table 4 and Figure 5 illustrate the proposed method also satisfied the main goal of a wrapper FS method, namely, decreasing the number of features and maintaining the accuracies intact or at least a little worse. The obtained results in Table 4 show that the selected optimal features subset by GABGWO give accuracies near to accuracies obtained using whole features.

Scenario 2 The performance of the LIDS based on GABGWO is compared with its performance based on pure GA, GWO, and two other recent FS algorithms such as FWP-SVM-GA [20] and BGWO [60] with the same fitness function and parameter settings according to the Table 5. The average of the achieved results is summarized in Table 5.

As the experiment results show in Table 5, the GABGWO provides better performance than the original GA and GWO and other existing FS techniques with respect to various evaluation measures. GABGWO through choosing lesser optimal features than BGWO and choosing more informative features than GA, GWO, and FWP-SVM-GA, has presented high anomaly detection accuracy which these are clear in Figure 6 and 7.

Furthermore, as Figures 8 and 9 illustrate, choosing the more informative and related features by GABGWO causes more decrease in the FAR and computational time of the anomaly detection method than other techniques.

The GABGWO is also compared with GA, GWO, and FWP-SVM-GA metaheuristic FS algorithms in...
7 Conclusions and Future Work

In this paper, we have presented an SVM-based lightweight anomaly detection model named LIDS, using combination concepts of GA and mathematical equations of GWO (GABGWO) for WSNs in IoT. Generally, the proposed model involves three main stages: to speed up the GABGWO and convert the complex traffics of the AWID Wi-Fi intrusion dataset into a readable format for SVM, a set of preprocessing operations are performed in stage 1. For selecting the optimal wireless traffic data to incorporate in anomaly detection, a wrapper FS technique based on the combination of GA and GWO is developed in five steps at stage 2. In the first step, all the parameters related to GABGWO, are initialized, in the second step, a primary search space is generated, quality of the solutions are evaluated in the third step, the best solutions are determined as parents in the fourth step, and finally in the fifth step, producing new solutions based on GABGWO’s crossover and mutation operators is performed. In stage 3, the optimal traffic features selected by GABGWO are given the SVM classifier to distinguish between anomaly and normal traffics. Experimentations are done under two different scenarios. In the first scenario, the proposed anomaly detection method is conducted with all features and in the second scenario, it conducted with features subset resulted by the GABGWO. The empirical results obtained from experiments proved that not only the GABGWO provided good performance for LIDS, but also it outperformed the pure GA and GWO and the existing FS algorithms for various evaluation metrics. In future work, the proposed hybrid algorithm will be extended with lower runtime to address different optimization problems.

References

Figure 10. The Presented FS’S Time by GABGWO and Other Algorithms.


A Lightweight Anomaly Detection Model using SVM ... — A. Davahli, M. Shamsi, et al.


Azam Davahi received her B.S. degree in Software Engineering and M.S. degree in Computer Networks from Islamic Azad University, Qazvin, Iran, in 2006, 2009 respectively. Currently, she is a Ph.D. student in Software Engineering at Islamic Azad University Qom, Iran. Her research interests are the security of wireless and internet of things networks, cryptography, feature selection, optimization algorithms, and soft computing.

Mahboubeh Shamsi received her B.S. degree in Mathematics from Isfahan University, Isfahan, Iran, and her M.S. degree in Software Engineering from Islamic Azad University, Najafabad, Iran, and her Ph.D. degree in Software Engineering from University Technology Malaysia, Malaysia, in 2003, 2006 and 2011, respectively. Currently, she is an assistant professor in the faculty of Electrical and Computer Engineering at the Qom University of Technology, Qom, Iran. Her research interests include image processing, soft computing, cloud computing, and central and distributed database.

Golnoush Abaei received her M.S. in Software Engineering from the University of Mysore, Mysore, Karnataka State, India, and her Ph.D. degree in Artificial Intelligence from University Technology Malaysia, Malaysia, in 2008 and 2015, respectively. Currently, she is an assistant professor in the faculty of Electrical, Computer, and Biomedical Engineering and Head of Research in Shahabdanesh Institute of Higher Education in Iran. Her research interests are artificial intelligence, soft computing, software testing, and software defect prediction.