International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 7 Issue 1 –MARCH 2014.

Zombie Based Attack Detection and Prevention in Web Applications

Saranya A
Department of IT,
MIT, Anna University
saranyamit12@gmail.com

Dr. Kola Sujatha P (Sr.Gr)
Assistant Professor
MIT, Anna University,
kola_sujatha@yahoo.com

Dr. Kannan A
Professor, Anna University
kola_sujatha@yahoo.com

***Abstract*- Most of the e mail servers and website loggers are facing many problems in their activities. Zombie based vulnerabilities are injected the attacks into the system through network by taking the availability of the server and response return to the clients. Zombie systems sends spam messages to the target websites and it grabs necessary information through the mails and also it injects the DoS (Denial of service) attack in the system. Every mail inputs are recorded and captured using KeyLogger (KL) algorithm. The captured inputs are given to the Principal Component Analysis (PCA) and enhanced tag extraction algorithm which gives the extracted features to identify the behaviors of spam. We proposed a ZDAP Spam Detection Algorithm (Zombie based Attack Detection and Prevention). This algorithm is executed in MIME (Multipurpose Internet Mail Extension) type protocol, tags are extracted and compared with existing tags by near duplicate matching algorithm. Some of the efficient network parameters are used for evaluating the delay or modification of the content using ZDAP Denial of Service (DoS) algorithm. The attack detection method needs to classify the attack types with the help of predicted parameters.**

***Key Words*- Spam, Denial of Service, Zombie Vulnerabilities, Key Logger Algorithm, Principle Component Analysis.**

I. INTRODUCTION

Network intrusions that were caused due to incoming packets in the network, which performs the malicious activities such ascending spam, denial of service attacks, or even attempts to crack into computers. To address these [2] vulnerabilities intrusion detection and prevention system is implemented. Zombie is a compromised system that is coupled with the internet that access and responses are taken by a cracker, computer virus or Trojan horse under the remote direction. Botnets of zombie computers are defined as group of compromised machines used to spread the e-mail spam and launch denial of service attacks. Most [9] of the owners of zombie computers are unaware that their system is compromised and it makes to other computer to compromise. E-mail spam, [16] also known as Junk E-mail or Unsolicited Bulk E-mail (UBE) or Unsolicited Commercial E-mail (UCE), is a subset of electronic spam involving nearly identical messages sent to numerous recipients by e mail. Spam is flooding the Internet with many copies of the same message, in an attempt to force the message on people who would not otherwise choose to receive it.

Spams messages are formed and sent by the un authorized users and spammers. The main objective of spammers is to advertise their products to make profits like using some great words, lottery prize winning and click the below link messages. This is done using the concept of a-mail appending or epending, it leads to collects or records the personal information. An unwanted email messages increase the total cost of operating the networks of computers which form the Internet. Spammers inject the spam to disrupt a network by the intension of crashing mail servers and filling up hard drives. On the other hand, if any web servers or organizations being classified as spam sites, form that sites cannot receive any mail or information to others sites. These actions are noted and inform the impact of the spammers and spam. The main challenge of spam and [17] Dos attack detection problems are filtered by using new ways it leads to improve the economic benefits of sending spam. A Denial-of-Service attack (DoS attack) or Distributed Denial-of Service attack (DDoS attack) is a challenge to make a machine or network resource unavailable to its planned users and to the targeted web sites. DoS [12] attacks normally achieved in large network and also the high profile web servers that are bank, healthcare institutions, credit card payments and root servers. Many spam detection techniques are used to filter the e-mails, [2] even though the spammers inject the spam messages through some of the wholes which is presented in network and resources handled by the users. A large amount of packets are (useless information) are sent to the particular or targeted web sites to launch the denial of service attacks. This type of attacks creates an unwanted traffic in the network and takes the response of the system out.

II. RELATED WORK

Network based attack injection techniques were used to test the system response and effectiveness. That proposes an automatic discovery of vulnerabilities in software components using AJECT (Attack Injection Tool) tool. This AJECT attack tool is used to replicate the irregularity and to assist the removal of the error. To evaluate the worth of this approach, is to obtain the several attack injections were performed with 16 widely available POP (Post of Protocol) and IMAP (Internet Message Access Protocol) servers but it gather that three of the flaws were related to client resource management only handled. This system is used to overcome the types of attacks presented in the web server.

Z. Duan et al,[7] Behavioral characteristics of spammers and their network reachability properties proposes More importantly, they also discard glow on the design of future email delivery architecture that can be used to inherently resist spam activities. In the current SMTP (Simple Mail Transfer Protocol) based email delivery architecture is not suitable for detect the spam of holding longer time in online, spammers can disappear the spam results after pushing (go offline) the overflow of spam to receivers. This is confirmed by our results is to a large part of spammers send spam only within a short period of time, and more disturbingly, some difficult spammers utilize their short lived networks for spamming activities. Our conclusion is to suggest that in order to effectively control spam, we must hold spammers responsible, force them to stay online for longer periods of time while throttling their spamming rates, and limit the spammer’s elasticity in commonly varying their locations and/or Internet Service Providers. Design of new pull-based email delivery architectures, such as DMTP

Tsung Huan Cheng et al, [25] Clustering analysis of network traffic for protocol and structure-independent botnet detection defines the independent of botnet C&C protocol and structure, and it does not requires any of prior knowledge of botnets to analyses the attack behavior. This is used for notice real-world botnets (IRC-based, HTTP-based, and P2P botnets) and has a very low false positive rate. Since bots within the same botnet will display the similar communication pattern and similar malicious activity patterns, it is not possible to separate the attacks.

Z. Duan et al [6], DMTP: Controlling spam through message delivery differentiation July 2007 proposed a methodology, which grants the receivers to achieve better control over the network and it defines how the messages from different senders should be delivered on the Internet in a particular time. This concept presents the platform for e-mail system can compel spammers to stay online for longer periods of time, which may considerably get better the performance and DTMP can be effortlessly deployed on the Internet incrementally. But it leads the unwanted Internet traffic in the network due to the attack phase.

I.S. Kim M.H. Kim et al[14,] Agent-based honeynet framework is used for defending servers in campus networks proposes honeypots that proceed to baits for attackers and can spot zero-day attacks and supply researchers intending to develop the security that is used for protects servers from new types of internet worms successfully, not including the use of signatures but it requires the client’s assets does not protect servers from indefinite internet worms.

This system proposes [4] a multiphase distributed vulnerability detection, measurement, and countermeasure selection mechanism, which is built on attack graph-based systematic models. It significantly improves attack detection and moderate. In host-based IDS solutions are wanted to be included and to cover the whole spectrum of IDS in the system.

In this technique, the trouble of detecting the machines that perform a sending spam has been addressed. This move toward to involves very low isolation intrusion as only the boundary of network flow data will be defined. Sending back to a TCP/SYN-ACK packet, and waiting for a TCP/ACK [13] packet in response from the sender address. However, the sender address is fake there is response in no way to comes true. These half-open connections oversupply the number of accessible connections the server is clever to make, observance it from responding to genuine requests until after the attack ends.

Spam Zombie attack detection system named SPOT by monitoring outgoing messages in a network. SPOT (Sequentially Probability Ratio Test) was designed [27] based on a simple and powerful statistical tool named sequential probability ratio test to detect the compromised machines. SPOT tool is based on the count (CT) and percentage (PT) of spam messages sent by internal machines. This system was suitable for spam detection and also it leads network traffic tends to poor responsiveness.

III. PROPOSEDD WORK

KeyLogger (KL) algorithm has been established for recording and [10],[15] capturing the keystrokes, inputs, mouse clicks. The captured inputs are given to the Principal Component Analysis (PCA) which extracts the features to identify the behaviors of spam. With these behaviors, whether the particular message is spam or not is identified by using ZDAP Spam Detection algorithm.

For MIME (Multipurpose Internet Mail Extension) type protocol, tags are extracted and compared with existing tags by near duplicate matching algorithm. This helps to detect [19] the message as spam or not. The extracted tags stored in tree manner for easy storage and retrieval. Some of the efficient network parameters are used for evaluating the delay or modification of the content using ZDAP Denial of Service algorithm. Attack detection method needs to classify the attack types with the help of predicted parameters. The detected (infected) messages are removed or reduced based on the level of severity. Then the notification is send to both victim and offenders for creating awareness. Denial of service attack is identified by the following network parameters such as, Number of Packets (np), Number of Bytes (bt), Average Packet Size.

Packet Rate per second (PRs), Byte Rate per second (BRs), Time Interval Variance (tc) and Packet Size Variance (pc) are calculated as follows,

$$\text{PRs} = np \times 1 \div (te - ts)$$

$$\text{BRs}= bt \times \ 1 \div (te - ts)$$

$$tc = \sqrt{\sum} \ (tn - t^{\circ})2 \div n$$

$$pc = \sqrt{\sum}(pn - p^{\circ})2 \div n$$

The zombie based attack detection and prevention system is explained in Fig 1. here with the appropriate component description.

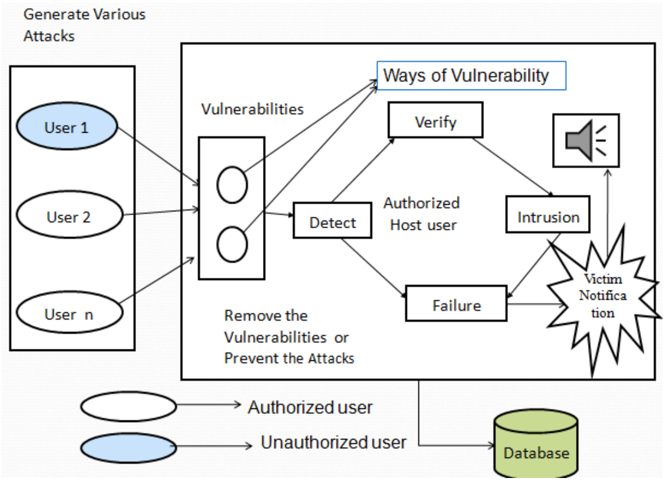


Fig 1. Architecture Diagram for ZDAP System

Various numbers of users are entered into the system with or without malicious activities, once the message has been detected the key logger algorithm captures the inputs and forwards to the verification state. Then the message operation is verified that is any abnormal activities are presented or not. If any of the malicious activities presented then messages are move to the intrusion. In this state message inputs are deeply analyzed with the extracted features, and if any of the features presents in the message input it becomes an intrusion so its forwards to failure. Every action are taken to account for future reference that is the same types of activities or same users enter into the system ip alarm indicate this particular user may leads to intrusion. Then the notification action is performed to both the victims and offenders.

A. DEFINITION OF <anchor>

The tag <anchor> is used to transform the mail id and url from the web reference. It is mainly used for domain name analysis and transformation. Anchor tag is to reduce the false positive when the email abstraction is limited. [22] Anchor tag is represented in this format <a href="http://www.ExampleOnly.com/">. The email abstraction is matched with known spam mail or url. A predefined threshold value is used to classify the spam emails.

The corresponding e-mails are compared with a threshold value, if the tag is smaller than the threshold value it named as spam mail. Tag length is [5] calculated by how many tags are presented in the e-mail abstraction. The e- mail abstraction are matched with the previous pattern if they are exactly matched that particular mail id is separated and recorded.

B. PREFIXING OF <anchor>

This tag is used for reduce the probability of matched unsuccessful spam messages. The prefixed tags [21] are reordered in the process of reordering the original tag position. In this process if two e-mails contain the same tag length for worst case they differ from the last tag only. For this type of problems destroy the original abstraction and rearrange the position of tags. This process ensures that the newly assigned position numbers of e-mail abstractions with the same number of tags are completely identical. As such, the matching process can be accelerated without violating the definition of near-duplicate in this system. Fig 2 describes the data flow diagram of zombie based attack detection and prevention system with all sub modules and actions.

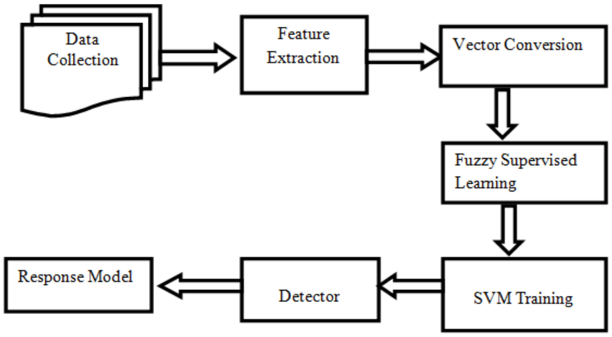


Fig.2 ZDAP Data Flow Diagram

C. FEATURE EXTRACTION

Tag Extraction is performed based on the Principal Component Analysis and Map Reduce function. The necessary [23] tags extracted in e-mail defined as <html>, </html>, <head>, </head>, <body>, </body> are removed. The long e- mails also extracted with the appropriate tags without affecting the effectiveness. of near-duplicate.

D. FUZZY SUPERVISED LEARNING

Fuzzy learning is based on the machine learning approach [1],[18]. It infers a function from the labeled training data. In supervised learning, the input samples and output set results are known by the user but training the samples is difficult to achieve the correctness. If the corresponding training data is trained again until it obtains the correct result. Support Vector Machine (SVM) is used to [11] classifies two different types of labels that is attack and normal e-mails. SVM is highly differentiating the process with the hyperplane integration.

IV. ALGORITHM FOR ZDAP SYSTEM

1.1 ZDAP SPAM DETECTION ALGORITHM

Input: Features, the e-mail with *text/html* MIME type

Output: Attack Types, Attacker details, the E-mail Abstraction (EA)

```
 Process:
      //Tag Extraction Phase
 Step 1: Transform each tag to <tagname>
 Step 2: Transform each paragraph of text to <mytext />
 Step 3: AnchorSet = union of all <anchor> tags
 Step 4: EA = concatenation of <tagname>
 Step 5: Preprocess the tag sequence of EA
      //Tag Reordering Phase
 Step 6: for (each tag of EA) // pn - position number
 Step 7: tag.new_pn = ASSIGN_PN(EA.taglength , tag.pn)
 Step 8: Put the tag to the new position tag.new_PN
 Step 9: EA = concatenation of <tagname> with new_PN
      // Prefixing <anchor> Phase
 Step 10: if (EA.tag_length < Lth_short) then
 Step 11: Append AnchorSet in front of EA
 Step 12: return EA
```

E. NEAR-DUPLICATE MATCHING SCHEME

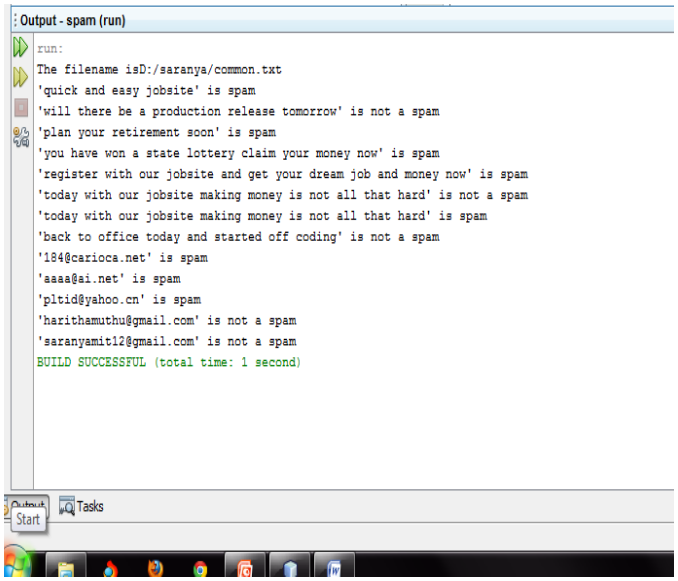
The near-duplicate matching algorithm is defined for spam detection is to maintain a known spam database, with the help of users' feedback, to block spam with the similar content. With the help of spam database the users can easily identify what spam messages occur subsequently. New techniques are used to detect the spammers efficiently and also detect new types of worms and spam activities.

F. BLACKLIST & WHITELIST

Blacklist is used to block or stop the spammers from performing spammer task like posting to forums, blogs, and sending emails. It collects the sender addresses contained spam mail and also the spam message content of the mail. Whitelist is used to allow the messages that not affected by spammers. It prevents the e-mails from the unwanted threats and attacks.

V. RESULTS AND PERFORMANCE ANALYSIS

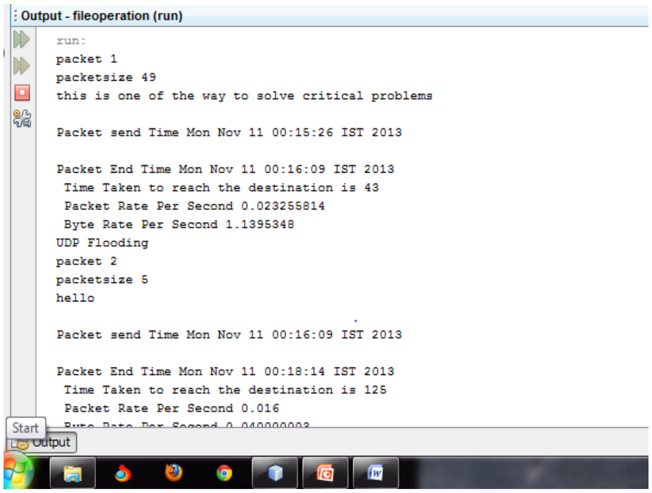
The results obtained in ZDAP system is significantly improved the performance of attack detection and mitigation rate by using efficient techniques.



```
run:
The filename isD:/saranya/common.txt
'quick and easy jobsite' is spam
'will there be a production release tomorrow' is not a spam
'plan your retirement soon' is spam
'you have won a state lottery claim your money now' is spam
'register with our jobsite and get your dream job and money now' is spam
'today with our jobsite making money is not all that hard' is not a spam
'today with our jobsite making money is not all that hard' is spam
'back to office today and started off coding' is not a spam
'184@carioca.net' is spam
'aaaa@ai.net' is spam
'pltid@yahoo.cn' is spam
'harithamuthu@gmail.com' is not a spam
'saranyamit12@gmail.com' is not a spam
BUILD SUCCESSFUL (total time: 1 second)
```

Fig3. Spam Message Analysis with Features

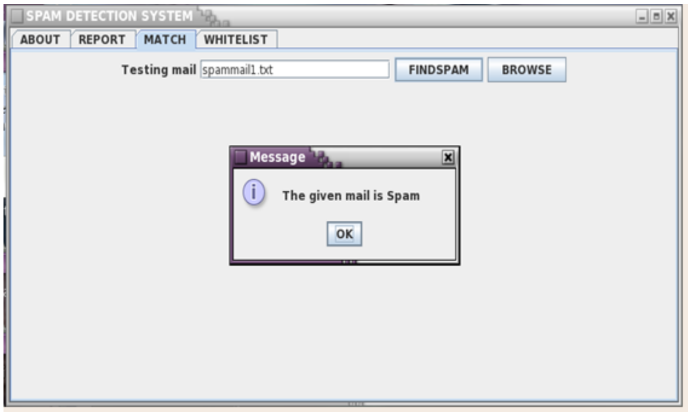
E –mail contents are deeply analyzed with the extracted features shown in Fig 3.



```
run:
packet 1
packetsize 49
this is one of the way to solve critical problems

Packet send Time Mon Nov 11 00:15:26 IST 2013

Packet End Time Mon Nov 11 00:16:09 IST 2013
 Time Taken to reach the destination is 43
 Packet Rate Per Second 0.023255814
 Byte Rate Per Second 1.1395348
UDP Flooding
packet 2
packetsize 5
hello

Packet send Time Mon Nov 11 00:16:09 IST 2013

Packet End Time Mon Nov 11 00:18:14 IST 2013
 Time Taken to reach the destination is 125
 Packet Rate Per Second 0.016
 Byte Rate Per Second 0.040000003
```

Fig 4. Denial of Service Attack

Each packet is analyzed and calculated the parameters of DoS attacks is given in Fig.4 If the parameter values are changed in the receiver end it will indicate the DoS attack is presented.



Testing mail spammail1.txt

The given mail is Spam

Fig 5. Spam Mail Matching Handler

The spam detection system has been tested with around 440 e-mails illustrated in Fig 5. The true positive rate i.e. the possibility that a real spam is classified as a spam was about 96 %. The false positive rate i.e., the possibility that a real ham is misclassified as spam was about 4 % which is a negligible amount. The system performance is shown in table1.

Table 1 System Performance Analysis

E- Mail Type	No. of E- Mails		Spam	Ham
	For Testing	For Training		
Text /html	320	100	95	5
Text plain	120	80	77	3

True Positive rate (TP): 96%

False Positive rate (FP): 4%

Compared to the existing system our proposed work is efficiently detected 96% of spam mails.

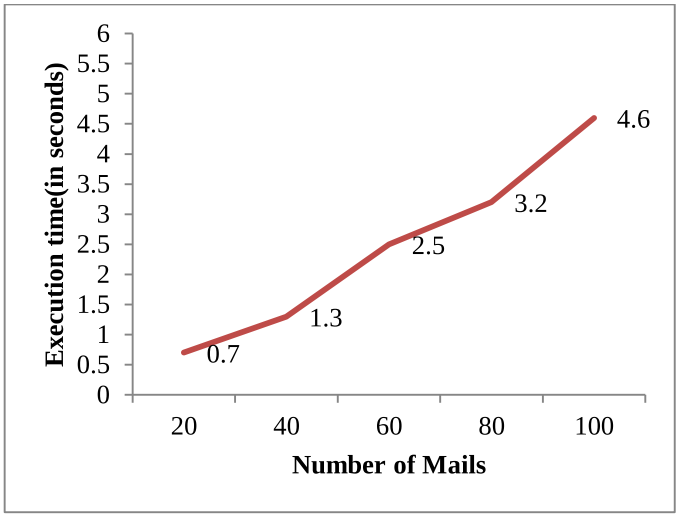


Fig 6. Execution Time of e- mail Abstraction Generation of the System

In fig 6 the execution time of 100 e-mails is monitored for generating the e-mail abstraction and it took about 4.6 seconds to generate e-mail abstraction of 100 e-mails.

VI. CONCLUSION AND FUTURE WORK

An effective way to detect and prevent the zombie based vulnerabilities by monitoring outgoing messages in web the application using key logger algorithm and ZDAP Spam and DoS (Denial of Service) is proposed. This system automatically detects the compromised machine in a network or a targeted server. This detection tool mainly implemented in E-mails and web applications like website monitoring, Web logging.

Every input message is analyzed and the abnormal behavior is detected. Any of the illegal activities are presented in the system it is automatically retransmits the message to the source point and records the ip address or mail id of the owner system. Then the notification is sent to the both ends. In future if the same person may enter into the system the ip alarm will indicates that the message is being vulnerable. There are many new types of attacks are formed while sending and receiving the advertisement. Most of the intrusion detection and prevention systems are not familiar with the steganography based issues and encrypted messages. So the extension of this work is to identify and rectify these issues by using efficient algorithms

REFERENCES

[1] Ahmed H. Fares and Mohamed I. Sharawy “Intrusion Detection: Supervised Machine Learning” Journal of Computing Science and Engineering, Vol. 5, No 4, pp. 305-313 December 2011

[2] Alexander G, Tartakovsky Aleksey, Polunchenko S, “ Efficient Computer Network Anomaly Detection by Changepoint Detection Methods” IEEE Journal, Vol. 7, No. 1, February 2013

[3] Chen Y, Hwang K, and Ku W, “Collaborative Detection of DDoS Attacks over Multiple Network Domains,” IEEE Transations on Parallel and Distributed Systems, 2007

[4] Chun-Jen Chung, Pankaj Khatkar ,“NICE: Network Intrusion Detection and Countermeasure Selection in Virtual Network Systems” IEEE Transactions Vol. 10, No 4, july/august 2013

[5] Clayton R, “E-mail Traffic: A Quantitative Snapshot,” Proceedings of the Fourth Conf. E-mail and Anti-Spam (CEAS), pp. 253-267, 2007.

[6] Duan Z, Dong Y and Gopalan, "DMTP: Controlling spam through message delivery differentiation". Computer Networks July 2007.

[7] Duan Z, Gopalan K, and Yuan X. “Behavioral characteristics of spammers and their network reachability properties.” Department of Computer Science, June 2006

[8] El-Atawy A, Al Shaer E, Tran T and Boutaba R “Adaptive early packet filtering for defending firewalls against DoS attacks”,in Proc IEEE INFOCOM, pp. 2437–2445, Apr. 2009.

[9] Françcois J, El Atawy A, Al Shaer E and Boutaba R, “A collaborative approach for proactive detection of distributed denial of service attacks,” in Proceedings, Vol. 11, 2007.

[10] Information about keylogger activities and algorithms Available at http://coolhackingtricks.blogspot.in/20 11 III Iwhat-is-keylogger.html

[11] Ivor W, Tsang, James Kwok T, Pak-Ming Cheung, “Core Vector Machines: Fast SVM Training on Very Large Data Sets” Journal of Machine Learning Research 6 2005.

[12] Jerome François, Issam Aib, Member IEEE, And Raouf Boutaba “Firecol: A Collaborative Protection Network For The Detection Of Flooding Ddos Attacks” IEEE/ACM Transactions On Networking, Vol. 20, No 6, December 2012

[13] Kavisankar L, Chellappan C, “A Mitigation model for TCP SYN flooding with IP Spoofing” IEEE-International Conference on Recent Trends in Information Technology, ICRTIT June 2011

[14] Kim I.S, Kim M.H “Agent-based honeynet framework for protecting servers in campus networks” School of Computer Science and Engineering, May 2012.

[15] Mohammad Avita Katal , Goudar R.H, Singh D.P and Asit Tyagi, “ A Framework for Detection and Prevention of Novel Keylogger Spyware Attacks” Proceedings of7'h International Conference on Intelligent Systems and Control ISCO 2013

[16] Pera M.S and Ng Y.K , “Using Word Similarity to Eradicate Junk E-mails,” Proceedings of 16th ACM International Conference Information and Knowledge Management (CIKM), pp. 943-946, 2007

[17] Pooja Bhoria, Kanwal Garg, “Determining feature set of DOS attacks” International Journal of Advanced Research in Computer Science and Software Engineering Vol. 3, Issue 5, May 2013

[18] Pradipta Maji and Partha Garai, “Fuzzy–Rough Simultaneous Attribute Selection and Feature Extraction Algorithm” IEEE Transactions On Cybernetics, Vol. 43, No. 4, August 2013

[19] Radhika Goel, Anjali Sardana, and Ramesh C. Joshi “Parallel Misuse and Anomaly Detection Model” International Journal of Network Security, Vol. 14, No 4, PP.211-222, July 2012

[20] Razieh Baradaran, Mahdieh Haji, Mohammad Hosseini, “Intrusion Detection System based on Support Vector Machine and BN-KDD Data Set” 7th SASTech 7-8 March, 2013.

[21] Rohan M, Julie J.C.H. Ryan, René van Dorp J, “Detecting Targeted Malicious Email” Copublished by the IEEE Computer and Reliability Societies May/June 2012

[22] Rung-Ching Chen, Kai-Fan Cheng and Chia-Fen Hsieh, “Using Rough Set And Support Vector Machine For Network Intrusion Detection” International Journal of Network Security & Its Applications (IJNSA),Vol. 1, No 1, April 2009

[23] Shailendra Singh, Sanjay Silakari and Ravindra Patel “An efficient feature reduction technique for intrusion detection system” International Conference on Machine Learning and Computing IPCSIT Vol. 3, 2011

[24] Thwe Thwe Oo, Thandar Phyu , “A Statistical Approach to Classify and Identify DDoS Attacks using UCLA Dataset” International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Vol. 2, No 5, May 2013

[25] Tsung Huan Cheng, Ying-Dar Lin, Yuan-Cheng Lai, “Evasion Techniques: Sneaking through Your Intrusion Detection/Prevention Systems” IEEE Communications Surveys & Tutorials, Vol. 14, No 4, 2012

[26] Xiao Feng Wang and Michael Reiter K, “Using Web-Referral Architectures to Mitigate Denial-of-Service Threats” in IEEE Transactions On Dependable And Secure Computing, Vol.7, No. 2, April-June 2010.

[27] Zhenhai Duan, Peng Chen, Fernando Sanchez, “Detecting Spam Zombies by Monitoring Outgoing Masseges”, IEEE Transactions Depentable and Secure Computing Vol. 9 No.2, pp 198-210, March/April 2012.