Detection of Spoofed IP nodes using BAT Algorithm and Extreme Learning Machine

Sabitha Banu. A, Padmavathi Ganapathi

Abstract: IP spoofing is known as the most important cyber-attack which is the source for DoS or DDoS attacks where the attacker is hidden inside the network and makes the computer resource services unavailable to the users. The attacker once done with spoofing the IP address will start to flood the system with keeping on sending requests and make the network bandwidth slow to the extent. This paper contains the literature study of the different types of defence mechanisms from different authors used few decades before to detect and mitigate the Spoofed IP nodes at router, host level and recently some author come up with ideas of using computational intelligence methods for detecting the different types of attacks in wireless communications which results in accurate prediction. This paper provides creating a threat model of detecting the Spoofed IP nodes among 105 network wireless communication scenario using computational intelligence algorithm, the features are selected from the simulated raw data and preprocessed by using BAT optimization algorithm and features are converted to ELM readable format and then they are trained and learned using Extreme learning machine algorithm to predict the accurate detection of the Spoofed IP nodes in the wireless communication network scenario. The proposed method provides high accuracy in detection of Spoofed IP nodes with respect to some performance metrics like end to end delay, throughput, packet delivery ratio, packet drop ratio and it is compared with the KNN-SVM exiting model proved the results.

Keywords: IP Spoofing, Feature Selection, BAT algorithm, Extreme Learning Machine.

I. INTRODUCTION

In Today’s world Internet has been acting like a pillar for distributed applications where the client and server need not be in the same location for communication with each other. There are large number cyber-attacks increasing on regular basis. The major threat in cyber-attack world is IP Spoofing. According to CAIDA study from March 1, 2015-Feb. 28, 2017 there are 30,000 number of spoofing attacks have been detected on every day and approximately Twenty-one million attacks targets on around six million IP addresses.

IP address spoofing is where the attacker modify the source address to conceal the sender's identity to mimic the system. Internet Protocol address spoofing is obtained from the model which only depends on internet packet forwarding in routers to the packet destinations. Internet Protocol address spoofing utilizes the flaws in the protocol stack layer 3 and layer 4 to make the goal inaccessible.

IP address spoofing is launched when the victim negligence of the IP packets validation in source address to validate the authenticity of the sender. It is the vulnerable point of IP spoofing. By spoofing the IP addresses, attacker bombards

- DDoS attacks
- Smurf Attacks
- Network Time Protocol (NTP) synchronization reflection
- DRDoS(Distributed Reflection DoS)
- DNS request reflection
- TCP-SYN Flooding attack

The three different kinds of scenarios where the IP spoofing can be launched on:
- Host
- Router
- Flow

Of all network scenarios the flow-based attack is considered as MITM (“Man in the Middle Attack”) which means the attackers capturing the packet flow through the routing devices like wireless access points, through that they can able to alter or change or modify the IP address to do their purpose.

Configuration and Services prone to IP spoofing are
- Remote Process Calls
- IP address authentication services.
- The R suites like Remote login (rlogin), Remote shell(rshell)
- X windows system

The Objective of this paper is to discover accurate detection of IP spoofed nodes among various wireless nodes in a network scenario using BAT algorithm to optimize the features needed for ELM.

There are many methods available in the literature to handle IP Spoofing. Section 2 is about the review of literature about IP spoofing using different detection and prevention mechanism. Section 3 is about taxonomy of computational intelligence techniques and few literatures on applying CI techniques to detect and prevent IP Spoofing. Of all the suggested methods, Computational Intelligence methods are very effective in providing accurate results.
Section 4 is about the proposed methodology. Section 5 discusses about the Experiments and analysis of the outcomes and Section 6 ends with Conclusion.

II. REVIEW OF LITERATURE

There are different defence mechanisms used to prevent the network from IP Spoofing both in wired and wireless networks. Different kind of mechanisms deployed in the past is specified in the following Table 1.

Table 1: Literature Survey of IP Spoofing Mechanisms

<table>
<thead>
<tr>
<th>Author</th>
<th>Defence Mechanisms</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>B.Liu, J. Bi ,and A.V. Vasilakos[3]</td>
<td>Ingress and Egress Filtering</td>
<td>Router Level</td>
</tr>
<tr>
<td>Fegnuson, M.J. Baker and P. Savola[5]</td>
<td>Filters the path by using Access Control List(ACL) and unicast reverse path forwarding(uRPF)</td>
<td></td>
</tr>
<tr>
<td>Yao et al[6]</td>
<td>Virtual Anti Spoofing (VASE)</td>
<td>Router Level</td>
</tr>
<tr>
<td></td>
<td>Filtering on path, end to end authentication.</td>
<td></td>
</tr>
<tr>
<td>K. Park and H. Lee.[7]</td>
<td>Distributed Packet Filtering(DPF)</td>
<td>Router level</td>
</tr>
<tr>
<td>Z. Duann, X. Yuan, and J. Chandrashekar.[8]</td>
<td>If the packets are transmitted in an unexpected route they are dropped. filtering rules inside the domain are framed using “valley free feature” and filtering rules of BGP announcement.</td>
<td></td>
</tr>
<tr>
<td>A. Bremer-Bair and H. Levy[9]</td>
<td>Spooing Prevention Mechanism (SPM)</td>
<td>Router Level</td>
</tr>
<tr>
<td></td>
<td>Autonomous system tag is attached with the packet denoting key (Source, Destination). Destination key is checked and deleted after receiving.</td>
<td></td>
</tr>
<tr>
<td>X. Lin, A. Li, X. Yang, and D. Wetherrill[10]</td>
<td>Packet Passport Method</td>
<td>Router Level</td>
</tr>
<tr>
<td></td>
<td>Follows symmetric cryptography. Verifying the tokens placed in every packets for the valid source address. Distribution of symmetric keys for verification is taken care by the routing system.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Provides information to routers that helps for verifying the source address. Router builds a table which filters it using the prefix and information of the path.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Every packets from the BGP router is marked by giving a key and filtering the incoming packets by using that key.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Validates the source prefix bound with hop count value. Produces False negatives. By modifying the TTL value HCF can be bypassed.</td>
<td></td>
</tr>
<tr>
<td>A. Yaar, A. Perrig, and D. Song[14]</td>
<td>Stack Path Identifier(Pi)</td>
<td>Host and Router</td>
</tr>
<tr>
<td></td>
<td>Every router assigns packet marking and Packets traveling the same route will be marked the same way. Even valid packets gets chance to drop due to attackers travelling in the same path.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cryptography using hash values</td>
<td></td>
</tr>
</tbody>
</table>

All the existing IP spoofing defence mechanisms are applied either in host or router level and in few methods cryptographic puzzles are used. There are some more common preventing mechanisms like packet filtering, monitoring networks for malicious activity, configuring router and firewalls, providing robust verification authentication of all IP addresses, using network attack blocker, using IPv6 protocol, network protocols based on cryptography such as “HTTPS(HTTP Secure), SSH(Secure Shell) and TLS(Transport layer Security)”.  

III. COMPUTATIONAL INTELLIGENCE TECHNIQUES

General Structural outline of Computational Intelligence Techniques is shown in the Figure 1. CI methods are basically classified in to five types [16] namely Fuzzy Sets, Artificial Neural Networks, Evolutionary Computing, Swarm Intelligence Artificial Immune System has given in Fig 2.. And Some Review of Literatures has been given below on CI techniques to detect IP spoofing and DoS attacks.

Fig. 1 Computational Intelligence Techniques Taxonomy

There are Few Literatures available for the detection of IP spoofed nodes applying computational Intelligence Techniques and it yields the high accurate results with minimal error rate show in the below Table 2.

Table II: Literature survey on CI Techniques to detect IP Spoofing

<table>
<thead>
<tr>
<th>Author</th>
<th>CI Techniques</th>
<th>Attacks and Accuracy rate</th>
</tr>
</thead>
</table>
Combination of Extreme Machine Learning of Feed Forward Networks from ANN and BAT Algorithm of Swarm Intelligence can bring the accurate results in detection of the IP spoofed nodes in a wireless network and the results are compared relating to some performance Metrics like “packet drop ratio, packet delivery ratio, throughput ratio, end-to-end delay”.

IV. PROPOSED IP SPOOFED NODES DETECTION APPROACH

There are various techniques and tools to detect the IP spoofed nodes in the wireless network. But our proposed hybrid detection of IP spoofed nodes is very simple and efficient with Fast Learning rate, provides accurate results with minimal error and the most important it is very easy to implement. Let see the approach in detail.

A. Bat optimization approach

Detection of IP spoofed nodes are done by using one of the Nature inspired algorithms called BAT algorithm based on SI techniques. The bat algorithms works on finding the location of the prey and bats by raising some echo signals. BAT algorithm is very efficient and is applied for optimization, classification, and feature selection.

BAT Algorithm works on the basis of echolocation which shows the bats behaviour to discover way out for single and multi-objective optimization problem. BAT Algorithms is used in order to distinguish unusual traffic from ordinary traffic in the proposed strategy.

B. Behavior of bats

Even in full darkness bats discovers its food and can differentiate among different kinds of insects. The BAT algorithm is an evolutionary algorithm based on population where each bat is a solution. It describes the behaviour of the bats according to the location of the echo emitted. This echolocation behaviour of bats empowers them to identify where the prey is located by emitting some frequency sound pulses and it carefully listens sound pulses which returns back from the place where the food/prey is located or the neighbouring phenomenon’s. During the time when a bat move towards close to the prey, the echo emittance is decreased and frequency sound pulse rate is increased.

The 3 general rules of BAT algorithms:

1. Entire Bats use echo based on location to feel distance, and in addition they predict the variation between prey and other unnecessary barriers in a few magical approach.

2. Bats arbitrarily fly with Speed S at position P, with specific Frequency $F_{\text{min}}$, fluctuating Wavelength $\lambda$ and echo vibration $B_0$ to hunt its food. Bats naturally fine-tune the frequency of the radiated pulses & fine-tune the pulse rate emitted signals $R$ which depends on closeness of the object.

3. In spite of vibration of the echo can be differed in various ways, it is pretended that the vibration of the echo differs against maximum (+ve) $B_0$ to a least constant rate $B_{\text{min}}$.

From the above past literatures it is observed that rather than Machine Learning Techniques, applying Computational intelligence methods brings the accurate results for the all kinds of complex problems. The Proposed Method Hybrid
Detection of Spoofed IP nodes using BAT Algorithm and Extreme Learning Machine

Initialize the population of bats. Modify the required parameters after initialization that are required for Fitness and finally the Fitness remains estimated for every individual Bat in the Population. Bat algorithm used for classification of IP Spoofing attacks from normal traffic is given below.

**Input:** Bat population $x_i = (x_{i1}, x_{i2}, ..., x_{iN})^T$ for $i = 1, 2, ..., N$, $\mathbf{p}$. velocity $v_i$, pulse rates $r_i$, loudness $L_i$, pulse frequency $f_i$ at $x_i$, and maximum number of generations $Max_{Gen}$.

**Output:** The best solution $x_{Best}$ and its corresponding fitness value $f(x_{Best})$

1: Initial bat $x_i$, $i = 1, 2, ..., N$
2: Evaluate fitness for each bat $f(x_i)$
3: while $(t < Max_{Gen})$
4: Generate new solutions by adjusting frequency, and updating velocities and locations/solutions
5: if $(\text{rand} > r_i)$
6: Select a solution among the best solutions
7: Generate a local solution around the selected best solution
9: end if
10: Generate a new solution by flying randomly
11: if $(\text{rand} < L_i$ and $f(x_i) < f(x_{Best}))$
12: Accept the new solutions
13: Increase $r_i$ and decrease $L_i$
14: end if
15: Rank the bats and update the current best solution $x_{Best}$
16: Increase the generation number $t$
17: end while

BAT optimization algorithm selects optimum features from the raw data by feeding the behaviour or pattern of DDoS attacks for detection purpose and those features are used for training and testing.

The advantages of using the BAT algorithm is it is very efficient, frequency is tuned randomly, automatic zooming on the nodes where promising solutions are found. Parameter tuning is done automatically which increases the performance of the system, solves complex network problems efficiently, accurate results in quick time, algorithms converges easily at the starting stage.

C. ELM approach

Extreme Machine Learning (ELM) can be used for selection of features at the first and the single layer of hidden nodes with parameters of the hidden nodes don’t need to be tuned. Input weights are randomly assigned for the hidden nodes and output weights are learned in a single step. ELM has a high speed of learning rate compared with the other gradient neural networks. It produces good performance and learns thousand times faster than others neural networks. ELM works according the following steps:

- Data Pre-processing: Converting the raw TCP/IP data in to ELM readable format.
- Training Phase: ELM is trained with the feature selection data’s.
- Testing Phase: Trained data is compared with tested data to find out the accuracy detection.

ELM shows a good performance when it is deployed in the proposed work which contains huge volume of network traffic data. Combination of BAT and ELM has produces high accuracy and minimal error.BAT algorithm optimizes the parameters so that the detection is accurate.

Pseudo Code for ELM

**ELM training algorithm.**

**Input:** training set $((\mathbf{a}_i, \mathbf{t}_i), \mathbf{a}_i \in \mathbb{R}^d, \mathbf{t}_i \in \mathbb{R}^n, i = 1, \ldots, N)$. **Output:** SLFN parameters.

1: Randomly generate hidden node parameters $(a_i, b_i)$, $i = 1, \ldots, L$ where $a_i$ and $b_i$ are the input weight and bias values and $L$ is the hidden node number.
2: Calculate the hidden layer output matrix $\mathbf{H}$.
3: Calculate the output weight vector, $\beta = \mathbf{H}^T$.

D. Proposed methodology

The wireless network is created and the network traffic generated between mobile nodes are simulated in ns2.TCP, UDP connection setup is done among all wireless nodes. Packets are exchanged between the nodes when the wireless nodes are in the range of the gateway and also when the nodes move away from the hearing range of the gateway the packets are dropped.

![Overview of Proposed Methodology](image)

V. EXPERIMENTAL ANALYSIS AND RESULTS

A. Experimental Setup

An Experimental Setup of proposed works was made up by creating a wireless network with 105 mobile nodes and 4 gateways in NS2. NS2 is an open source simulator to design communication networks for research purpose. It is the best tool for implementing wired and wireless networks like VANET, MANET with different routing protocols like TCP, UDP, FTP, DSR, HTTPs, TELNET,CBR,VBR. All mobile nodes should be configured before using them. They are given in Table 3.

Network components of mobile node Parameters are

- Type of Antenna
- Channel Type
- Link Layer(LL)
- MAC Layer
• Ad-hoc routing protocol
• Interface Queue (iFQ)
• Radio propagation Model

Additionally,

<table>
<thead>
<tr>
<th>Table III. Simulation Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>option</td>
</tr>
<tr>
<td>Prop Type</td>
</tr>
<tr>
<td>Wired Routing</td>
</tr>
<tr>
<td>txPower</td>
</tr>
<tr>
<td>topology instance</td>
</tr>
<tr>
<td>rxPower</td>
</tr>
<tr>
<td>routerTrace</td>
</tr>
<tr>
<td>propInstance</td>
</tr>
<tr>
<td>phyType</td>
</tr>
<tr>
<td>MPLS (Multi protocol Label Switching)</td>
</tr>
<tr>
<td>mobileIP</td>
</tr>
<tr>
<td>macType</td>
</tr>
<tr>
<td>macTrace</td>
</tr>
<tr>
<td>ifqType</td>
</tr>
<tr>
<td>initialEnergy</td>
</tr>
<tr>
<td>Adhocrouting</td>
</tr>
<tr>
<td>Idle Power</td>
</tr>
<tr>
<td>energyModel</td>
</tr>
<tr>
<td>channel</td>
</tr>
<tr>
<td>antType</td>
</tr>
<tr>
<td>agentTrace</td>
</tr>
<tr>
<td>iiType</td>
</tr>
<tr>
<td>addressType</td>
</tr>
</tbody>
</table>

By setting up all the simulation parameters a wireless network is created and IP spoofing attacks are detected and the simulated network traffic is collected and the data is loaded in MATLAB for accuracy detection. MATLAB is a powerful tool for data analysis, optimization and visualization.

B. Results

A wireless network with 105 mobile nodes and 4 gateways are created and network traffic is generated.
VI. PERFORMANCE COMPARISON

Comparing the proposed method and the existing Methods performance evaluated which is shown in Figure 2.8 and are outlined in the below sections using different metrics.

A. End-to-End Delay Ratio

Travelled time throughout the network. The time when the packet is generated by the sender till it reaches the destination’s application layer. The formula for calculating End to End Delay ratio is given below

\[ \text{Average End Delay} = \frac{TD}{TR} \]

Where TD=Total no of Packets
TR= Packets Received.
PDR is shown in the Figure 2.9

B. Packet Delivery Ratio (PDR)

PDR is measured by transmitted packets by a CBR traffic source and the received packets by a CBR traffic sink. The formula used to calculate PDR is given below

\[ \text{PDR} = \frac{PR}{PS} \times 100\% \]

PR stands for Packets Received and PS stands for Packets Sent.
PDR is shown in the Figure 2.10

C. Packet Drop Ratio

The Difference between Packets Sent and Packets Received. The formula used to calculate Packet Drop Ratio is given below

\[ \text{PDR} = \frac{\text{Sent Packets} - \text{Received Packets}}{\text{Sent Packets}} \]

PDR is shown in the Figure 2.11

D. Throughput Calculation

The time taken to send data from sender to receiver. Throughput denotes the effectiveness of the routing protocol. The formula used to calculate throughput is given below
TP-PR * PZ/SE
Where PR is Received Packets
PZ is Packet Size
SE is Simulation End Time
Throughput can be shown in the Figure 2.12

Fig 2.12 Throughput ratio

VII. CONCLUSIONS

In this paper a hybrid combination of BAT based ELM approach is used to detect IP spoofed nodes in a wireless network and the simulated traffic dataset is generated. Few features like latency, simulation time, No of TCP packets sent and received, total no of bytes of a packets and packet loss is selected to train our ELM model. ELM approach randomly places the inputs and learns it in a “single layer feed forward neural network” and provides 100% detection accuracy results compared with SVM and KNN respectively under optimized parameter tuning. Our Proposed BAT Based ELM approach have got increased performance metrics in the form of Packet Delivery ratio, Delay, throughput and Packet Drop.

REFERENCES


AUTHORS PROFILE

Sabitha Banu, A., received her MCA in 2007 from Avinashilingam Institute for Home Science and Higher Education for Women University, Coimbatore. She is currently pursuing her Ph.D. at Avinashilingam Institute for Home Science and Higher Education for Women University, Coimbatore. Her areas of interest are Network Security, Machine to Machine Communications, Cryptography and Wireless Communications.

G. Padmavathi, is the Professor in the Department of Computer Science at Avinashilingam Institute for Home Science and Higher Education for Women University, Coimbatore. She has 29 years of teaching experience and one year of industrial experience. Her areas of interest include Real Time Communication, Wireless Communication, Network Security and Cryptography. She has significant number of publications in peer reviewed International and National Journals. Life member of CSI, ISTE, WSEAS, AACE and ACRS.