International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 3 Issue 4, April - 2014

# Firewall Optimization Model With Malicious Participants Detection

Miss. Pooja R. Kadam
Computer engineering department
BSIOTR(W),Wagholi
Pune, India

***Abstract*—Firewalls are widely getting used for securing the private network. Firewalls check each incoming and outgoing packets and according the rules given by network administrator and it will take the decision whether to accept or discard the packet. As per the huge requirement of services on internet the rule set becomes large and takes more time to process one packet and it affects the throughput of firewall. So firewall optimization has a great demand to get good performance. Previously lot of work has been done in the area of firewall optimization like single firewall optimization, cross domain firewall optimization with privacy protection. In this paper we propose a model for firewall optimization within two administrator domain with malicious participant detection. We are detecting such users who are revealing the firewall policies while working in different administrative domain and also detecting the users who are trying to reveal the other parties policies by issuing the sequence of inputs by using anomaly detection technique. So our model first monitors the behavior of user and records this audit data and try to detect the malicious content hidden in single network packet and find the malicious users.**

***Keywords—Malacious Participants; Optimization; Firewall Policies; Domain***

## I. INTRODUCTION

Firewalls are widely used in securing private networks of organizations, corporate world, and personal networks. These firewalls are keeping at the entry point of private network to our secure network. Firewall checks each incoming and outgoing packet and according to policies set by the network administrator it will gives the decision whether to accept or discard the packet. These policies are nothing but the access control list and each firewall or router have two types of access control list1) for filtering incoming packet 2) for filtering outgoing packets. Firewall checks packet according to first match semantics means packet checks each rule sequentially until it found the match point and so on. So performance of firewall depends on rule set. Due to huge services available on internet network administrator set rules according to need of user and this rule set becomes bulky and packet processing time gets decreased. Optimization of firewall is needed to increase the firewall performance. Various different techniques are used to optimize the firewall like optimization of rules [1], inter firewall optimization [5], [6], intra firewall optimization [2], [3], [4] etc. Intra firewall optimization works on single administrative domain so no need of privacy protection and inters firewall optimization works on two different administrative domains with privacy protection. While working in two different administrative domains we have to look after the privacy protection of firewall policies. Previous work has been done in the area of inter firewall optimization by protecting the privacy of firewall [5], [6].But in author's threat model they consider the inputs to their protocol are honest and giving input sequence to disclosing the firewall policies are apart from this threat model. Dishonest employee in the organization try to disclose the firewall policies of other parties or other can disclose the policies by issuing the firewall input sequence. The threat model assume that one firewall policies are remain constant and another requirement of the protocol is whenever we want to execute the protocol, disclosing significant amount of firewall rules is very expensive. So we have to find such a malicious activities created by the dishonest employee. Such activities can be identified by using other parties using anomaly detection technique.

### *A. Intrusion Detection Technique*

Intrusion detection techniques used to find the malicious activities because of which security of system get compromise. In intrusion detection techniques we have to consider three key elements.1) Resources protection definition of genuine action on resources 3) Effective methods which act as a real time model of assigning activity as an intrusive. These techniques are based on two types first is signature based works on comparing collected data with predefined signatures and other is anomaly based works on legal behavior. There are two different techniques used in anomaly detection system Misuse detection and Anomaly detection systems. Misuse detection works on known attacks, in this first we find the previously known intrusions and predefined rules. But the disadvantages for this are we cannot detect the unknown attacks. Anomaly detection techniques works on d expect behavior. The action which deviates from normal behavior is an intrusion. But sometimes it raises so many false alarms .Following table will show us the difference between the misuse and anomaly detection techniques.

TABLE I. DIFFERENTIATION BETWEEN MISUSE AND ANOMALY DETECTION TECHNIQUE

| Technique | Advantage | Disadvantage |
|---|---|---|
| Misuse Detection | It generates accurately false alarms and fewer in numbers. | It is unable to detect novel attacks and threats. |
| Anomaly detection | Based on audit data it detects the unknown data | It generates very high false alarms |

### B. *Cross domain interfirewall optimization technique*

In this technique the focus is given on working of the firewall in different administrative domains. Optimization is done by removing redundant rules with preserving the privacy of the rules. Let us consider the two different administrative domains CO and IT, F1 denotes the policy on firewall one's outgoing interface and F2 denotes the policy on firewall two's incoming interface given in Fig1. The physical interfaces are denoted as I1, I2 connecting two routers respectively. For any rule in F2,if suppose any packet that match rule r but not matches any rule above r in F2 is discarded by F1.Such a packet never cones to F2 and rule becomes the inter firewall redundant rule. While removing the inter firewall redundant rule we have to consider the privacy of the policies designed at the different administrative domains. One should not disclose the firewall policies to other.

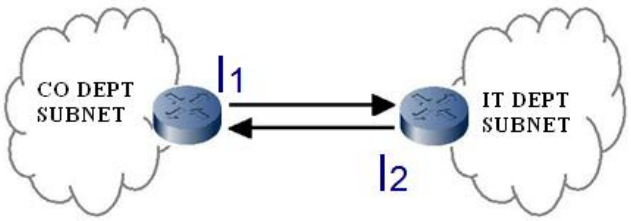


Fig. 1. Interfirewall Connectivity

### C. *Technical challenges*

We have to design such a model which will detect malicious participant in inter firewall optimization technique. When two different administrative domains communicate with each other, threat model is considered as semi honest [7] means there is no any guarantee that corrupted employees present in the organization try disclose the policies of each other or by checking the sequential input they try to know the firewall policies. So our first target is to remove such an attack. We try to find the malicious participant present in the organization without disclosing the privacy of the policies.

Intrusion is nothing but an action which tries to destroy the integrity, privacy of the network related d information so we can find malicious users present in the system by using the one of the intrusion detection technique. As this technique which analyses the audit data from some point of computer system and if any wrong behavior got then they throws it as a intrusion. Intrusion detection is hardware software combination and there are various algorithms are proposed for anomaly classification. For finding malicious activities we can use the ID3 decision tree algorithm [16] in inter firewall optimization technique that differentiate corrupt employee from normal one.

### D. *Classifications Methods for Anomaly Detection*

Anomaly detection first creates the normal behavior data set by observing the network. This technique molds into various categories supervised anomaly detection technique first know a classifier by using labeled instances which belongs to normal and beyond normal case. After this it assign normal and abnormal label to the activity. In this the actual classification goes through following tasks.

- Build Data-In this normal communication in employee and prediction of abnormal actions of employee get build.
- Decision tree-It provides our model transparency.

In Semi-supervised anomaly detection technique uses model first represent normal behavior from training data set and second it test nearly matching test instances which getting generated by learnt model. It uses both labeled and unlabeled data. It comes between supervised learning with properly labeled training data and unsupervised learning but without labeled training data.

Machine learning is one of the scientific discipline that which is related with the design and build out of algorithms and which allow computers to grasp based on data, such as from databases. In machine learning research major focus is given on automatic recognition of complex patterns and on based data makes intelligent decisions.

Unsupervised anomaly detection is a technique which detects anomalies in an unlabeled test data set but it assumes that most of the instances in the data set are normal. Unsupervised functions in data mining are nothing but the association rule grasping. This is a most popular and well researched method for detecting interesting relations between variables in large databases.

Next data mining technique is Clustering. In this technique is used to place data elements into correlated groups without any previous knowledge of definitions of the group. Association model is frequently used for market analysis, which tries to discover relationships among a set of items.



### E. *Proposed technique*

In our technique we try to detect the malicious user in the system which tries to reveal private firewall policies. We first collect the audit data and according to collected data we try to find the malicious user by using anomaly detection technique. Various techniques have been developed for anomaly detection which proposes ID3 Decision tree classifiers. This model differentiates behavior of intruder from the normal behavior. K. Hanumantha Rao, G. Srinivas [16] proposes the implementation of Implementation of Anomaly Detection Technique Using Machine Learning Algorithms and uses the clustering technique with combined approach.

We are going through two steps:

- Implement cross domain optimization protocol.
- Collect the audit data after implementation of optimization protocol by using log files.
- Create the normal behavior data set.
- Afterwards check the normal behavior data set with the audit data.

The rest of paper is organized as follows: Section II tells about related work done in the area of firewall optimization and various clustering techniques. Section III will tell us about the proposed model for malicious user detection. Section IV gives us advantages of or implementation and in section V we conclude our paper.

## II. BACKGROUND AND RELATED WORK

### A. Firewall Redundancy removal

Previously lot of work has been done in the area of inter firewall optimization and intra firewall optimization technique. In the intra firewall redundancy removal technique [2], [3], [4] authors aims to remove redundant rules in the single administrative domain. As the work has been done in single administrative domain the privacy of the policies designed is not concerned. The backward and forward redundant rules identification has been done by Gupta [11].

Prior work also shows that the Inter firewall redundancies removal [5], [6].In this techniques work focused on the redundancy removal in the two different administrative domain. When work has been done on the two different administrative domains the policy privacy should be concerned. Therefore it is applicable to one administrative domain only.

Firewall compressor proposed by X. Liu, E. Torng, and C. Meiners, [12] give us a framework and this framework remarkably reduces the rules in firewall. After implementing this technique firewall semantics gets unchanged. They proposed dynamic programming which gives us optimal solution which compresses one dimensional firewall and in second approach he gives a systematic compression of multidimensional firewall.

To get better performance from firewall Tihomir Katic, Predrag Pale [1], proposed logic to optimize firewall rules. As rules of the firewall get set by the network administrator he have to regular check of the newly designed rule with existing rules. In large organization as there is very huge design of rules is present it is not possible to check the so new rule with existing rule. And in case of less experienced administrator, he finds more difficulty to do this. Administrator finds difficulty in finding the rule redundancies. The proposed technique by the author use log rules and other parameters related to rules in replacement of using IP address, protocol and ports. Authors developed software called FIRO which a command tool related to firewall. The work of firewall is related with IP tables of LINUX Platform.

Many other firewall optimization techniques are proposed by researcher in different areas. Some of the famous firewall optimization algorithms are Trie tree-based algorithm [13], Decision Tree-based algorithm [14], and TCAM-based algorithm [15].

Alex X. Liu Fei Chen [23], proposed us a new technique which removes redundant rules present in inter firewall without having any knowledge of each other's policies. They proposed a protective framework. In this model they work collaboratively and enforce the firewall policies. This solution is far better than proposed Cross Domain Cooperative Firewall (CDCF) because, the encryption technique used in CDFC is little bit slower than three magnitude order proposed by Alex X. Liu Fei[23].

Fei Chen,Bezawada Bruhadeshwar, and Alex X. Liu [7] proposed a cross domain optimization technique with preventing the privacy of firewall policies in cooperative environment. To get the goal of this they propose two methods. 1) They propose a novel approach and give a protocol which detects inter firewall redundancy removal in one firewall. 2) They implemented the protocol and got tremendous result in removing of redundant rules. When they designing this protocol they consider one threat model in that they consider two firewalls are semi honest. For preserving the privacy of the firewall they use the encryption techniques and encrypt the policies of firewall and use Pohling-Hellman Algorithm as encryption technique.

### B. Clustering Technique

Cluster is a group of similar objects grouped together. We can also say that it as the organizing of dataset into similar or with respect to distance or equivalently similarity measure we can create well separated groups. Cluster is nothing but collection of points in the test space in such a way that distance between two points of the cluster is less than the distance between any another two points in the cluster and point not present in it. There are two different types of clustering attributes first is numerical and another is categorical attribute. Numerical attributes are related with ordered values and Categorical attributes are related with unordered values. Ordered values we can say weight or height of person and unordered values means brand of car. Clustering can be used in two different forms:

- Hierarchical clustering.
- Partition (non Hierarchical) clustering.

#### 1) Hierarchical clustering

In hierarchical clustering data are not partitioned into a particular cluster in one step but instead a sequel of partitions takes place. These partitions may get run from a single cluster which contains all objects to n clusters intern each cluster containing a single object [15].Hierarchical Clustering again get subdivided into agglomerative methods, which get extended by series of fusions of different n objects into groups. Secondly the divisive methods, finer grouping of successive n objects get created.

#### 2) Partition (non Hierarchical) clustering

For this technique partition can be of K-means [15] and K-mediod. In the purposed solution by K. Hanumantha Rao, G. Srinivas, Ankam Damodhar and M. Vikas Krishna[16] is based on K-means (Unsupervised) clustering with combination of Id3 Decision Tree type of Classification (Supervised).

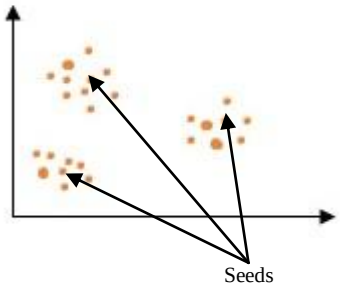


Fig. 2. Un-clustered Data Instances

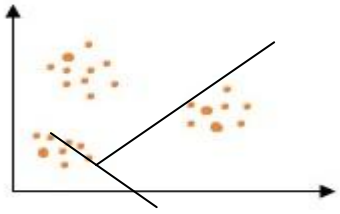

Fig. 3. Resultant Clusters

## III. PROPOSED MODEL

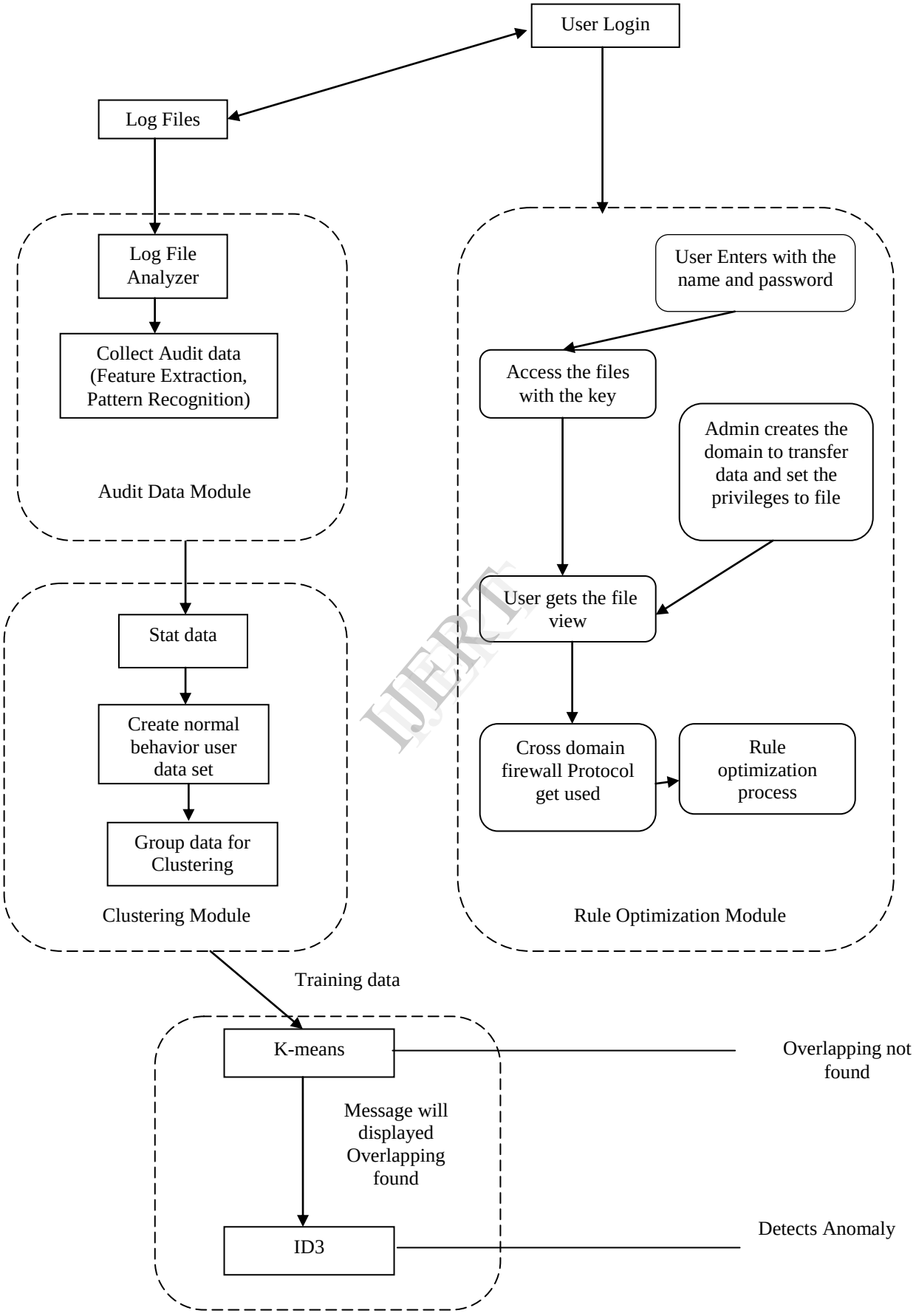


Fig. 4. Proposed Model

In the proposed model, when user login for the system he has its own username and password and gets logged for the system. Administrator set the privileges to the file access for the security of the file. When the key of user get match with the password already saved in database .File get downloaded by the user. This is nothing but the inter firewall cross domain communication process. In this the protocol gets executed and finds the optimized rules present at the firewall and remove the redundancies in the firewall policies. After this we got the increased processor time for the packet. Now we will move forward for the audit model in this model, first we are collecting the audit data in from log files by using log file analyzer and audit data will contain pattern recognition records. In the clustering model by using stat data we are doing following things,

- Collect the normal behavior of users and create set of normal behavior clusters.
- We are grouping similar clusters with respect to behavior of user.

In the future approach which is used by combining K-means and Id3 Decision tree is a unique thing in this,

- We are taking few assumptions on the log file analyzers which is a data driven method based on data.
- K means is a greedy search method gives us a guarantee of at least local criteria functions and gets converted into larger dataset.

On the training dataset K-means clustering get performed and we obtain different K clusters. They represent the similar instances region and cluster's centroid. In next state the k means and ID3 decisions tree gets combined using instances. When computer get connected to network and packet get send and we search whether it is normal or anomaly packet is there. In our proposed approach we are giving input ad feature extracted as normal behavior of users to analyze K means clusters groups the data into clusters like K1,K2,…….,Kn and depend on Euclidean Distance data get group into appropriate clusters. In the K means clusters the data which is not having normal behavior is not get eliminated but in next step the ID3 decision tree is classify the each user clusters. Finally combined result will declare as anomaly or not. If K means never found the overlaps in behavior points then data will directly send to the server and if founds overlap it will pass to decision tree as shown in Fig. 4.

This combined approach of K means and ID3 method get completed in two steps,

1) Training data:In this partining of traning space into different K clusres Like C1,C2,C3,…….,Cn and then Id3 Decesion tree is trained with each K means cluster instance.

2) Testing data:-In this state if any overlapp or subgroups get found the ID3 Decesion tree polishes the desesion boundries by instances partinioning by using if then rules. This may get included in the future scope of the system of malacious user finding.

The advantage of the technic is removing anomalies and disadvantage is time required to find the anomaly. This may get included in the future scope of the system of malacious user finding.

## IV. APPLICATIONS

By proposed design we can find the malicious activities done by the user. If any user try to find the policies of another user and if they are checking the user sequences then this malicious activities get find. This technique will increase the performance of cross domain firewall optimization technique. After optimizing the firewall packet processing time get increased and the cost for communication get also decreased and we are getting good performance firewalls. As the anomaly detection technique we are using is easy to understand it is well suited for the detecting firewall optimization protocol.

## ACKNOWLEGEMENT

First of all I thank my guide to giving me strong support and feedback for the topic. I thank my all colleagues of BSIOTR Wagholi, Pune, for their constant feedback on the paper. We also thank the anonymous reviewers for their valuable comments.

## REFERNCES

[1] Tihomir Katic, Predrag Pale, "Optimization of firewall rules" in IEEE Trans. Information Technology Interfaces, 2007.
[2] Q. Dong, S. Banerjee, J. Wang, D. Agrawal, and A. Shukla, "Packet classifiers in ternary CAMs can be smaller," in Proc. ACM SIGMETRICS, 2006, pp. 311–322.
[3] A. X. Liu and M. G. Gouda, "Complete redundancy removal for packet classifiers in TCAMs," IEEE Trans. Parallel Distrib. Syst., vol. 21, no. 4, pp. 424–437, Apr. 2010.
[4] C. R. Meiners, A. X. Liu, and E. Torng, "Topological transformation approaches to optimizing TCAM-based packet processing systems," in Proc. ACM SIGMETRICS, 2009, pp. 73–84.
[5] E. Al-Shaer and H. Hamed, "Discovery of policy anomalies in distributed firewalls," in Proc. IEEE INFOCOM, 2004, pp. 2605–2616.
[6] L. Yuan, H. Chen, J. Mai, C.-N. Chuah, Z. Su, and P. Mohapatra, "Fireman: A toolkit for firewall modeling and analysis," in Proc. IEEE S&P, 2006, pp. 199–213.
[7] Fei Chen, Bezawada Bruhadeshwar, and Alex X. Liu. "Cross-domain privacy-preserving cooperative firewall optimization" In Proceedings of the IEEE/ACM TRANSACTIONS ON NETWORKING, volume 21, pages 857 – 868, 2013.
[8] E. Al-Shaer and H. Hamed, "Discovery of policy anomalies in distributed firewalls," in Proc. IEEE INFOCOM, 2004, pp. 2605–2616S. M. Metev and V. P. Veiko, Laser Assisted Micro technology, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.
[9] Acharya, S., Wang, J., Zihui Ge; Znati, T.F., Greenberg "Traffic-aware firewall optimization strategies", IEEE ICC, 2006.
[10] Qi Duan and Ehab Al-Shaer., "Traffic-Aware dynamic firewall policy management: techniques and applications" IEEE ICC, 2013.
[11] P. Gupta, "Algorithms for routing lookups and packet classification," Ph.D. dissertation, Stanford Univ., Stanford, CA, 2000.
[12] A. X. Liu, E. Torng, and C. Meiners, "Firewall compressor: An algorithm for minimizing firewall policies," in *Proc. IEEE INFOCOM*, 2008.
[13] Fengjun S, Yingjun P, and Xuezeng P. "Study on an absolute aon-collision hash ip classification algorithms". In Proceedings of the Journal of Communications, 2005.
[14] Singh S, Baboescu F, and Varghese G. "Packet classification using multidimensional cutting". In Proceedings of the ACM SIGCOMM, 2003.
[15] Text Book of Data mining Techniques by Arun K Pujari Universities Press (India) Private Limited.
[16] K. Hanumantha Rao, G. Srinivas, Ankam Damodhar and M. Vikas Krishna, "Implementation of anomaly detection technique using machine learning algorithms' in International Journal of Computer Science and Telecommunications in Volume 2, Issue 3, June 2011.
[17] Lazarevic, A. Ozgur, L. Ertoz, J. Srivastava, and V. Kumar, "A comparative study of anomaly detection schemes in network intrusion detection," Proc. SIAM Int'l Conf. Data Mining, May 2003.
[18] W. Lee, S. J. Stolfo Data Mining Approaches for Intrusion Detection.

[19] Rui Xu, Donald Wunsch II, "Survey of Clustering Algorithms", IEEE in Neural Networks 16(3) (2005).
[20] W. Lee and D. Xiang, "Information-theoretic measures for anomaly detection," in Proc. IEEE S&P, 2001, pp. 130–143.
[21] R. Sekar, A. Gupta, J. Frullo, T. Shanbhag, A. Tiwari, H. Yang, and S. Zhou, "Specification-based anomaly detection: A new approach for detecting network intrusions," in Proc. ACM CCS, 2002, pp. 265–274.
[22] C. Kruegel, T. Toth, and E. Kirda, "Service specific anomaly detection for network intrusion detection", in Proc. ACM SAC, 2002, pp. 201–208.
[23] A. X. Liu and F. Chen, "Collaborative enforcement of firewall policies in virtual private networks," in Proc. ACM PODC, 2008, pp. 95–104.