



8th International Congress of Information and Communication Technology (ICICT-2018)

A Novel Intrusion Detection System based on IABRBFSVM for Wireless Sensor Networks

Dai Jianjian^{a,一}, Tao Yang^a, Yang Feiyue^a

^aCollege of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Abstract

With the rapid development of wireless sensor technology, the application of Wireless Sensor Networks (WSNs) is more and more extensive, and has important military value and broad commercial application prospect. However, due to the limited resources of terminal equipment, wireless communication environment and other reasons, it faces severe security problems. This paper mainly proposes an intrusion detection algorithm based on improved AdaBoost-RBFSVM, and designs an intrusion detection system (IDS) for WSNs denial of service (DoS) attack based on the proposed method. In order to make the RBF-SVM algorithm as the AdaBoost weak classifier, the effect of training is achieved. Using the influence of parameter σ to RBF-SVM and the effect of model training error e_m on the smoothness of AdaBoost weights, the IABRBFSVM algorithm is proposed. On the other hand, after analyzing the DoS attack, the eigenspace for the attack is proposed, and the corresponding intrusion detection system is designed. Through simulation, the proposed IDS can significantly improve the network performance by detecting and removing malicious nodes in the network, from the perspective of detection rate, packet delivery rate, transmission delay and energy consumption analysis, and has the characteristics of simple structure, short computation time and high detection rate.

© 2018 The Authors. Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Selection and peer-review under responsibility of the scientific committee of the 8th International Congress of Information and Communication Technology.

Keywords: Wireless Sensor Networks; Intrusion Detection System; AdaBoost; SVM

1. Introduction

The open characteristics of WSNs deployment area and the broadcast characteristics of wireless communication cause the network to be vulnerable to various external attacks, which seriously threaten the entire network

* Corresponding author. Tel.: +86-131-0121-0702.

E-mail address: s150131019@stu.cqupt.edu.cn

information security and normal use. The solution to this problem is to develop a wireless sensor networks intrusion detection system to ensure the normal operation of the network.

In recent years, with the development of machine learning and deep learning, the AdaBoost algorithm has been successfully applied in intrusion detection. AdaBoost algorithm is proposed by Freund and Schapire¹, Sun X and Yan B use this algorithm to combine multiple classifier cascade structures to implement the intrusion detection application of WSNs². Aljawarneh S and Aldwairi combine AdaBoost, random forest and other algorithm³, using the idea of ensemble learning to get high detection rate in intrusion detection, but it lacks pertinence and complex models, and the generalization is not high. Yu Ren⁴ uses AdaBoost and Support Vector Machine (SVM) to classify intrusion attacks. Experimental results show that this algorithm is more balanced than SVM alone. SVM was originally proposed by Vapnik and other⁵, and Aburomman⁶ proposed an effective method to optimize the parameters of SVM with Particle Swarm Optimization (PSO) to improve intrusion detection accuracy. Compared with the method of Shams, such as⁷, the method has little error and improves the detection speed for the DoS attack. Murugan and other⁸ put forward the method of deploying the SVM weak classifier to the node by using the wireless sensor networks layer structure and detecting the attack by the joint node. This method has a higher detection rate than the method of deploying IDS to the base station. However, due to the variety of routing protocols in WSN, the algorithm is not universality and detection is not high, and it doesn't solve the problem of how SVM is adjusted to AdaBoost weak classifier.

In order to improve the precision of intrusion detection and adjust the RBF-SVM algorithm as a AdaBoost weak classifier, a new intrusion detection algorithm and an intrusion detection system based on the DoS attack are proposed in this paper, mainly reflected in 2 aspects: adjusting the parameter sigma of RBF-SVM and changing the updating rule of AdaBoost weight, the IABRBFSVM algorithm is proposed. For the DoS attack in wireless sensor networks, the corresponding eigenspace is put forward, and the intrusion detection system with alarm threshold is designed by using the algorithm proposed.

2. DoS attack based on AODV routing protocol

WSNs is composed of a large number of sensor nodes, which are with large amounts of computation, storage and energy consumption, and these sensor nodes may operate autonomously in complex and harsh environment without manual intervention. One of the basic goals of WSNs is to collect data from the physical world, but its broadcast characteristics are vulnerable to a variety of network attacks. DoS is the most common in WSNs attacks⁹, this attack seriously affect the quality of service (QoS) parameters, such as energy consumption, throughput, transmission delay, packet delivery ratio, and limited characteristics of node resources in WSNs, these make security become a key problem. DoS is a kind of attack that makes network resources unable to normally be used for normal nodes. One of the most common ways of such attacks is to use unsolicited, overwhelming data packets to overwhelm network nodes, thus occupying bandwidth and depleting the target system resources. Based on the AODV routing protocol, the nodes of the WSNs send different types of control packets to ensure the topology of the connection. When a node receives a RREQ packet, the node broadcasts the HELLO packet, which continues until the destination node receives the packet. When a node knows the routing path, it will send RREP packets to the source node. Through this process, a routing path is established for the actual data transmission. By using the DoS attack of the RREQ flood under the AODV protocol, the QoS parameters of WSNs is affected. When malicious nodes receive such packets, it will generate multiple routing request packets, deplete the limited resources of wireless sensor nodes, and seriously affect the network performance. In addition, malicious nodes send RREQ messages ceaselessly, causing neighbor nodes to fail to handle requests from other nodes.

2.1. Energy consumption model

Because of the limited resource characteristics, the most important problem is how to reduce the energy consumption and prolong the life cycle of WSNs. At present, the energy consumption analysis of WSNs often uses the Radio energy dissipation model (REDM), assuming that the distance between the sending node and the receiving node is d , and the default threshold is d_0 . The energy consumed by sending 1 bits:

$$E_{TX}(k, d) = \begin{cases} E_{elec} \times k + \xi_{fs} \times k \times d^2, & d < d_0 \\ E_{elec} \times k + \xi_{amp} \times k \times d^4, & d \geq d_0 \end{cases} \quad (1)$$

The energy consumed by receiving 1 bits:

$$E_{RX}(k) = E_{elec} \times k \quad (2)$$

k is the length of the data packet, E_{elec} is each emit or receive 1 bits of data of energy consumption. When $d < d_0$, the transmission is short range transmission, using a free space model ξ_{fs} . When $d \geq d_0$, transmission is long distance transmission, multi path transmission attenuation model ξ_{amp} will be used.

2.2. DoS attack model based on AODV

In the case of AODV routing protocol, the control packets have HELLO packets, RREQ packets and RREP packets. Under AODV routing protocol, nodes know the adjacent nodes by broadcasting HELLO packets, and use RREQ packets to find the route to the destination. DoS attacks use RREQ packets and Hello packets to achieve flooding attacks, which can destroy network resources such as destroying routing or consuming node computing power, bandwidth, memory and energy.

RREQ attacks. Malicious nodes send RREQ or RREP packets in a large number of ways without following the AODV protocol retransmission mechanism, so that the RREQ packets cannot be reached by the normal nodes, so that the middle routing nodes continuously broadcast the RREQ packets and deplete their resources.

HELLO attacks. Malicious nodes use HELLO packets to deceive adjacent nodes. Because malicious nodes have higher transmission power than ordinary nodes, they can broadcast their HELLO messages to a wider range, deceiving ordinary nodes as the next hop. When the common node needs to carry out the routing request, the transmission power does not reach the next hop of the malicious node, which leads to the loss of the packet.

3. Improved AdaBoost-RBF-SVM algorithm

AdaBoost is an iterative algorithm which uses the idea of lifting to update the weight of the sample many times. By improving the weights of the previous classification error samples, the next training will pay more attention to the error samples and improve the classification effect. Finally, the weak classifier which is trained every time is combined with a strong classifier.

3.1. RBF-SVM

In order to meet the conditions of AdaBoost training, the kernel function of SVM is selected. The kernel function is an eigenspace to kernel space mapping, SVM find a hyperplane maximum geometric space in kernel space, so as to realize nonlinear separation in eigenspace. According to the relationship between the size of the training set and the eigenspace, when the eigenspace dimension is smaller than the size of the training set, the use of the Gauss kernel function will have a good effect¹⁰.

Gauss kernel function:

$$K(x, z) = \exp\left(-\|x - z\|^2 / 2\sigma^2\right) \quad (3)$$

RBF-SVM Decision function:

$$f_i(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^N a_i^* y_i \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{2\sigma^2}\right) + b^*\right) \quad (4)$$

Compared with a single SVM, when using RBF-SVM as an AdaBoost weak classifier, it is necessary to solve the problem that how to train and set the parameter. As mentioned above, the parameter σ , which must be pre-set before RBF-SVM training, is required. The intuitive way is to apply single parameter σ to all RBF-SVM classifiers simply. Because the SVM classifier with weak or too strong performance will be trained in the process of weight distribution promotion, which will lead to the poor performance of the trained AdaBoost model. Although single best parameter σ may exist, the use of a single best parameter, through cross validation, will lead to the poor generalization performance of the model. Therefore, according to the trained error e_t and the weights distribution, the parameter σ of each training RBF-SVM are modulated each time to make the final training model reach their best performance.

3.2. Improved AdaBoost with RBF-SVM

To solve the problem of training RBF-SVM strong classifier as AdaBoost weak classifier, we need to weaken these strong classifiers appropriately, so as to achieve the purpose of combining multiple classifiers to enhance the performance of the final training model. Therefore, the parameter RBF-SVM above each training, initially set to a large value, each training process, in addition to the weights update, also need to update the parameters σ , to achieve the condition that the training error e_t is less than 0.5. Although the uncertainty of parameter σ leads to less precision than RBF-SVM under the best parameter σ , RBF-SVM under this condition has greater diversity, so that the final AdaBoost model has better generalization ability. On the other hand, the traditional AdaBoost algorithm in every round of training will increase the weight of the classified error sample and reduce the weight of the correct sample, these will cause uncommon samples to allocate too much weight, and the current RBF-SVM model training without good generalization ability, to cover these unusual samples. Therefore, the weight updating rule is improved, and the model error is added to the rule of updating weights, so as to achieve the effect of updating weights slowly, so as to satisfy the condition that RBF-SVM is used as a weak classifier and the parameter σ gradually increases generalization ability. The improved AdaBoost with RBF-SVM (IABRBF-SVM) is as follows:

Algorithm: IABRBF-SVM

1. **Input:** a set of training samples with labels $D = \{(x_1, y_1), \dots, (x_N, y_N)\}$; the initial σ_{step} ; Maximum number of iterations T ; the RBF-SVM weak classifier of the t iteration $f_t(x)$;

2. **Output:** the ensemble of the weak classifiers $F_{strong}(x)$

3. **Initialize:** the weights of training data $W_1 = (w_{11}, \dots, w_{1N})$, $w_{1i} = 1/N$, for all $i = 1, 2, \dots, N$, and σ_{init}

4. For $t = 1, 2, \dots, T$

Step1: Train weak classifier by distribution W_1

Step2: Get weak classifiers $f_t(x)$ of RBF-SVM by the σ , Calculate training error e_t and Model error e_m of $f_t(x)$

$$e_t = \sum_{i=1}^N w_{it} I(f_t(x_i) \neq y_i) \quad e_m = \frac{1}{N} \sum_{i=1}^N I(f_t(x_i) \neq y_i) \quad I = \begin{cases} 1 & f_t(x_i) \neq y_i \\ 0 & f_t(x_i) = y_i \end{cases} \quad e_m < \frac{1}{TN} \sum_{j=1}^T \sum_{i=1}^N I(f_j(x_i) \neq y_i)$$

Step3: If $e_t > 0.5$, decrease σ by σ_{step} and goto Step 1

Step4: Update the weights: $W_{t+1} = (w_{t+1,1}, \dots, w_{t+1,N})$

$$w_{t+1,i} = \left(w_{it} / Z_t \right) \exp(-a_i y_i f_t(x_i)) \quad Z_t = \sum_{i=1}^N w_{it} \exp(-a_i y_i f_t(x_i)) \quad a_i = \frac{1}{2} \log(1 - e_t \times e_m) / e_t \times e_m$$

5. **End for**

Output the final hypothesis:

$$F_{strong}(x) = \text{sign} \left(\sum_{t=1}^T a_t f_t(x) \right), \quad F_{strong}(x) = \{-1, +1\}$$

4. IDS based on IABRBF-SVM

IDS is usually composed of three modules: Data collection module, detection and response module. Data collection module takes the statistical information of WSNs as the input of the data analysis module of the detection

module, and finally executes the response module according to the output of the detection module, taking into account the needs of IDS, each module is designed at different stages. When nodes receive or send data packets, packets will pass through wireless transmission module, receive, sending buffer, detection module, response module and routing module. The whole process will consume MCU computing resources and energy of nodes. Because malicious nodes in the process of simulation is randomly selected, the malicious node will close the IDS, when it receives the request packet, the AODV routing module will not abide by the restrictions, a large number of sending a response packet. Because malicious nodes cannot be collected data, nor the deployment of IDS, when there are nodes of DoS attacks in WSNs, the target node through the deployment of IDS to detect whether or not its attack, and take corresponding actions to remove the malicious nodes in WSNs, restore the normal operation of the network performance.

4.1. Data collection module

Data collection module acquisition parameters are based on the observation of DoS attacks, through simulation of nodes in the absence of attack and attack under different features of the impact of network behavior, such as the attacked nodes can be observed that there are obvious traffic increment and other features. When malicious nodes began to attack the target node, they will lead to a significant increase in the number of packets received or transmitted within each unit time by the destination node and the neighbor node so as to exceed the bandwidth limit, resulting in frequent packet loss. Through experiments, based on AODV routing protocol, features of DoS attack is as follows.

The node receives the data packet $C_{RX}(n, \tau)$ is equal to the number of packets actually received by the node n within a unit time τ (not included the number of packet loss), defined as formula 5, RP (Receive, Packets) is equal to all of the data packet received, PL (Packets, Loss) is equal to the number of packets lost.

$$C_{RX}(n, \tau) = \sum_{\tau} RP - \sum_{\tau} PL \quad (5)$$

Packet delivery rate (PDR) $R_{DR}(n, \tau)$ is equal to the number of packet loss packets received per unit time τ and data packet and unit node n in unit time received in the period of time and the ratio of, defined as follows.

$$R_{DR}(n, \tau) = \sum_{\tau} RP / \left(\sum_{\tau} RP + \sum_{\tau} PL \right) \quad (6)$$

The energy loss (LE) $E_{LE}(n, \tau)$ is equal to the node n in unit time τ node energy consumption value. According to formula 1 and 2, $E_{LE}(n, \tau)$ is defined as shown in formula 7. E_{idle} , the consumption of energy per second standby node.

$$E_{LE}(n, \tau) = \sum_{\tau} E_{TX} + \sum_{\tau} E_{RX} + \tau E_{idle} \quad (7)$$

End To End Delay (EED) $T_{EED}(n, \tau)$ is equal to the average time of node n successfully send packets to the destination node in unit time τ , DP (Receive, Packets) is equal to successfully send packets, defined as follows.

$$T_{EED}(n, \tau) = \sum_{\tau} EED / \sum_{\tau} DP \quad (8)$$

Node caching utilization $R_{cache}(n, \tau)$ defined as formula 9, including L for the data packet length, and Cache Size (CS).

$$R_{cache}(n, \tau) = \sum_{\tau} L / (CS \times C_{RX}(n, \tau)) \quad (9)$$

In addition, in order to improve the detection rate of IDS, we add the number of adjacent nodes and the number of node routing tables as additional feature parameters, so as to adapt to the characteristics of DoS attack and improve the detection rate of the algorithm.

4.2. Detection module

The core of the IDS is the detection module, which has a great impact on the network performance. The network simulation software NS2 is used to simulate the DoS attack of WSN based on the AODV routing protocol. The trace files generated by the simulation model are analyzed, and the eigenspace obtained above is extracted and counted. The collected data set is randomly divided into training set, validation set and test set. The ratio is 0.6:0.2:0.2, which is used to train IABRBFSVM algorithm, get the final classifier, and deploy it as a module to the node.

4.3. Response module

The output of IABRBFSVM is sent from the detection module to the response module to make the final conclusion, and the final response is generated after analyzing the input, and corresponding actions are taken accordingly. In order to reach the best decisions before the detection module is output, two situations need to be considered: the accuracy of the detection module and the possible model of the attack. Because the false alarm rate and false alarm rate of IABRBFSVM can reduce the accuracy of the detection module, setting the alarm threshold can improve the detection precision of the module. With the continuous check module detect node DoS attack, when the alarm threshold reaches its maximum threshold response module according to the data acquisition module data, it will identify malicious nodes according to the node routing table to find the maximum number of records request node for malicious nodes, and send the message to the IDS decision and its neighbor nodes removed from the routing table malicious node. After the malicious node is removed from the system, the response module reset the threshold for the IDS to detect the next malicious node. The alarm threshold is set as a variable parameter. The relationship between the alarm threshold and the F-Score based on IABRBFSVM intrusion detection system is simulated, and the optimal alarm threshold is selected.

5. Experiment and analysis

By simulating the DoS attack based on AODV routing protocol, the collected data set is used to train IABRBFSVM-IDS and test the performance of IDS under various environments and parameters. The experiment is divided into 7 steps

Step 1, the routing protocol is determined, and the corresponding network parameters are set up to create a simulation model.

Step 2, under the AODV protocol, the RREQ flood simulation DoS attack is used to implement the C++, and it is embedded into the NS2 simulation platform as a component.

Step 3, the parameters of simulation model based on the set, using the TCL language implementation of AODV routing protocol and the data acquisition module, using the attack stage component generated by using the Setdest simulation of random attacks of malicious nodes adjacent node per unit time, and then randomly selected a malicious node attack adjacent node.

Step 4, run the TCL script program, generation and analysis of Nam file and Trace file.

Step 5, use Python to write IABRBFSVM model and train model by data sets above.

Step 6, the optimal model is written with C++ as a component of NS2, which is written as a detection model and a response model. Modify the previous simulation model, use TCL to add the detection module and response module to the simulation model, adjust the number of nodes, the number of malicious nodes and node speed, run the TCL script program, and generate Trace files.

Step 7, repeat the upper stage process 20 times, extract the trace files generated by each simulation process using python, and use Matplotlib to display the performance parameters of statistical performance.

5.1. Experimentation

The system used in the whole simulation process is CentOS 7, the software is NS 2.35, and python2.7 based on the IABRBFSVM model training. The simulation parameters are shown in Table 1.

Table 1. The simulation parameters.

parameters	value
Simulation area	800m*800m
Number of nodes	100
Transmission source	15
Mac protocol	IEEE 802.11
Routing protocol	AODV
Bandwidth	2Mbps
Initial energy of node	0.5J
Simulation time	400s
E_{elec}	$5 \times 10^{-8} J$
f_s	10-11J
amp	$1.3 \times 10^{-13} J$
d_0	87m
Traffic model	CBR
Node speed	10m/s
Traffic Rate (Normal/Attack)	200/2048Kbps
Packet Length	512bytes
AlarmThreshold	3

In the fourth step, the TCL script file is run, and the trace file are generated. And the nam run diagram is shown in Fig. 1. The graph is a DoS attack network animation demonstration screenshot based on the AODV routing protocol. The total simulation time is 400s, and the trace file records 1850137 transmission data. And 3600 data sets were obtained through statistics, calculation and data cleaning in the unit time.

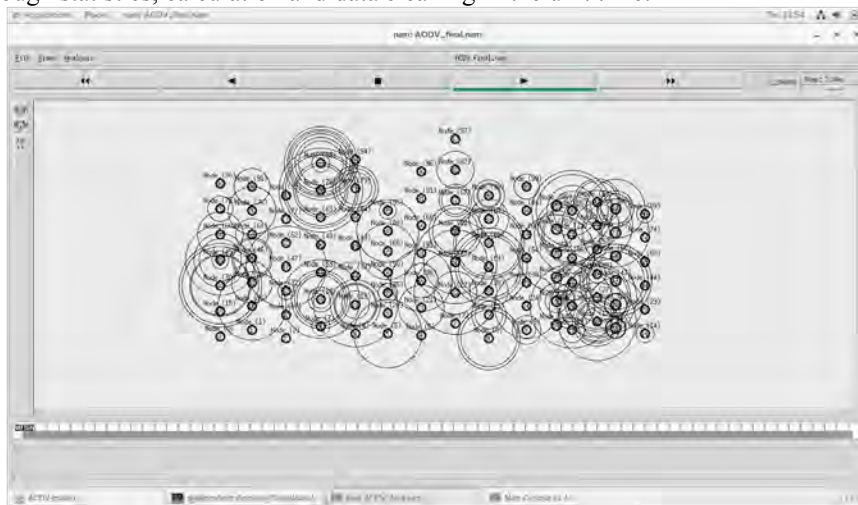


Fig. 1. Simulation of DoS attack based on AODV routing protocol.

5.2. Detection rate

First, we use the Detection Rate (DR) as the most important performance parameter of IABRBFSVM-IDS. We compare the detection results with AdaBoost-SVM⁸, Random Forest¹⁰ and SVM⁷. Fig. 2 (a) shows the IDS in different situations with the increase of malicious nodes in the network, the network topology more complex, the network suffered more DoS attacks, the IDS average DR decreased; the IABRBFSVM algorithm is better than AdaBoost, Random Forest and SVM algorithm based on IDS, the average DR is 95.72, even when malicious nodes under many circumstances and the malicious node number is 10, the DR algorithm has good performance, which is better than other algorithms, 2.08%, 2.21%, 2.91%.

5.3. Delay

The end to end delay (EED) is defined as the destination of the delay of the total packet in the transmission and the number of packets received. Fig. 2 (b) shows the impact of the number of malicious nodes on the average end-to-end delay. Obviously, with the increase of malicious nodes in the network, the average network EED is also significantly increased. Fig. 2 (b) shows, with malicious nodes increased from 0 to 10 when the end-to-end delay is increased from 49ms to 1631ms, but IDS can effectively detect and remove from malicious nodes in the network, and when there is a large number of malicious nodes in WSN, network performance of IDS WSN increased more significantly. And from Fig. 2 (b), it can be seen that IABRBFSVM-IDS has a better effect on EED than other IDS, so it can meet the transmission task of wireless sensor network better.

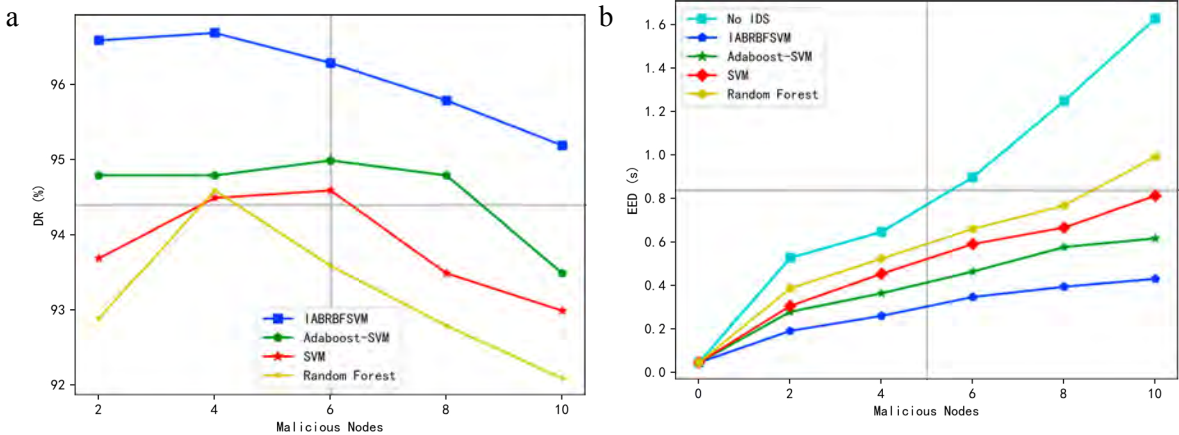


Fig. 2. (a) The relationship between the number of malicious nodes and the detection rate of IDS under DoS attack based on AODV protocol; (b) The relationship between the number of malicious nodes and network delay under DoS attack based on AODV protocol

5.4. Energy consumption

The residual energy of nodes is equal to the ratio of the total residual energy to the number of surviving nodes in the current time. The key to calculate the residual energy is the energy consumption of nodes in the unit time. When no sensor nodes running the intrusion detection system, the ability of communication cost is far greater than the energy consumption of the nodes according to the above model, the energy consumption of communication network based on AODV, in the DoS attack, the network of 100 nodes, including 10 malicious nodes, each node of the initial energy of 0.5J, respectively, IABRBFSVM, AdaBoost-SVM⁸, SVM⁷, Random Forest¹⁰, no defense five under the environment of network average node residual energy, as shown in Fig. 3. Because the IABRBFSVM algorithm has a higher detection rate for DoS attacks than other algorithms, and can detect and remove malicious nodes faster, making the average node residual energy slow down. And as the energy of some nodes is exhausted, the data packet transmission task is gradually completed, and the energy consumption of the average node is gradually reduced.

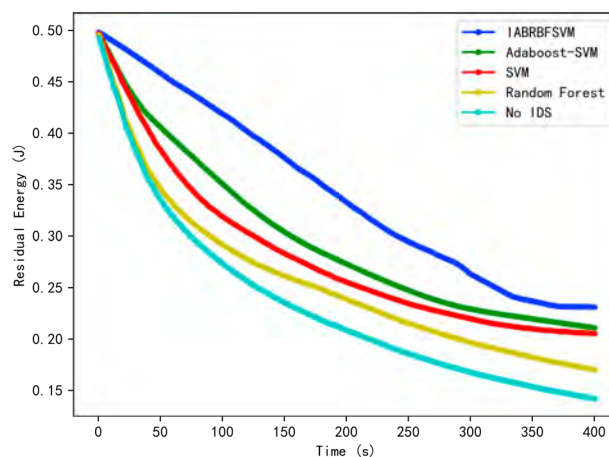


Fig. 3. The node residual energy under DoS attack based on AODV protocol.

6. Conclusions

In this paper, the IDS based on IABRBFSVM algorithm is proposed, which enables the network system to resist DoS attacks effectively. The experimental results show that the proposed IABRBFSVM-IDS integrated into the system can improve the network performance and make the performance closer to the ideal system. A large number of simulation results show that the proposed IABRBFSVM algorithm has a high detection rate for detecting network attacks in a short time and removing malicious nodes from the system. Future work needs to be focused on other attacks in the network.

Acknowledgements

This work was partially supported by Chongqing basic science and frontier technology research Project (cstc2017jcyjAX0135 7-2017-01).

References

- Freund Y, Schapire R E. A decision-theoretic generalization of on-line learning and an application to boosting[J]. *Journal of computer and system sciences*, 1997, 55(1): 119-139.
- Sun X, Yan B, Zhang X, et al. An Integrated Intrusion Detection Model of Cluster-Based Wireless Sensor Network.[J]. *Plos One*, 2015, 10(10):e0139513.
- Aljawarneh S, Aldwairi M, Yassein M B. Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model[J]. *Journal of Computational Science*, 2017.
- Yu Ren. An Integrated Intrusion Detection System by Combining SVM with AdaBoost[J]. *Journal of Software Engineering & Applications*, 2014, 7(12):1031-1038.
- Vapnik V. The nature of statistical learning theory[M]. *Springer science & business media*, 2013.
- Abuomman A A, Reaz M B I. A novel SVM-kNN-PSO ensemble method for intrusion detection system[J]. *Applied Soft Computing*, 2016, 38: 360-372.
- Shams E A, Rizaner A. A novel support vector machine based intrusion detection system for mobile ad hoc networks[J]. *Wireless Networks*, 2017:1-9.
- Murugan K, Suresh P. Ensemble of Ada Booster with SVM Classifier for Anomaly Intrusion Detection in Wireless Ad Hoc Network[J]. *Indian Journal of Science and Technology*, 2017, 10(21):1-10.
- Lotfy P A, Azer M A. Performance evaluation of AODV under dos attacks[C]// *Wireless and Mobile NETWORKING Conference*. IEEE, 2013:1-4.
- Belavagi M C, Muniyal B. Performance Evaluation of Supervised Machine Learning Algorithms for Intrusion Detection [J]. *Procedia Computer Science*, 2016, 89:117-123.