Network Intrusion Detection technique for Regular Expression Detection using DPI in wireless network.

Mr. Girish M. Wandhare
Student of Master of Engineering
SKNCOE, Pune
Maharashtra, India,
girishwandhare@gmail.com

Prof. S. N. Gujar
IT Department, SKNCOE, Pune
Maharashtra, India
satishgujar@gmail.com

Dr. V. M. Thakare
S.G.B. Amravati University, Amravati
Maharashtra, India
vilthakare@yahoo.co.in

***Abstract*— Deep Packet Inspection (DPI) is becoming more widely used in virtually all applications or services like Intrusion Detection System (IDS), which operate with or within a network. DPI analyzes all data present in the packet as it passes an inspection to determine the application transported and protocol. Deep packet inspection typically uses regular expression matching as a core operator. Regular expressions (RegExes) are used to flexibly represent complex string patterns in many applications ranging from network intrusion detection and prevention systems (NIDPSs). Regular expressions represent complex string pattern as attack signatures in DPI. It examine whether a packet's payload matches any of a set of predefined regular expressions. There are various techniques developed in DPI for deep packet inspection for regular expression. We survey on these techniques for further improvement in regular expression detection in this paper. We implement technique to block regexp packet such as DOS attack. In the result we found that it is possible to reduce RegExp transaction memory required in network intrusion detection. We implement this technique with possible use of DPI techniques in the wireless network.**

Keywords*— Deep Packet Inspection(DPI); Regular Expression(RegExp); Deterministic Finite Automata(DFA); LaFA; StriFA; CompactDFA; Tcam; DFA/EC; Snort; Bro.

1. INTRODUCTION

In most of the applications regular expressions (RegExes) are used to flexibly represent complex string patterns in many applications, such as network intrusion detection and prevention systems (NIDPSs), Compilers and DNA multiple sequence alignment [1].

Intrusion detection is the process of monitoring the events occurring in a computer system or network and analyzing them for signs of possible incidents, which are imminent threats of violation of computer security policies and standard security practices. Intrusion detection and prevention systems (IDPS) are primarily focused on identifying possible incidents in packet, logging information about these incidents, attempting to stop and reporting them to security administrators [3].

Deep Packet Inspection (DPI) is a technology that enables the network owner to analyze internet traffic, throughout the network, in real-time and to differentiate them according to their payload [3]. Newer DPI systems, such as Snort [11], and Bro [10], use rule-sets consisting of regular expressions, these systems are more expressive, efficient, and compact in specifying attack signatures [4]. Regular expressions represent complex string pattern as attack signatures in many applications. There is no current regular expression detection system is capable of supporting large RegExp set ; and even larger RegExp sets are expected in future with high speed demand [1].

LaFA [1] requires less amount of memory due to these three contributions: 1) providing specialized and optimized detection modules to increase resource utilization; 2) systematically reordering the RegExp detection sequence to reduce the number of concurrent operations; 3) sharing states among automata for different RegExes to reduce resource requirements.

The TCAM introduces three novel techniques to reduce TCAM space and improve RE matching speed: Transition sharing, table consolidation, and variable striding. These three techniques can achieve potential RE matching throughput of 10–19 Gb/s [2]. StriFA [6] technology presents the stride finite automata; it's a novel finite automata family, used to accelerate both string matching and regular expression matching. This technique implemented in software and evaluated based on different traces.

Compact DFA [6] proposed method is to compress DFAs by observing that the name used by common DFA encoding is meaningless. Compact DFA technique applies to a large class of automata, which can be categorized by simple properties. With a TCAM [2] the throughput of compact DFA reaches to 10 Gb/s with low power consumption. Extended Character set DFA [3] focused on reducing the memory storage requirement of DFAs, and it can be divided into the following categories: reducing the number of states, reducing the number of transitions, reducing the bits encoding the transitions, and reducing the character-set.

We did survey on above intrusion detection techniques to understand the literature related to Intrusion detection using deep packet inspection techniques for RegExp matching, used to improve IDS technique in wireless networks and improved regexp detection technique implementation and comparatively evolution of existing technique with the new propose technique by considering different parameter such as bandwidth requirement, speed of intrusion detection etc. The details describe in following section.

2 PAPER ORGRNIZATION

The rest of the paper organized as follows. An overview of DPI is given in Section 3. Here we describe Detail working of DPI and its levels. Section 4 describes the use of regular expression in DPI. Section 5 gives related DPI techniques limitations in RegExp detection. Ad-hoc network describes In section 6. Dos attacks and its types describes in section 7. Our proposed system details describe in section 8. Section 9 includes advantages of our proposed system. Result and comparison with existing system is given in section 10. Section 11 conclude our work and contains our future work.

3 DEEP PACKET INSPECTION (DPI)

Deep Packet Inspection (DPI) is a technology that enables the network owner to analyses internet traffic, through the network, in real-time and to differentiate them according to their payload. Since, this has to be done on real time basis at the high speeds it cannot be implemented by processors or switches on software running. It has only become possible in the last few years through advances in computer engineering and in pattern matching algorithms [2].

Originally the Internet protocols required the network routers to scan only the header of an Internet Protocol (IP) packet. The packet header contains the origin and destination address and other information relevant to moving the packet across the network. The “payload” or content of the packet, which contains the text, images, files or applications transmitted by the user, was not considered to be a concern of the network operator. DPI allows network operators to scan the payload of IP packets as well as the header. Figure 1[8] shows the domain of packet inspection required in internet protocols and in DPI [8].

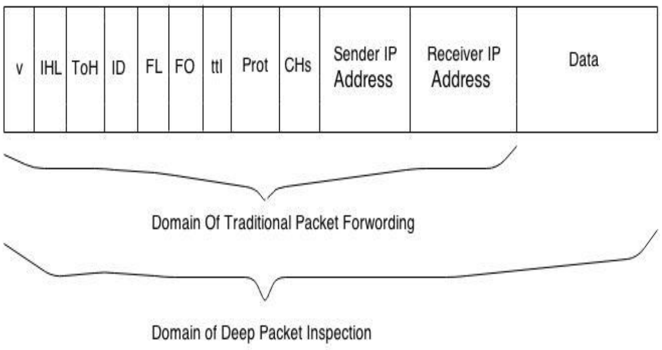
v
IHL
ToH
ID
FL
FO
ttl
Prot
CHs
Sender IP Address
Receiver IP Address
Data
Domain Of Traditional Packet Forwording
Domain of Deep Packet Inspection

Figure 1: Domain of Deep Packet Inspection [8]

DPI systems use expressions to define patterns of interest in network data streams. The equipment is programmed to make decisions about how to handle the packet or a stream of packets based on the recognition of a regular expression or pattern in the payload. This allows networks to classify and control traffic on the basis of the content, applications, and subscribers [8].

3.1 Levels of Packet Inspections

Many of the functions provided by DPI technology have been available before to limited extent depending on the level of packet analysis [1]. Packet inspection technologies that have been in use in networking environments can be classified in three classes. .

Shallow Packet Inspection: Shallow packet inspection (SPI) examines the headers of the packets (which is the information placed at the beginning of a block of data, such as the sender and recipient's IP addresses), as opposed to the body or “payload” of the packet [8].

Medium Packet Inspection: Medium Packet Inspection (MPI) is typically used to refer to ‘application proxies’, or devices that stand between end-users’ computers and ISP/Internet gateways [8].

Deep Packet Inspection: Deep Packet Inspection (DPI) technologies are intended to allow network operators precisely to identify the origin and content of each packet of data that passes through the networking hubs.

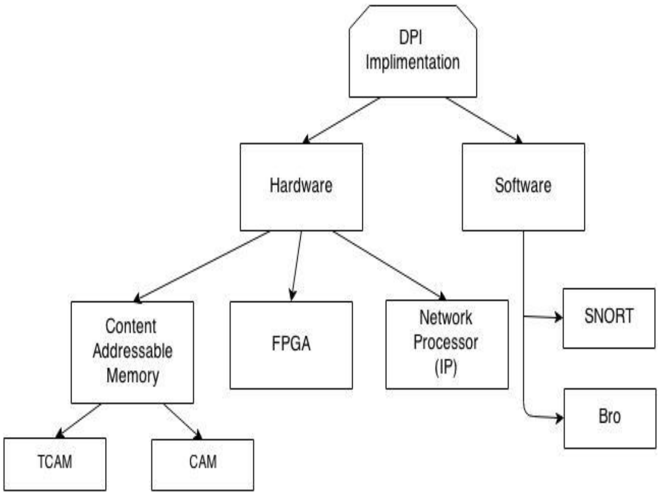

Figure 2: DPI Implementation [7].

4 REGULAR EXPRESSION

Deep packet inspection typically uses regular expression (RE) matching as a core operator. It examine whether a packet's payload matches any of a set of predefined regular expressions. REs are fundamentally more expressive, efficient, and flexible in specifying attack signatures. Prior RE matching algorithms are either software base or field-programmable gate array (FPGA) based [1].

RegExes consist of a variety of different components such as character classes or repetitions [1]. Due to this variety, it is hard to identify a method that is efficient for concurrently detection of all these different components of a RegExp. Most RegExes share similar components.

In the traditional FA, a small state machine is used to detect a component in a RegExp. This state machine is duplicated since the similar component may appear multiple times in different RegExes. Furthermore, most of the time, RegExes sharing this component cannot appear at the same time in the input. As a result, the repetition of the same state machine for different RegExes introduces redundancy and limits the scalability of the RegEx detection system. Figure 3 shows example illustrating the transformation from a RegEx set R into the corresponding LaFA technique [1].

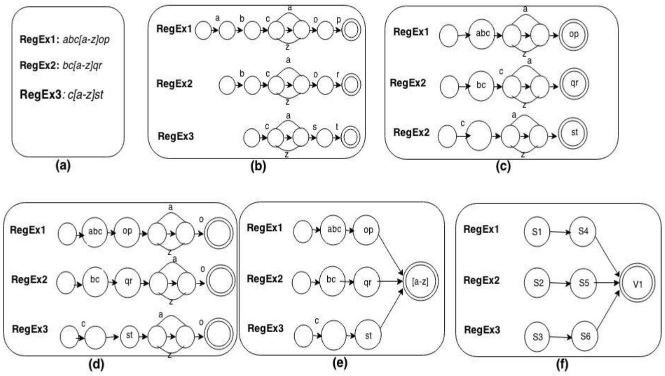

Figure 3: Example illustrating the transformation from a RegEx set R into the corresponding LaFA. (a) RegEx set. (b) NFA corresponding to. (c) Separation of simple strings. (d) Reordering of the detection sequence. (e) Sharing of complex detection modules. (f) LaFA representation of the RegExes [1].

4.1 Why Regular Expression Detection

Currently, regular expressions are replacing explicit string patterns as the pattern matching language of choice in packet scanning applications. Their widespread use is due to their expressive power and flexibility for describing useful patterns. For example, in the Linux Application Protocol Classifier (L7-filter), all protocol identifiers are expressed as regular expressions. Similarly, the Snort intrusion detection system has evolved from no regular expressions in its rule set in April 2003 to 1131 out of 4867 rules using regular expressions as of February 2006. Another intrusion detection system, Bro [10], also uses regular expressions as its pattern language [9].

As regular expressions gain widespread adoption for packet content scanning, it is imperative that regular expression matching over the packet payload keep up with the line-speed packet header processing. Unfortunately, this requirement cannot be met in many existing payload scanning implementations. For example, when all 70 protocol filters are enabled in the Linux L7-filter [1], we found that the system throughput drops to less than 10Mbps, which is well below current LAN speeds. Moreover, over 90% of the CPU time is spent in regular expression matching, leaving little time for other intrusion detection or monitoring functions. On the other hand, although many schemes for fast string matching have been developed recently in intrusion detection systems, they focus on explicit string patterns only and cannot be easily extended to fast regular expression matching [9].

4.2 Working of Regular Expression in DPI

A regular expression describes a set of strings without enumerating them explicitly. Table 1 lists the common features of regular expression patterns used in packet payload scanning. For example, consider a regular expression from the Linux L7-filter for detecting Yahoo traffic: “^(ymsg|ypns|yhoo).?.?.?.?.?.?.?[lwt].*\xc0\x80”. This pattern matches any packet payload that starts with *ymsg*, *ypns*, or *yhoo*, followed by seven or fewer arbitrary characters, and then a letter *l*, *w* or *t*, and some arbitrary characters, and finally the ASCII letters *c0* and *80* in the hexadecimal form [9].

Syntax	**Meaning**	**Example**
^	Pattern to be matched at the start of the input	^AB means the input starts with AB. A pattern without ‘^’, e.g., AB, can be matched anywhere in the input.
I	OR relationship	A|B denotes A or B
.	A single character wildcard	
?	A quantifier denoting one or less	A? denotes A, or an empty sting.
*****	A quantifier denoting zero or more	A* means an arbitrary number of As.
{}	Repeat	A{100} denotes 100 As.
[]	A class of characters	[lwt] denotes a letter l, w, or t.
[^]	Anything but not n	[^\n] denotes any character except \n.

Table 1: Features of Regular Expressions [9].

5 Limitation of Existing Techniques

From the survey of Dpi techniques we found that, the approach describe in these techniques may require a large number of transitions for some cases, leading to an increase in the number of memory accesses per input byte. In addition, DFA construction is complex and requires significant resources [1]. There is very few network intrusion detection techniques discover in wireless networks.

CompactDFA technique used in architecture requires several TCAM working in parallel, Due to its small memory and power requirements. NBA technologies have some significant limitations. They are delayed in detecting attacks because of their data sources, especially when they rely on flow data from routers and other network devices [3]. DFA/EC does not combine with the existing transition compression and character-set compression techniques, and perform experiments with more rule-sets [4]. One of the problems for StriFA is how to choose an appropriate tag. Since in both the rules and the incoming traffic, the occurrence probabilities of different characters vary from each other, it is a problem to choose an appropriate tag from the rule set [5].

Following table shows comparison of existing network intrusion detection techniques.

Intrusion Detection Techniques	**Throughput**
LaFA	34 Gb/s
CompactDFA	10 Gb/s
Small TCAM	18.6 Gb/s
StriDFA	26.5 Gb/s

Table 2: Comparison of deep packet intrusion techniques

The Deterministic Finite Automata (DFA) consists of a finite set of input symbols (which are denoted as P), a finite set of states, and a transition function to move from one state to the other denoted as @. In contrast of NFA, DFA has only one active state at any given time [4] [9].

The regular expression is required as a need for packet payload inspection to different protocols packets. It introduces a limited DPI system to deal with all packets structures. As the result of this limitation, state-of-art systems have been introduced to replace the string sets of intrusion signature with more expressiveness regular expression (regexp) systems. Therefore, there are several content inspection engines which have partially or fully migrated to regexps including those in Snort [11], Bro [10], and Cisco systems’. However, using the regexp to represent patterns includes converting this regexp to Deterministic Finite Automata (DFA). This DFA is represented in the DPI systems as table. This table represents the states and transitions of DFA as records which mean that the expansion of memory table of DFA of regexp depends on the size of DFA [9].

Experimentally, DFA of regexp that contains hundreds of pattern yields to tens of thousands of states which mean memory consumptions in hundreds of megabytes. As a solution of one of the common problems of HW based DPI solutions is the memory access because the memory accesses for the contents of the off chip memory are proportional with the number of bytes in the packet [9].

6. WIRELESS AD-HOC NETWORK

An ad-hoc network is a local area network (LAN) that is built spontaneously as devices connect. Instead of relying on a base station to coordinate the flow of messages to each node in the network, the individual network nodes forward packets to and from each other. In the Windows operating system, ad-hoc is a communication mode (setting) that allows computers to directly communicate with each other without a router.

Opposed to infrastructure wireless networks, where each user directly communicates with an access point or base station, a

Mobile ad hoc network, or MANET, does not rely on a fixed infrastructure for its operation shown in figure. The network is an autonomous transitory association of mobile nodes that communicate with each other over wireless links. Nodes that lie within each other's send range can communicate directly and are responsible for dynamically discovering each other. In order to enable communication between nodes that are not directly within each other's send range, intermediate nodes act as routers that relay packets generated by other nodes to their destination.

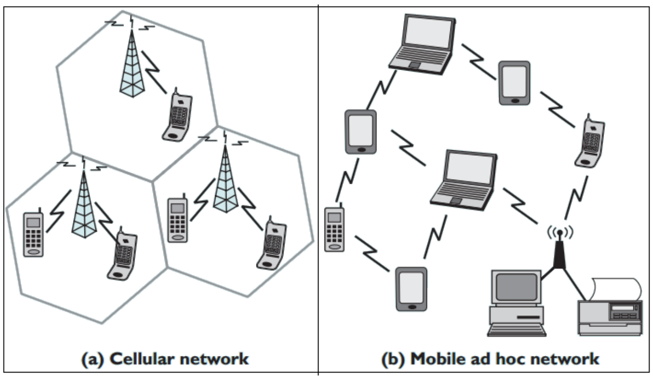

Figure 4: Adhoc Vs Cellular wireless network

The rapid evolution in the field of mobile computing is driving a new alternative way for mobile communication, in which mobile devices form a self-creating, self-organizing and self-administering wireless network, called a mobile ad hoc network. Its intrinsic flexibility, lack of infrastructure, ease of deployment, auto-configuration, low cost and potential applications makes it an essential part of future pervasive computing environments.

7. DOS ATTACK

A Distributed Denial of Service (DDoS) attack is an attempt to make an online service unavailable by overwhelming it with traffic from multiple sources. They target a wide variety of important resources, from banks to news websites, and present a major challenge to making sure people can publish and access important information.

Attackers build networks of infected computers, known as 'botnets', by spreading malicious software through emails, websites and social media. Once infected, these machines can be controlled remotely, without their owners' knowledge, and used like an army to launch an attack against any target. Some botnets are millions of machines strong.

Botnets can generate huge floods of traffic to overwhelm a target. These floods can be generated in multiple ways, such as sending more connection requests than a server can handle, or having computers send the victim huge amounts of random data to use up the target's bandwidth. Some attacks are so big they can max out a country's international cable capacity.

Four common categories of attacks:

TCP Connection Attacks - *Occupying connections*

This attack attempts to use up all the available connections to infrastructure devices such as load-balancers, firewalls and application servers. Even devices capable of maintaining state on millions of connections can be taken down by these attacks.

Volumetric Attacks - *Using up bandwidth*

This attack attempts to consume the bandwidth either within the target network/service, or between the target network/service and the rest of the Internet. These attacks are simply about causing congestion.

Fragmentation Attacks - *Pieces of packets*

These send a flood of TCP or UDP fragments to a victim, overwhelming the victim's ability to re-assemble the streams and severely reducing performance.

Application Attacks - *Targeting applications*

These attempt to overwhelm a specific aspect of an application or service and can be effective even with very few attacking machines generating a low traffic rate (making them difficult to detect and mitigate).

7.1 Difference between DOS and DDOS Attack

It is important to differentiate between Denial of Service (DoS) and Distributed Denial of Service (DDoS) attacks. In a DoS attack, one computer and one internet connection is used to flood a server with packets, with the aim of overloading the targeted server's bandwidth and resources.

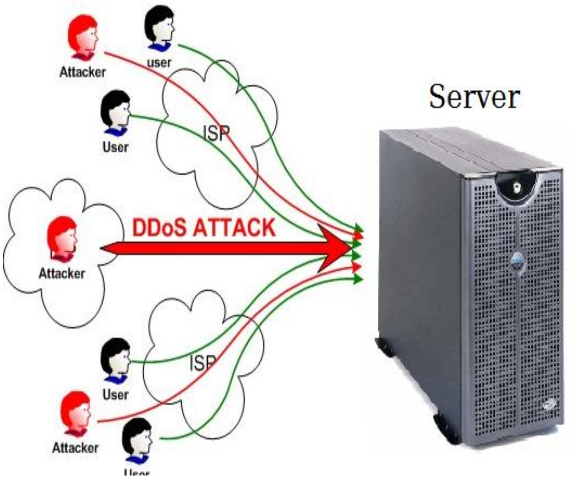

Figure 5 DoS and DDoS Attack

DDoS attack, uses many devices and multiple Internet connections, often distributed globally into what is referred to as a botnet. A DDoS attack is, therefore, much harder to deflect, simply because there is no single attacker to defend from, as the targeted resource will be flooded with requests from many hundreds and thousands of multiple sources.

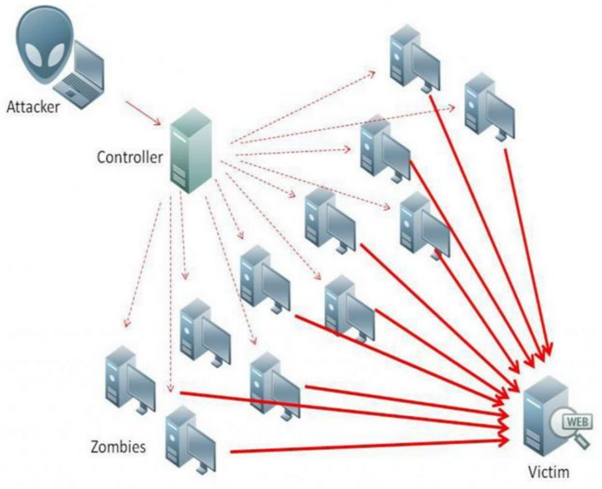

Figure 6: DDOS attack

8. PROPOSED METHODOLOGY

In most of the applications Regular expressions (RegExes) are used to flexibly represent complex string patterns in many applications, such as network intrusion detection and prevention systems (NIDPSs), Compilers and DNA multiple sequence alignment [1]. Intrusion detection is the process of monitoring the events occurring in a computer system or network and analyzing them for signs of possible incidents, which are imminent threats of violation of computer security policies and standard security practices. Intrusion detection and prevention systems (IDPS) are primarily focused on identifying possible incidents in packet, logging information about these incidents, attempting to stop and reporting them to security administrators [3].

Deep Packet Inspection (DPI) is a technology that enables the network owner to analyze internet traffic, throughout the network, in real-time and to differentiate them according to their payload [3]. Traditional packet inspection algorithms have been limited to comparing packets to a set of strings. Newer DPI systems, such as Snort [11], and Bro [10], use rule-sets consisting of regular expressions, these systems are more expressive, efficient, and compact in specifying attack signatures [4].

8.1 Major Functions

- Improvement in LaFA technique.
- Increase in throughput of network intrusion detection.
- Increase in complex RegExp detection speed.
- Minimize memory and resource requirements.
- Implement this technique for wireless network intrusion detection.

8.2 Proposed system Dataflow

Following figure 7 shows dataflow diagram for our proposed system.

When client send request on network, before reaching server it accepted by DOS filter. Dos filter validate the request packets for DOS attack packets, if it found dos contaminated the it will block the packet and terminated shown in figure. When client send request on network, before reaching server it accepted by DOS filter.

If pocket does not contain DOS attack pattern it forwarded to IDPS module. The IPDS used regular expression detection technique in DPI for further intrusion analysis. In this analysis it will search for what type of malicious code used and how it will harm to server analyzed.

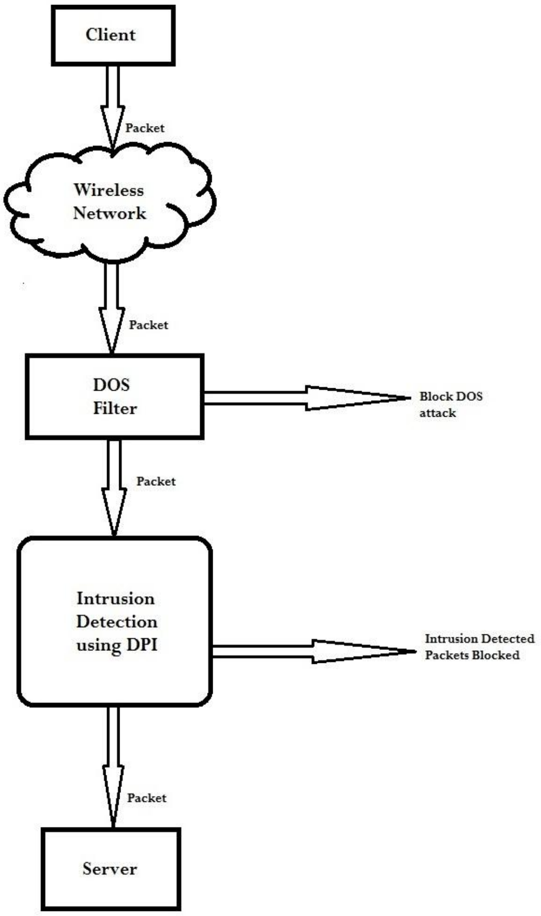

Figure 7: Dataflow diagram for proposed system

If packet matched the regexp then packet is blocked, else it passed to the server.

8.3 DOS filter Dataflow

Dos filter dataflow diagram shoe in following figure.

In DOS filter when packet is received it first calculated the number of request come from each IP per second (RPS). RPS compared with the minimum request to be come from any client (X). X is the minimum request can be sent by any human client.

If RPS is greater than X then packet rejected as it DOS attack packet. Else the packet is forwarded to IDPS module. In IDPS module the packet searches for other intrusion or, malicious code in it.

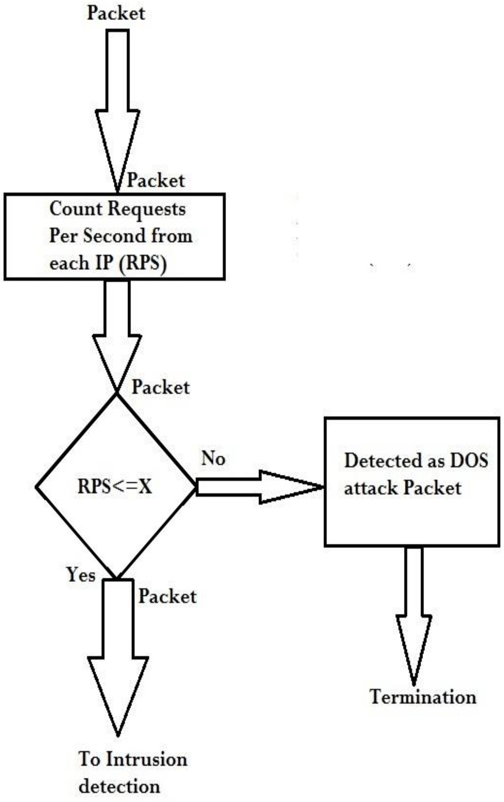
Packet
Packet
Count Requests Per Second from each IP (RPS)
Packet
RPS<=X
No
Detected as DOS attack Packet
Yes
Packet
To Intrusion detection
Termination

Figure 8: Dataflow diagram for DOS filter

8.3.1 Working of Dos Filter

When any user connects to the server it start to counting request from one ip per second if it more than limit of request can be made by human user, then the packets from this ip is blocked. In this methodology the attacker packets containing malicious code can be blocked and protect the server.

If the attacker uses the system that changes ip of the client continuously then the proposed methodology scan the packets same structure and blocked it.

9 Advantages of Proposed System

With this proposed methodology improve network intrusion detection throughput with use of DPI techniques. The Dos filter remove or blocked all dos attack malicious packet it filters the dos attack packets due to which it improve malicious packet detection.

A LaFA technique is used for effective detection of evaluating RegExp on the network. Higher throughput in network intrusion detection can be possible with the use of proposed methodology. Above discuss techniques could have better performance in memory requirements, speed of detection, detection of evaluating RegEx detection. There is very few intrusion detection techniques work on wireless network.

10 Result and Comparison

Table 3 shows the comparison of above deep packet intrusion techniques with respect to the maximum throughput.

Intrusion Detection Techniques	**Throughput**
LaFA	34 Gb/s
CompactDFA	10 Gb/s
Small TCAM	18.6 Gb/s
StriDFA	26.5 Gb/s

Table 3: Comparison of deep packet intrusion detection techniques

11. Future work and Conclusion

In our proposed methodology we work on DOS filter to improve speed of network intrusion detection. We are working on DPI technique improvement for further intrusion detection. We conclude that network intrusion techniques discuss in this paper can be improved for better results and performance. In future we try to improve DPI techniques in wireless networks. We try to improve limitations on existing DPI techniques discuss in section 5 and try to reduce memory consumption required for regular expression matching. Furthermore we will implement or improve Dpi technique which will work on the wireless networks for RegExp detection.

References

[1] Masanori Bando, N. Sertac Artan, and H. Jonathan Chao., "*Scalable Lookahead Regular Expression Detection System for Deep Packet Inspection*", IEEE Transactions on Networking, Vol. 20, No. 3, June 2012.

[2] Chad R. Meiners, Jignesh Patel, Eric Norige, Alex X. Liu, and Eric Torng., "*Fast Regular Expression Matching Using Small TCAM*", IEEE/Acm Transactions On Networking, Vol. 22, No. 1, February 2014.

[3] Tiwari Nitin, Solanki Rajdeep Singh and Pandya Gajaraj Singh, "*Intrusion Detection and Prevention System (IDPS) Technology- Network Behavior Analysis System (NBAS)*", ISCA Journal of Engineering Sciences, Vol. 1(1), 51-56, July 2012.

[4] Cong Liu, Yan Pan, Ai Chen, and Jie Wu., "*A DFA with Extended Character-Set for Fast Deep Packet Inspection*", IEEE Transactions On Computers, Vol. 63, No. 8, August 2014.

[5] Xiaofei Wang, Yang Xu, Junchen Jiang, Olga Ormond, Bin Liu, and Xiaojun Wang, "*StriFA: Stride Finite Automata for High-Speed Regular Expression Matching in Network Intrusion Detection Systems*", IEEE Systems Journal, Vol. 7, No. 3, September 2013.

[6] Anat Bremler-Barr, DavidHay, and Yaron Koral, "CompactDFA: Scalable Pattern Matching Using Longest Prefix Match Solutions", IEEE/Acm Transactions On Networking, Vol. 22, No. 2, April 2014.

[7] Tamer AbuHmed, Abedelaziz Mohaisen, and DaeHun Nyang., "*A Survey on Deep Packet Inspection for Intrusion Detection Systems*", Information Security Research Laboratory, Inha University, Incheon 402-751, Korea, March 2008.

[8] Klaus Mochalski, and Hendrik Schulze,"*White paper on Deep Packet Inspection*", ITU-T study groups com13.

[9] Fang Yu, Zhifeng Chen, Yanlei Diao, T. V. Lakshman, and Randy H. Katz, "*Fast and Memory-Efficient Regular Expression Matching for Deep Packet Inspection*", ACM 580-0/06/0012, December 3–5, 2006.

[10] Bing Chen, Lee, J., and Wu, A.S., "*Active event correlation in Bro IDS to detect multi-stage attacks*", Fourth IEEE International Workshop on Information Assurance, 13-14 April 2006.

[11] Rafeeq Ur Rehman, "*Intrusion Detection Systems with Snor*t", ISBN 0-13-140733-3, Library of Congress Cataloging-in-Publication Data, Prentice Hall PTR Upper Saddle River, New Jersey 07458.