ISSN (Print) : 0974-6846
Indian Journal of Science and Technology, *Vol 8(35), DOI: 10.17485/ijst/2015/v8i35/86652, December 2015* ISSN (Online) : 0974-5645

Hybrid Evolutionary Algorithmn based Intrusion Detection System for Denial of Service Attacks

S. Mourougan[1]* and M. Aramudhan[2]
[1]Periyar University, Salem - 636011, Tamil Nadu, India; smourougan@yahoo.com
[2]Department of IT, Perunthalaivar Kamarajar Institute of Engineering and Technology, Nedungadu, Karaikal - 609603, Puducherry, India; aranagai@yahoo.co.in

Abstract

Background/Objectives: In current days, administrating security effectively for computer resources has become a difficult task for the administrator. One of the security problems is Denial of Service (DoS), is a type of attack that tries to prevent legitimate users from accessing either the services or resources, by generating large number of artificial packets send towards the victim resource. In turn, the victim resource is unable to extend the service to legitimate user. To meet this type of attacks, numerous detecting and preventing systems have been proposed, but they were suffering from low detection accuracy and high false alarms. **Methods/Statistical Analysis:** A new computational technique was proposed to perform the classification task andextracting features from KDDCUP 99 datasets using genetic algorithm and Particle Swarm Optimization. This research work focuses on the identification of DoS attacks with high detection accuracy and less false alarms. **Findings/Conclusion:** In the proposed Intrusion Detection System model, attacks are identified by training the Particle Swarm Optimization classifiers with Genetic-Particle Swarm Optimization based on wrapper feature selection which is superior to those classical intrusion feature selection. The proposed work was implemented in MATLAB. The result shows high detection accuracy and fewer false alarms compared to the existing models.

Keywords: Denial of Service, Genetic Algorithm, Intrusion Detection System, Particle Swarm Optimization

1. Introduction

Intrusion is defined as the set of actions that attempt to compromise the confidentiality, integrity and availability of the resources[1]. It can be detected by absorbing the deviations from user's or historical pattern of behavior. Developing a high performance IDS is a very complicated research challenge for the researchers and mainly classification accuracy depends on the features extraction from the dataset. Computation intelligence techniques are used to classify the inward network traffics as either normal or malicious. A lot of computational intelligence approaches have been proposed by the researchers, for example artificial neural network, fuzzy sets, evolutionary computation, expert system approach, rule based approach, artificial immune systems etc[2]. Denial of Service (DoS) is a type of attack that tries to prevent legitimate users from accessing either the services or resources. Neptune, smurf, Pod and Teardrop are the types of DoS attacks that have been present in KDDCUP99 datasets. DoS attacks have been emerging attacks that create threat to business and Internet providers around the world. Intelligent computational mechanism is needed to encounter this type of attacks and extend safety environment that increase the confident of the users to use Internet business. IDS are the more prominent approach to detect the DoS type attacks with higher detection accuracy.

* *Author for correspondence*

IDS are previously examined with large amount of data that shows slow training, testing process and low detection rate. So, feature extraction is the challenging task in developing IDS[3]. Generally, the implementation of IDS consists of three phases such as data preprocessing, features extraction and classifier. The tasks that are carried out in preprocessing phases are: 1. Identifies the attributes and their value. 2. Converts categorical to numerical data. 3. Data normalization and 4. Compute redundancy check and handles about null value. Feature extraction process is a preprocessing step when constructing IDS, used to reduce the dimensionality of the dataset by removing irrelevant, redundant features and improving the prediction accuracy of the classifier using selected features from the dataset. Classifier module finds the conditions of the traffics as either legitimate or malicious attack. Classifier is faced with a problem when it has to generate rules with many attributes or features. Obviously, the time required to generate rules is proportional to the number of features. In addition, irrelevant and redundant features can reduce both the predictive accuracy and comprehensibility of the induced rule and degrade the classifier speed. Thus, selecting the most relevant features is necessary, this strategy is implemented to simplify the rules and reduce its computational time while retaining the quality of classifier, as it represents the original features set.

Particle Swarm Optimization (PSO) is a well-known biologically inspired computational search and optimization algorithm which is based on the social behaviors of bird's flocks or schools of fish[4]. It is suitable to solve difficult problems because of its stochastic nature. Genetic Algorithm (GA) is an efficient search method based on principles of natural selection and population genetics. It is being effectively applied to problems in business, engineering and science. GA uses three operators' namely selection, crossover and mutation. The selection operator identifies the fittest individuals of the current population to serve as parents of the next generation. Cross over operator combines the second half of the first record with the first half of the second record. Mutation operator randomly changes the bits from '0' to '1' and vice versa. The concept of PSO applied to continuous nonlinear functions obtains reasonable solutions in short computational time. PSO is a type of heuristic search and self repeatable process that helps to brings optimal solutions. Hence, it is used as a classifier in the proposed IDS.

The rest of this paper is structured as follows. Section 2 discusses the related works of existing IDS algorithms for DoS attacks; Section 3 illustrates the proposed genetic-PSO algorithm. Section 4 describes the implementation of the proposed algorithm using KDDCUP 99 dataset. The last sectiondeals with conclusion.

2. Existing Methods

There are many methods and frameworks which are proposed in order to detect the SYN flood attacks. The authors of the detected SYN flooding attacks at leaf routers that connect end hosts to the Internet, utilize the normalized difference between the number of SYNs packets and the number of FIN (RST) packets in a time interval[5]. If the rate of SYNs packets is much higher than that of FIN (RST) packets by a non-parametric cumulative sum algorithm, the router recognizes that some attacking traffic is mixed into the current traffic. Similar work is presented in[6], where the fast and effective method was proposed for detecting SYN flood attacks. Moreover, a linear prediction analysis is proposed as a new paradigm for DoS SYN flood attack detection. The proposed mechanism makes use of the exponential back off property of TCP used during timeouts. By modeling the difference of SYN and SYN&ACK packets, it is shown that this approach is able to detect an attack within short delays. Again this method is used at leaf routers and firewalls to detect the attack without the need of maintaining any state. However, considering the fact that the sources of attack can be distributed in different networks, there is a lack of analysis for the traffic near the sources and also the detection of the source of SYN flooding attack in TCP based low intensity attacks is missing. Moreover, a quite similar approach was used in[4], which also considers a non-parametric cumulative sum algorithm, apply to measure the number of only SYN packets, and by using an exponential weighted moving average for obtaining a recent estimate of the mean rate after the change of SYN packets. In three counters algorithms for SYN flooding defense attacks were proposed and includes detection and mitigation[3,7,8]. The detection scheme utilizes the inherent TCP valid SYN–FIN pairs behavior, hence is capable of detecting various SYN flooding attacks with high accuracy and short response time. The mitigation scheme works in high reliable manner for victim to detect the SYN packets of SYN flooding attack. Although the given schemes are

stateless and required low computation overhead, making itself immune to SYN flooding attacks, the attackers may retransmit every SYN packet more than one time to destroy the mitigation function. It is necessary to make it more robust and adaptive.

The authors built a standard model generated by observations from the characteristic between the SYN packet and the SYN+ACK response packet from the server[9]. A method was proposed to detect the flooding agents by considering all the possible kinds of IP spoofing, which is based on the SYN/SYN-ACK protocol pair with the consideration of packet header information[10]. The Counting Bloom Filter is used to classify all the incoming SYN-ACK packets to the sub network into two streams, and a nonparametric cumulative sum algorithm is applied to make the detection decision by the two normalized differences, with one difference between the number of SYN packets, the number of the first SYN-ACK packets, another difference between the number of the first SYN-ACK packets and the number of the retransmission SYN-ACK. It is a simple and efficient method to detect and defend against SYN flood attacks under different IP spoofing types[10]. The method makes use of a storage-efficient data structure and a change-point detection method for distinguishing complete three-way TCP handshakes from incomplete ones. The presented experiments in[11] consistently show that their method was both efficient and effective in defending against TCP-based flooding attacks under different IP spoofing types. However there was lack of process automation within the scheme setting the parameters. Additionally, the method was not evaluated in a reasonably large real network.

Moreover, there are also some other related studies such as SYN cookies, SYN filtering mechanisms[12], SYN cache, SYN proxy (firewall), SYN kill, D-SAT[13] and DiDDeM[14,15], and more related studies in [16–19]. In the [16,17] an Early Stage Detecting Method (ESDM) is proposed. The ESDM was a simple but effective method to detect SYN flooding attacks at the early stage. In the ESDM, the SYN traffic is forecasted by autoregressive integrated moving average model, and non-parametric cumulative sum algorithm is used to find the SYN flooding attacks according to the forecasted traffic. The ESDM achieves shorter detection time and small storage space. However, these exiting methods or defense mechanisms which oppose the SYN flooding attack are effective only at the later stages, when attacking signatures are obvious[17].

3. Proposed Genetic-PSO based IDS

The architecture of the proposed GA-PSO model is shown in Figure 1. The architecture contains two phases 1. Training phase 2. Testing phase. In training phase, the KDDCUP 99 datasets was used, Data pre-processing, feature selection using genetic algorithm and classifier using PSO were implemented in training stage and DoS patterns were identified. Second stage is testing stage, the captured traffic is evaluated as in training stage, pattern identified, matched with database and decision to be taken. New patterns were identified by analyzing the behavior of the traffic, if it was against the legitimate traffic, the pattern was captured and updated in the database.

The design and their implementation of Genetic-PSO based IDS have following phases such as, Preprocessing, Feature Selection, and Classifier.

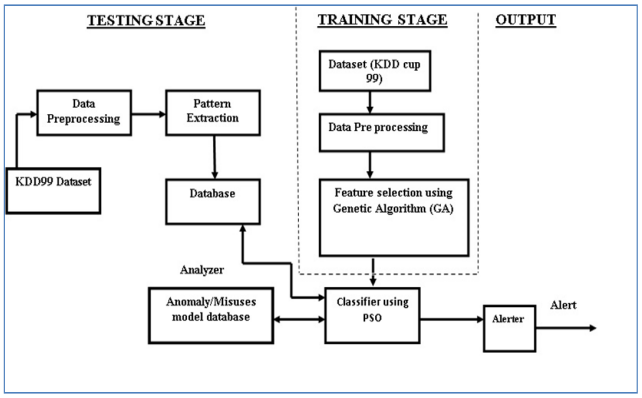

Figure 1. Architecture of genetic-PSO based IDS.

3.1 Data Preprocessing

Data preprocessing is an important step in the machine learning computing that eliminates out of range values, impossible data combinations, missing values etc. Generally data preprocessing includes learning, normalization, transformation, feature extraction and selection. The output of the data preprocessing is the final training set that extracts knowledge for the testing phase. The following steps used for data preprocessing:

- Identifying features and its related values.
- Converting original feature data value in to numerical data value.
- Applying data normalization based on min-max normalization.
- Perform similarity check and remove null values.

3.2 Feature Selection based on Genetic Algorithm

Accuracy of the classifier depends on the selection of optimum feature subset. Feature selection method mainly used for selecting subset of features from the original data set. There are two feature selection methods that are already proposed namely filter and wrapper methods. Filter method was mainly based on general characteristics of data features without involving machine language. These features are ranked based on certain criteria, where features with highest rank values are selected as optimal. The main advantages of filter method are low computational cost without involving any machine language algorithm for future selection. Frequently used filter method is information gain method. Wrapper method is mainly used for feature subset selection from the data set based on objective function and analysis of the performance of feature subset.

In this paper, Genetic Algorithm is used to select optimal feature subset from the datasets. GA reduces the KDDCUP 99 features from 41 attributes to 6 attributes features that are related to the characteristics of DoS attack, which reduces 85% of feature space. The six attributes are protocol_type, src_bytes, dst_bytes, count (No of connection to the same host), srv_count (No of connection requesting same service), serror_rate. KDCUP 99 dataset contains huge number of redundant records. 10% portions of the full dataset contains two types of DoS attacks (Smurf and Neptune). These two types constitute over 71% of the testing dataset which completely affects the evaluation. Brief Steps about Genetic Algorithm that selected features from dataset is presented as algorithm belows

- Initialize a population of Pre-processed data.
- Calculate objective function for each individual.
- Selection of individual solution.
- Perform mating of pair of individuals.
- Perform mutation operation.
- Calculate objective function for newly created population.
- If it satisfies stop the operation.
- Otherwise repeat step 3.
- Return the best features from KDD 99 dataset that reflects the properties of DoS

Algorithm 1. Genetic Algorithm based feature selection.

3.3 PSO Classifier

GA generates relevant feature from the data set and given as input to PSO based classifier. Using the available population (swarm) of individuals (particles) which are updated from iteration to iteration, the searches in PSO are being performed. The particles of PSO are composed of a set of attributes with Pbest. To determine the optimal solution, every particle moves in the direction of its previous best position (Pbest) and its global best position. If each particle is denoted by i and its dimension by j, it is assumed that = (1, 2….)^{T} denotes the current position of the ith particle and its fly velocity is denoted by = (1, 2….)^{T}. The velocity and position of particles can be updated by the following Equations (1) and (2).

$$V_{ij}^{t+1} = w.v_{ij}^{t} + c_1 rand_1 (pbest_{ij}^{t} - x_{ij}^{t}) + c_2 rand_2 (gbest_{ij}^{t} - x_{ij}^{t}) \quad (1)$$

$$X_{ij}^{t+1} = x_{ij}^{t} + v_{ij}^{t+1} \quad (2)$$

In the above formula, the evolutionary generation is given by t, the velocity of particle i on dimension j is given by and its value is limited to the range [-V_{max}, V_{max}], denotes the position of particle i on dimension j and its value is limited to the range [-X_{max}, X_{max}]. The global exploration and local exploration is balanced using the inertia weight w. The $rand_1$ and $rand_2$ are the random functions in the range [0, 1].

Each particle is considered as attacks that is composed of a set of attributes that prescribed the attacks. Every particle extracts further possible extended rule set in the direction of its behaviors and variations in the patterns of the attacks.

4. Simulations and Result Discussions

The proposed computational intelligence based Intrusion Detection System was implemented in MATLAB. During the evaluation, 10 percent labeled data of KDDCUP 99 was used for training the proposed IDS. This dataset contains three types of traffics and six types of DoS attack about four gigabytes and each traffic record has 41 features names whose values facilitate to identify the type category either as normal or attack. It contains a total of 24 attack types that fall into four major categories such as Denial of Service (DoS), probe, User to Root (U2R), Remote to User (R2L). DoS attacks are difficult to deal with because they are very easy to launch, difficult to track and also it is not easy to refuse the requests of the attacker. Back, land,

Neptune (Syn Flood), Pod (Ping of Death), smurf, teardrop are the six kinds of DoS attacks in KDDCUP 99. Back type of denial of service attacks against the Apache web server, an attacker submits requests with URL's containing many front slashes. As the server tries to process these requests it will slow down and becomes unable to process other requests. Back attack needs to know that requests for documents with more than some number of front slashes in the URL should be considered an attack. In the "smurf" attack, attackers use ICMP echo request packets directed to IP broadcast addresses from remote locations to create a denial-of-service attack. The Land attack occurs when an attacker sends a spoofed SYN packet in which the source address is the same as the destination address. Teardrop occurs due to IP fragmentation re-assembly code which does not properly handle overlapping IP fragments. This attack by looking for two specially fragmented IP datagram. The first datagram is a 0 offset fragment with a payload of size N, with the MF bit on (the data content of the packet is irrelevant). The second datagram is the last fragment (MF = 0), with a positive offset greater than N and with a payload of size less than N. Neptune attack describes that each half-open TCP connection made to a machine causes the 'tcpd' server to add a record to the data structure that stores information about all pending connections. This data structure is of finite size, and it can be made to overflow by intentionally creating too many partially-open connections. Neptune attack can be distinguished from normal network traffic by looking for a number of simultaneous SYN packets destined for a particular machine that are coming from an unreachable host. A host-based intrusion detection system can monitor the size of the tcpd connection data structure and alert a user if this data structure nears its size limit. Ping of Death attack has been reported when the systems react in an unpredictable fashion when receiving oversized IP packets. Possible reactions include crashing, freezing and rebooting. Ping of Death can be identified by noting the size of all ICMP packets and flagging those that are longer than 64000 bytes.

Based on the description above, the following rule structure derived from the KDDCUP 99 dataset and it is given in the Table 1. In this proposed model, the hidden related information from the features was observed. Learners discussed among others, about possible potential variations in traffic records which help to realize the prior knowledge of anomalous behaviors in advance. This proposed computational technique facilitates prompt detection and distinction of possible individual traffic records from crowd. There are 97,277 normal and 3, 91,450 DoS attacks traffic records in 10 percent labeled KDDCUP 99 data set. 2,80,790 smurf, 107201 Neptune, 2203 back, 979 teardrop, 21 land and 264 pod are in the10 percent labeled KDDCUP99. After removing duplicated instances class, 97277 normal, 641 smurf, 51820 Neptune, 994 back, 19 land, 918 teardrop, 206 pod are the traffic records considered for training the proposed IDS. The rule structure of six types of Dos attacks in KDDCUP 99 dataset is shown in Table 1. After PSO, the extended rule set identified with respect to each attack is shown in Table 2. Effectiveness of the IDS is evaluated by its ability to make correct predictions. Events are successfully labeled as normal and attacks. False positives refer to normal events being predicted as attacks. False negatives are attack events incorrectly predicted as normal events. Detection Accuracy (DA) is defined as the ratio of the sum of true negative and positive rate and sum of true and false positive and negative rate.

Table 1. Rule structure of Dos attacks in KDDCUP 99 datasets

S. No	Attack Description	Attack Type
1	protocol=ICMP,Service=ecr_i,src_byte=1032, flag=SF, host_count=255	smurf
2	protocol=tcp,service=private or ctf, flag=SO or SF, serror_rate=1	Neptune
3	protocol=tcp,service=http, flag = SF or RSTFR,src_byte=54540,dst_byte=7300 or 8314, same_srv_rate=1, srv_count>=5	back
4	protocol=UDP,service=SF,src_byte=28,wrong fragment=3,dst_host_count=255	teardrop
5	protocol=tcp,service=finger,flag=SO,land=1,srv_count=2,dst_host_srv_error_rate>=0.17	land
6	Protocol=ICMP,service=ecr_i,flag=SF,src_byte=1480,wrong_fragment=1,dst_host_count=255, dst_host_diff_srv_rate=0.02	Pod

Table 2. Extended rule set observed from the proposed techniques

S. No	Attack Description	Attack Type
1	If (Duration <3) and (protocol_type=icmp) and(dst_byte=125016) Then Buffer overflow	smurf
2	if {the connection has following information: source IP address 124.12.5.18; destination IP address:130.18.206.55; destination port number: 21; connection time: 10.1 seconds } Then {stop the connection	Neptune
3	protocol=tcp,service=http, flag = SF or RSTFR,src_byte=54540,dst_byte=7300 or 8314, same_srv_rate=1, srv_count>=5	back
4	If (source_bytes> 265616) and(source_bytes<= 283618) Then Warezmaster Attack	teardrop
5	If (Duration 0 to 25) and (protocol_ type = tcp and UDP) and (service=ftp OR private OR other domain)	land
6	If(duration<10seconds) of an FTP connection /session, there are many Hot indicators (hot > 20) being set by a logged user then it is highly likely that is being executed	Pod

The simulation results show that performance variations among evolutionary algorithms that where used as computational intelligence in IDS are less. Clustering based algorithms performance is better compared to non-clustered. Results reveal that no evolutionary algorithm performs better for all type of DoS attacks. Simulation results of the proposed techniques are shown in Table 3. Compared to the existing, proposed technique is efficient. It reduces more false negative compared to the existing work that reveals in the simulation results in Table 3.

Table 3. Results obtained from the simulation

Test Data	Training Data	Test data	Deduction Accuracy (%)	
			GA-PSO	Fuzzy Clustering
Normal	97277	60255	99.5	99.2
Smurf	641	400	83	96
Neptune	51820	20500	96	98
Back	994	714	98	96
Teardrop	918	300	96	96
Land	19	`07	94	99
Pod	206	101	99.5	98

5. Conclusion

In this paper, new computational technique was proposed by extracting the role of Genetic and PSO. The proposed method performs the classification task and extracts required knowledge using Genetic and PSO. The proposed systems are high reliability and adequate interpretability, and are comparable with several well-known algorithms such as Fuzzy clustering. Results on intrusion detection data set from KDDCUP 99 repository show that the proposed approach would be capable of classifying intrusion instances with high accuracy rate in addition to adequate interpretability of extracted rules. The results of PSO are better than fuzzy clustering technique.

6. References

1. Jain YK, Upendra. An efficient intrusion detection based on decision tree classifier using feature reduction. International Journal of Scientific and Research Publication. 2012; 2(1):1–6.
2. Joo D, Hong T, and Han I. The neural network models for IDS based on the asymmetric costs of false negative errors and false positive errors. Expert System with Applications. 2003; 25(1):69–75.
3. Gavaskar S, Surendiran R, Ramaraj E. Three counter defense mechanism for SYN flooding attacks. International Journal of Computer Applications. 2010; 6(6):12–5.
4. Siris VA, Fotini P. Application of anomaly detect algorithms for detecting SYN flooding attack. Elsevier Computer Communications. 2006; 1433–42.
5. Wang H, Zhang D, and Shin KG. Detecting SYN flooding attacks. In Proceedings of Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM); 2002 Jun 23-27. p. 1530–9.
6. Divakaran DM. Murthy HA, Gonsalves TA. Detection of SYN flooding attacks using linear prediction analysis. 14th IEEE InternationalConference on Networks, ICON2006; Singapore, 2006 Sep. p. 1–6.

7. Eddy WM. TCP SYN flooding attacks and common mitigations. RFC 4987, August. 2007; Available from: http://tools.ietf.org/html/rfc4987
8. Eddy W. Defenses against TCP SYN flooding attacks. Cisco Internet Protocol Journal. 2006; 9(4). Available from: http://www.cisco.com/web/about/ac123/ac147/archived_issues/ipj_94/syn_flooding_attacks.html
9. Nakashima T, Oshima S. A detective method for SYN flood attacks. 1st International Conference on Innovative Computing, Information and Control; 2006.
10. Nashat D, Jiang X, Horiguchi S. Detecting SYN flooding agents under any type of IP spoofing. IEEE International Conference on e-Business Engineering; Xi'an. 2008 Oct 22-24 p. 499–505.
11. Chen W, Yeung D-Y. Defending against SYN flooding attacks under different types of IP spoofing. ICN/ICONS/MCL'06 IEEE Computer Society; 2006 Apr 23-29. p. 38-44.
12. Yaar A, Perrig A, Song D. StackPi: New packet marking and filtering mechanisms for DDoS and IP spoofing defense. IEEE Journal on Selected Areas in Communications. 2006. 24(10):1853–63.
13. Shin S-W, Kim K-Y, Jang J-S. D-SAT: Detecting SYN flooding attack by two-stage statistical approach. Applications and the Internet. 2005; 430–6.
14. Haggerty J, Berry T, Shi Q, Merabti M. DiDDeM: A system for early detection of SYN flood attacks. GLOBECOM; 2004 Nov 29-Dec 3.
15. Haggerty J, Shi Q, Merabti M. Early detection and prevention of denial-of-service attacks: A novel mechanism with propagated traced-back attack blocking. IEEE Journal on Selected Areas in Communications. 2005; 23(10):1994–2002.
16. Qibo S, Shangguang W, Danfeng Y, Fangchun Y. An early stage detecting method against SYN flooding attacks. China Communication. 2009; 4:108–16.
17. Wei G, Gu Y, Ling Y. An early stage detecting method against SYN flooding attack. International Symposium on Computer Science and its Applications; 2008. p. 263–8.
18. Peng T, Leckie C, Rammamohanarao K. Survey of network-based defense mechanisms countering the DoS and DDoS problems. ACM Computing Surveys. 2007; 39(1).
19. Xiao B, Chen W, He Y, Sha MEH. An active detecting method against SYN flooding attack. Parallel and Distributed Systems. 2005.