Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 02, February-2016

A Smart Packet Filter and Anamoly Detector using BDD and Apriori Algorithm

Deepak Bakhtiani , Sonali Bhatnagar
B.E. Student
Dept. of Computer Engineering,
Thadomal Shahani Engineering College,
University of Mumbai, India.

Sonal Shroff
Assistant Professor
Dept. of Computer Engineering,
Thadomal Shahani Engineering College,
University of Mumbai, India.

***Abstract*— Network threat has always persisted in every organization. Packet filtering is the one of the major contemporary firewall design techniques. An important design goal is to arrive at the decision at the packet level only. Implementation of such packet filter using Binary Decision Diagram (BDD) gives more advantages in terms of memory usage and look up time. In the traditional list-based packet filter firewall where rules are checked one by one for each incoming packet, the time taken to decide on a packet is proportional to the number of rules. The system is proposed to prevent the illegitimate use in an authorized group of networks, the packet filter acts as a network antivirus, where the administrator can define certain rules on the system that can levied on the organization. The term proctor is defined for the prototype. Given the system makes a comparison between list based and the BDD based approach, coupled with detection of novel set of attacks on the system. The system also uses a data mining model to generate real time rule for the firewall within the organization. Apriori Algorithm is used for this purpose, Apriori is the best algorithm used for mining association rules, which helps out to detect novel anomaly attack.**

***Keywords*— Intrusion Detection, BDD, Packet Filter**

I. INTRODUCTION

Security is the need of today’s net connected environment. With the increasing need of the internet, security remains in number of potential sources of attacks and network requires protection from all sorts of vulnerabilities to it. Hence to overcome this issue systems need to be carefully configured so that the systems become robust. A firewall is such device hardware/software which prevents the network from tampering, intrusion and maintaining the integrity of the system. Firewalls are basically installed to ensure the consistent policy across an organizations network. [1]

A. Firewall

A Firewall is an independent network device which is responsible for the security of the network; it basically acts as a proctor for the network environment. A Firewall is hardware or a software device or combination of both. Firewalls are broadly categorized into three sub categories namely: 1) Application Level Firewall,(which are protocol specific)[3] 2) Circuit Level Firewall, 3) Packet Filter Firewall. The proposed system makes the use of Packet Filtering approach.

A.1. Packet Filter Firewall

A Packet Filter Firewall is a core component of any network inspection tool available today. The application of this type of firewall is on the network layer of the protocol stack [2]. This type of Firewall makes a check on, each and every incoming and outgoing traffic packet by packet. IP packets contain information about the source and destination and internally it contains of transport layer information such as source and destination address, ports and protocols. TCP and UDP are such protocols under the transport layer whereas the TCP is the connection oriented and UDP is connectionless protocol. The packet filtering approach is widely used and adapted due to the following reasons.

i) The Development cost of such system is cheap as compared to other available systems.
ii) The approach is transparent and makes easier to use for development purpose [4]
iii) Most importantly these are very fast with regards the available counter parts.

The proposed system makes use of such type of approach for filtering the packets. The system focuses on the five available dimensions that are protocol, source port, source address, destination port, and destination address [5,6].

Consider a case of an organization where numerous machines are connected to a central server to which this firewall is in place which governs the overall activity of the network, this means all the traffic flow using the packet filtering approach in place will flow from that particular firewall. The general convention is that certain rules are levied on such type of arrangement so that the inflow of packets can be restricted on the basis of the rules. Traditionally a designer will follow the list based rule set, and when packet arrives at the firewall it gets checked from the rule set and whether to allow that packet or not is the look out of that firewall. Imagine millions of packets follow throughout the network organization and checking each and every bit of information that too on a packet level will be very bothersome and time consuming. Instead of a list based approach the system uses a BDD approach. BDD is a data structure that used generally to store the data in binary format. The approach used here [8] is the trie based approach where in the network policies are demonstrated as trie based format which is an ordered set of tuples maintaining the precedence of the rules and ensuring the integrity among them, further using a trie approach [7] makes it easy as it removes the multiple rules from the rule set making it easier for accessing. To show a uplift-ment on the traditional based approach that is the list based approach, as the disadvantage of list based approach were stated earlier [5], these types of traditional systems can be only adapted in static type of environment whereas the modern approach suggests the usage of BDD. Therefore a trie based system us is definitely good for an environment which is dynamic and also supports practical usage. So the main aim of fastening up the process of scanning the packets at firewall gets satisfied by using the above methodology by making the access time of the rule list.

B) Intrusion Detection

Intrusion detection system (IDS) is a tool which monitors real time and live traffic and guards the network from any malicious activity. In any network the most important thing is to protect the data and to avoid unauthorized access to any network which violates the network foundation. In a sense they make correct action to prevent all the possible set of attacks and in return they provide with set of intrusive activities. An intrusion is an activity that compromises the security for any system. In the proposed design the IDS works in the background, it functions on the basis of the supplied signatures of the intrusion activities and comes up with the activities leading an intrusion. These signatures are usually the pattern of attacks that can happen on any system, with the increasing need of computer and networking more and more number of possible attacks are possible, so one needs to keep constantly updating the signature database which will hold the signatures. The proposed system uses apriori algorithm, one of the classic data mining approach for item set mining. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. The frequent item sets are determined by Apriori and can be used to determine association rules which highlights the general trends in the database, this has application in market basket analysis and also in determining the known sets based on some probability factor, as the patterns of the attacks share same nature which can be determined by apriori at the first place.

II APPROCH

A. Binary Decision Diagram

A binary decision tree diagram BDD, is a tree type data structure used to represent Boolean values [9]. Processing over binary data is quite fast and efficient; BDD consists of one of the many decision nodes and two terminal nodes and edges which connect these nodes. A weighted node is marked with high (1) and a un weighted node is marked as (0). The BDD can be classified as follows:

Ordered BDD (OBDD): here each variable in BDD appears at most once in each complete path and if the variables appear in same order in all the paths [8].

Reduced OBDD (ROBDD): it is a form of ordered BDD but it does not have redundant nodes. ROBDD is the most used type of BDD in application [8]. Figure 1 demonstrates a BDD and Figure 2 demonstrate ROBDD which is in reduced fashion. TABLE 1 represents a table of variable ordering for the construction of BDD. As in the figure it represents a BDD tree in which a weighted cyclic graph is represented.

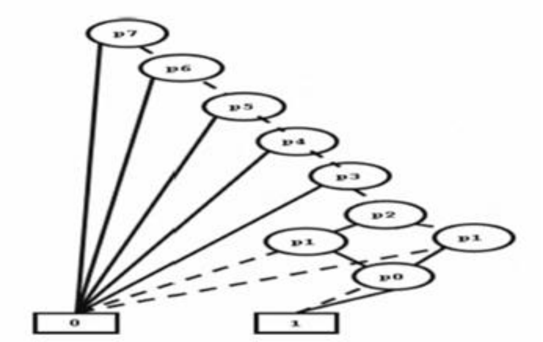


Figure 1: Example of BDD

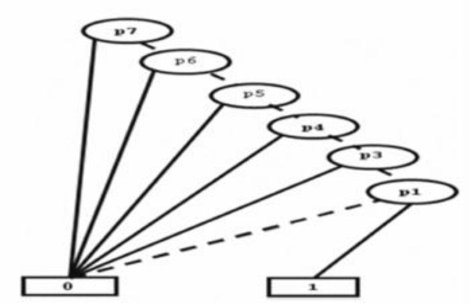


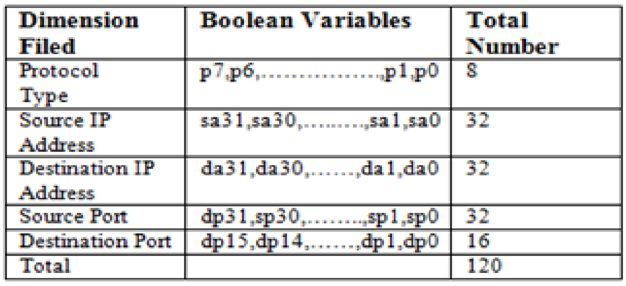
Figure 2: Example of ROBDD

Figure 2. Represents ROBDD, it clearly demonstrates that the number of states or rules gets reduced and that's the result the application of ROBDD is much in use as compared to traditional BDD.

B. Use of BDD as a firewall

Each packet contains header information such as a protocol, source address, source port, destination, and destination port. All these are represented through a binary format [10]. Suppose the protocol is TCP and the number is 6 that get converted to 00000110. To store these kind of variables are used for holding its place. The basic theory of BDD packet filter firewall is given in the report [10]. The actual traversal of the BDD is done from top to bottom. Example, number 6, 128,64 contains the following 8 bits.

Table I: Boolean variables required for the representation of access list

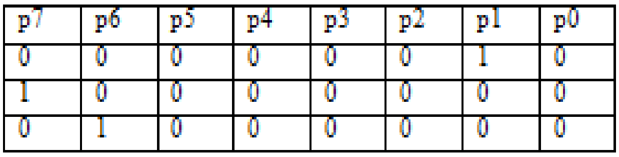


Dimension Filed	Boolean Variables	Total Number
Protocol Type	p7,p6,................p1,p0	8
Source IP Address	sa31,sa30,.........,sa1,sa0	32
Destination IP Address	da31,da30,.......,da1,da0	32
Source Port	dp31,sp30,.........,sp1,sp0	32
Destination Port	dp15,dp14,.......,dp1,dp0	16
Total		120

First line in Example I contains bits for the number 6. The above BDD (i.e. Fig. 2) represents both the numbers 6 and 3.

Therefore the number 6 would be traversed within the BDD by comparing each bit from top to bottom and reaches the leaf node 1 and that indicates the number and is accepted by the BDD mechanism. Bits that are in the second line are represented are for number 128. At node p7, if the bit is 0 it then traverses to the p6, but now the bit 1 which had reached, reaches the leaf node 0, indicating that 128 is rejected by the BDD. Number 128 then gets rejected by the BDD due to the mismatch at node number p6. The table also demonstrates variables that are required to store the access list in the BDD format. For the purpose of example and a broader perspective the below table mentions the rule set that are implemented in the system, which the admin of the system maintains.

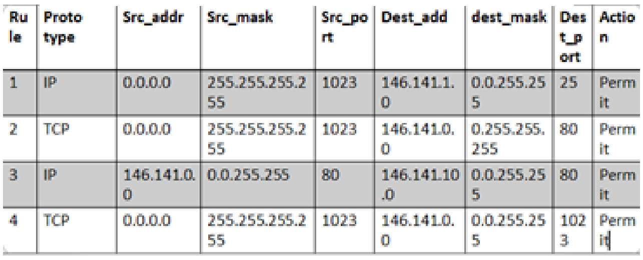
Example 1.



p7	p6	p5	p4	p3	p2	p1	p0
0	0	0	0	0	0	1	0
1	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0

Using the BDD way makes the system advantageous as the BDD has a non-redundant computation strategy, where the look up with regards the rules is done without any repetition. Therefore owing to BDD based approach redundant rules are curtailed and save the look up time where as in a traditional list based approach the look time is enormously large. BDD also supports multidimensional filtering where as in the five key elements mentioned in the above table states and the foremost important thing is that the entire operation is in binary making in efficient and faster [14].

Table II: Sample access list containing 4 rules



Rule	Proto type	Src_addr	Src_mask	Src_port	Dest_add	dest_mask	Dest_port	Action
1	IP	0.0.0.0	255.255.255.255	1023	146.141.1.0	0.0.255.255	25	Permit
2	TCP	0.0.0.0	255.255.255.255	1023	146.141.0.0	0.255.255.255	80	Permit
3	IP	146.141.0.0	0.0.255.255	80	146.141.10.0	0.0.255.255	80	Permit
4	TCP	0.0.0.0	255.255.255.255	1023	146.141.0.0	0.0.255.255	1023	Permit

C. Data Mining in Firewall

Data mining is a comprehensive approach to mine the frequent patterns of a data set to get an extensive result which is undiscoverable set of patters which current systems can't find at the first place. Data mining is a correlation technique of handling the data [12]. Data mining approach results a general pattern or a trend after analyzing the facts. There are several data mining approaches out of which the proposed system uses Apriori algorithm, Apriori algorithm is an association mining technique. Data mining approach in intrusion detection systems were used for auditing propose. Association rules are one of the data mining techniques. Association mining provides the set of information in the form of if-then statements, but they are based on certain probability [13]. Association has two numbers that expresses degree of uncertainty about that rule; these are referred as antecedent and consequent. The 'if' part is known as the antecedent and 'then' part is called as the consequent. They are simply an item set which don't have any item set in common. The two numbers that express the degree of uncertainty of the rule is one which demonstrates the support for the rule and the other is for the confidence of the rule

A Formal model of association is stated below:

$I = \{i_1, i_2, \ldots, i_n\}$ Set of items
D: Database of transactions
$T \varepsilon D$: a transaction, $T \subseteq I$
TID: unique identifier, associated with each T
X: a subset of I
T contains X if $X \subseteq T$
Association rule: $X \Rightarrow Y$ here $X \subset I$, $Y \subset I$ and $X \cap Y = \emptyset$
$\text{Supp}(X \cup Y)$ = Number of transactions in D contain $(X \cup Y)$

$$\text{conf}(X \Rightarrow Y) = \frac{\text{supp}(X \cup Y)}{\text{supp}(X)}$$

Apriori is a well suited algorithm for finding association on large sets of data that allows implication outcomes that consist of more than one item.

D. Apriori Algorithm

The main idea of apriori algorithm is to find frequent item sets whose occurrences exceed the predefined threshold. Apriori uses knowledge from previous item sets that occur frequently.

For creating frequent set lets define:

C_k as a candidate itemset of size k
L_k as a frequent item set of size k

Main steps of iteration are:

1. Find frequent set L_{k-1}
2. Join step: Ck is generated by joining L_{k-1} (Cartesian product $L_{k-1} * L_{k-1}$)
3. Prune step (Apriori Property) : Any (k-1) size itemsets that is not frequent cannot be the a subset of a frequent k size itemsets, hence it should be removed.
4. Frequent sets L_k has been achieved now.

Apriori uses a breadth-first search and has a hash like tree structure which makes candidate keys themselves. Candidate keys have a higher frequency than minimum support threshold that are qualified to be the frequent item sets. Pseudocode is given below.

```
Apriori(T, ε)
L_1 <= {large 1-itemsets that appear in more than ε transcations}
k <= 2
        while $L_{k-1} \neq \emptyset$
          Ck <= Generate($L_{k-1}$)
          For transactions t ε T
         Ct <= Subset($C_{kt}$)
        for candidate c ε $C_t$
          count[c] <= count [c] +1
Lk <= {c ε $C_k$ | count [c] ≥ ε }
K <= k+1
return $\bigcup_k L_k$
```

the above pseudo code demonstrates apriori algorithm which is used in the proposed system to detect the anomalies.

III DESIGN AND DEVELOPMENT

This section focuses on the implementation and its details, which leads to a successful completion of a packet filter and anomaly detector. The Table II contains the list of set of rules which states the rules for the functioning of the system, but in real life scenario there might be thousands of such rules that govern this type of a system. The implementation of this type of system is done by converting the parameters of rule field into binary data as to optimize the performance of BDD. The algorithm takes in the five dimensions namely protocol number, source address, port number, and destination address and port number, convert each of them into binary bits, store it to form a binary file in .blif file, for which the input is the rule list as stated in Table II.

A. Development of packet filter

The dot files also known as the CUDD file [11] which is the input to the system are the decider of the legitimate action based on the incoming and outgoing packets.

A.1 Algorithm: BDD Lookup algorithm

```
//input is the dot file
//output is the accepted or the rejected packet.
Check the head of the BDD given by the CUDD
If(solid) then
    final_op = 1
else
    final_op = 0
lookup_ptr = first_node of the BDD
while (lookup_ptr ==1 OR lookup_ptr == 0)
    if (header bits[lookup_ptr] = 1)
        lookup_ptr = high(lookup_ptr)
    else
        lookup_ptr= low(lookup_ptr)
    if (loopup_ptr == 1) then
        ACCEPT the packet
    else
        REJECT the packet
```

The above algorithm searches the BDD generated by the CUDD with respect to the rule set for which the BDD is generated. Thereafter it takes in each bit of the header and compares with the BDD generated. The CUDD is a package fro0m University of Colorado(CU), and is appropriately named as CUDD. It is supposedly the best BDD available package.

IV PROPOSED SYSTEM

The proposed system is a firewall which takes decisions automatically within the network and blocks the traffic coming from the outside world into the network based on the set of rule set which are explicitly provided to the system. This makes it a complete solution for the network based intrusion attacks as well as packet filtering approach to block a miss happening. Below is the pictorial representation of the system in board aspect.

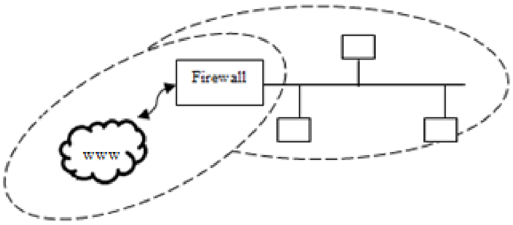


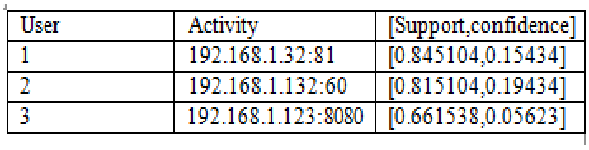
Figure 3: Proposed system

Figure 3.demonstrates the proposed system flow whereas the firewall acts like the proctor of the two entities that is the organization that is the within organist ion network and the outside world that is world wide web (www).

At the internet layer i.e. the WWW, its handling is done by the BDD approach, where in return to the incoming packet inside the organization the packets gets evaluated from the set of rules whose computation is done by BDD algorithm which is mentioned above in *A.1*. Here the purpose of redundant rules and faster processing of the rules is done, hence avoiding a bottle necking situation. For example there are 100 rules in the rule list and the 98th rule is the most frequent rule to be used under, with regards the list based approach for each and every packet it needs to do that many computations, but with BDD the number of computations reduces as it supports faster processing with its trie based structure . At the organization front it acts as a SNORT, which is the most popular used Intrusion detection software that performs content searching, and matching that also includes using rule sets and can be used to detect a variety of attacks which can happen within the network. Snort supports a variety of applications such as generating alert messages for the user based activities just as a log mechanism. In the current system same type of snort functionality is used to monitor the user behavior to catch any malicious activity on the network. For that training set of database has been attached which monitor these user activities and computes out a confidence factor for the possible form of an anomaly on a network. If the level of predefined malicious activity exceeds the limit, an alarm would be raised to the administrator notifying the intent of the attack and on that basis the administrator can take the necessary action. Figure 4 illustrates the scenario of the method. Misuse or the system verses the anomaly a user can create is taken into consideration.

A. Misuse vs. Anomaly Detection

In misuse detection system the system analyses recorded information by doing a comparison from the database, the system monitors the activities and records events, and each event is matched with the set of rules. An anomaly is some abrupt activity that violates the system if at all a deviation occurs, during an anomaly high false alarm is raised but it lacks training of the data as that event may or may not be malicious.



User	Activity	[Support,confidence]
1	192.168.1.32:81	[0.845104,0.15434]
2	192.168.1.132:60	[0.815104,0.19434]
3	192.168.1.123:8080	[0.661538,0.05623]

Table III: Sample Detection vs. Anomaly detection

For example in the user 1 in table III apriori generates a model which contains the fact:
192.168.1.32:81 => INTRUSION [0.845104,0.15434]
Means 84.5104% of the time user does activity (192.168.1.32:81) out of which 15.434% is the intrusion detected activity. Based on the current systems threshold the value can be determined. After collecting datasets the Apriori algorithm creates a model to detect activities (on the IP or on the port) which are malicious.

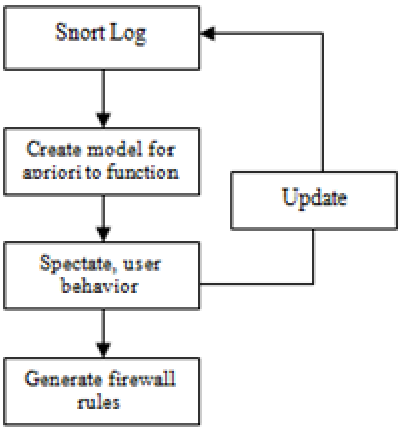


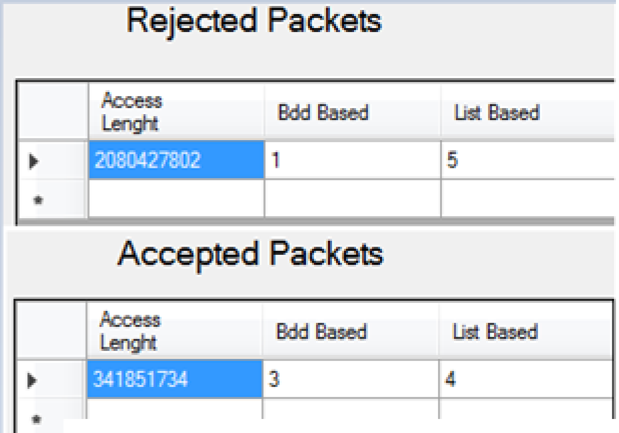
Figure 4: Aprioris' working Scenario

V. Result and Analysis

The experiment was carried out on the list based and the apriori based. For demonstration purpose List based packet filter was proven to be the best but it considerably lacks its efficiency when compared to BDD based approach, the BDD based prototype consumes lesser memory and provides a faster access. Here the accuracy of the firewall depends upon the number of acceptance and rejection rate of the packets that flow within the organization. Figure 5 shows the following comparison which proves that the BDD based approach is much better than the traditional list based approach. The anomaly based detection system classifies the type of intrusion on the system by association rule mining the Figure 6. Shows the comparison between the clusters based predicts and provides us the novel sets of attacks based on the given dataset to the system. The purpose of the system which suffices the need of a firewall and an intrusion detection system is satisfied in the given system approach and the association rule based approach, i.e. Apriori as the classification approach uses data mining approach.

VI. CONCLUSION

In this paper, a complete solution is provided which solves the problems that are faced within the network and outside the network. This model has also used data mining approach to tackle the snort related activities which tamper the organization within the internal means of the users. The comparison between the traditional list based approach and the BDD based approach is studied. Featuring as, Proctor for any organization. The results based on the filtering approach were done and BDD was found to be well suited. The association mining pattern proved that Apriori provides a better way to mine the novel set of attacks.



Rejected Packets

Access Lenght	Bdd Based	List Based
2080427802	1	5

Accepted Packets

Access Lenght	Bdd Based	List Based
341851734	3	4

Figure 5: Acceptance Rejection ratio

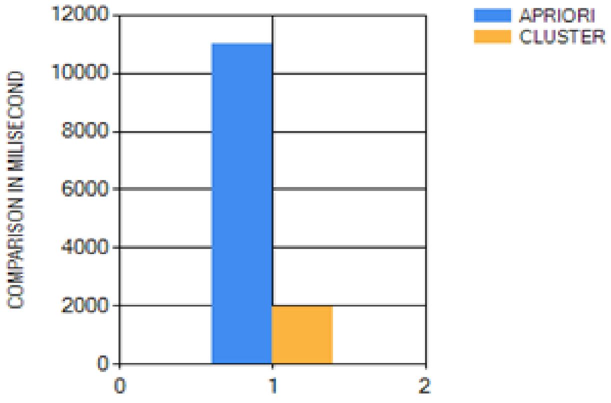


Figure 6: Comparison between Cluster and Apriori

References

[1] W.R. Cheswick and S.M Bellovin, "Firewalls and Internet Security: *Repelling the Hacker*", Addison-Wesley, Reading, MA, 1994.

[2] Dan Sttrom, "The Packet Filter: A Basic network ", SANS Institute 2000-2002.

[3] M. Christiansen and E. Fluery, "An MTIDD Based Firewall", *Tellecommunications System* 27:2-4, Page(s): 297-319,2004.

[4] H.Julkunen and C.Chow, "Enhance Network Security with Dynamic Packet Filter", Proceeings of ICCN, Page(s): 268-275,1998.

[5] S. Acharya, J. Wang, Z.Ge, T.Znati and A. Greeberg, "Simulation study of Firewall to Aid Improved Performance", Proceedings of ANSS, Page(s): 18-26, 2006.

[6] S. Acharya, J. Wang, Z. Ge, T. Znati and A. Greenberg, 'TrafficAware Firewall Optimization Strategies", Proceedings of ICC, Page (s): 2225-30, 2006.

[7] E. W. Fulp and S.J. Tarsa, “Trie-Based Policy Representation for Network Firewalls”, Proceedings of ISCC, Page(s): 434-441,2005.

[8] R. E. Bryant, "Graph-based algorithms for Boolean function Manipulation", Transaction on Computers, Vol. C-35(8), Page(s): 677- 691, 1986.
S. B. Akers, "Binary decision diagrams", Transaction on Computers, Vol. C-27(6), Page(s): 509-516, 1978.

[9] S. Hazelhurst, A. Fatti and A. Henwood, "Binary Decision Diagram Representations of Firewall and Router Access Lists", \jtp:/(ftp.cs. wits.ac.zalpublresearchlreportsITR-Wi, 1998.

[10] [F. Somenzi, "CUDD: CU Decision Diagram Package"; http://bessie.colorado.edu;'-fabioICUDD

[11] Abraham A, Thomas J. Distributed intrusion detection systems: a computational intelligence approach. In: Abbass HA, Essam D, editors. Applications of information systems to homeland security and defense. USA: Idea Group Inc. Publishers; 2005. p. 105–35 [Chapter 5].

[12] Lazarevic, A., Ertoz, L., Kumar, V., Ozgur, A., Srivastava, J., 2003. A comparative study of anomaly detection schemes in network intrusion detection. Proceedings of the SIAM International Conference on Data Mining.

[13] Gopal Paul, Amaresh Pothnal, c. R. Mandalt Bhargab B. Bhattacharya, “Design and Implementation of Packet Filter Firewall using Binary Decision Tree Diagra,”, Proceeding of the 2011 IEEE Students' Technology Symposium.

AUTHOR PROFILE

Deepak Bakhtiani is currently pursuing B.E. in Computer Engineering from Thadomal Shahani Engineering College. He is a tech enthusiast and keen in codin, he aims to pursue masters in the same field

Sonali Bhatnagar is currently pursuing B.E. in Computer Engineering from Thadomal Shahani Engineering College. She is interested developing and researching algorithms for cyber security.

Sonal Shroff Shroff received B.Sc. degree in Physics, B.Sc. (Tech) degree in Computer Technology and M.E. degree in Computer Engineering, from University of Mumbai

She has 15 years of experience in teaching, currently working as an Assistant Professor in Thadomal Shahani Engineering College. Her area of interests is satellite image processing, system security and data mining