International Journal of Scientific & Engineering Research, Volume 4, Issue 8, August 2013
ISSN 2229-5518

Pattern Matching Mechanisms for Packet Inspection

[1] Nikhila T. Suresh, [2] Mariam Varghese
[1] PG Scholar, Rajagiri School of Engineering and Technology,Kochi,India
[2] Assistant Professor,Department of Information Technology,Rajagiri School of Engineering and Technology,Kochi,India
[1] tonikhila@gmail.com, [2]mariamv@rajagiritech.ac.in

Abstract— Packet is an integral part of any network system. Packet holds information & also they are the smallest individual units of information passed through the network. As the attacks in the internet have taken a tremendous growth in the previous years, the protection mechanisms come to the play dramatically. Initially there were protection & authentication mechanisms like password authentication, cryptography, firewalls etc. Along with these all techniques, the introduction of IDS (Intrusion Detection Systems)&Intrusion protection Systems(IPS)also happened. Deep packet inspection is very critical processes in many networking applications. This inspects the packets flowing through the network to detect viruses, applications or intrusions. If it detects any intrusion, then those packets will be blocked. This is known as deep packet inspection. These inspection points are embedded with a large no. of pattern which can be expressed in terms of ordinary strings or regular expressions. Regular expressions are powerful than ordinary strings & can be effectively find the intrusion attempts. Packet payloads are matched against these patterns. I.e. they allow the packet to continue its flow only if the payload of that packet matches with the patterns. This is called deep packet inspection. The major application areas are Intrusion detection, virus scanning, content filtering, instant-messenger management, and peer-to-peer identification etc. The deep packet inspection can be done through string matching or regular expression matching. The string matching can be done using 3 different methods i.e. Automation based, Heuristic based &filter based string matching algorithms. This paper mainly focuses on the statistical study on the different string matching mechanisms present in this scenario for deep packet inspection.

Index Terms— content filtering, deep packet inspection, Intrusion detection, instant-messenger management,virus scanning, packet payload, regular expressions

1 INTRODUCTION

String matching has taken a rapid turn because of the deep packet inspection in different applications. In the case of deep packet inspection, String matching is one of the flavors among a lot. Ordinary string matching algorithms are not useful for packet inspection because it has find applications in intrusion detection, virus scanning, content filtering, instant-messenger management, and peer-to-peer identification. For ensuring security &efficiency it should require proper& powerful mechanisms for trapping any kind of attacks, intrusion attempts, content filtering & application/user identification. Deep packet inspection, as the name implies it deals with the packet which carries information in its payload as shown in Fig 1.1.

Rather than payload, inspection points also process the packet header too. E.g.-from the packet header inspection, we can extract the protocol identifier which enables us for protocol discovery. The packet payload is matched against particular built-in / well defined patterns of the string provided by the network administrator for finding a match. Here patterns represent the possible/allowed combinations of string that may be ruled by a grammar or organization rules or security policies of the network intrusion systems. Finding a match as much as fast considered as a remarking factor in today's world. These mechanisms also have to consider not only the speed but also the efficiency of algorithm, hardware parameters & utilization of memory. There are different variants of methods present for the deep packet inspection. This dissertation mainly discuss about the various kinds of string matching procedures in the networking area that deals with the packet oriented inspection concept.

Initially there were protection & authentication mechanisms like password authentication, cryptography, firewalls etc.

Along with these all techniques, the introduction of IDS (Intrusion Detection Systems)&Intrusion protection Systems(IPS)also happened for handling the various kinds of attack like viruses ,worms ,banned contents , spam etc. Here IDS deals with the analysis for security in forensic application tools where as IPS handles the protection policies for security in an automatical manner .Deep packet Inspection can be implemented in many of internet devices like proxy servers, packet filters, packet sniffers etc. present in the traditional application layer of ISO/OSI model .In another view, DPI can be regarded as the service model of networking system. It can be taken as a combination of IDS&IPS systems.

The deep packet inspection mainly deals with the string

- *Nikhila T. Suresh is currently pursuing masters degree program in network engineering in Mahatma Gandhi University,Kerala, India, PH-+919946675300. E-mail: tonikhila@gmail.com*
- *Mariam Varghese is currently working as assistant professor in Rajagiri School of Engg and Technology,Kerala, India, PH-+919496805528. E-mail: mariamv@rajagiritech.ac.in*

matching algorithms.

Initially deep packet inspection dealt with some of the challenges as well as the design issues [1]. The challenges categorized as: Matching algorithm complexity, multiple packet inspection, increasing population of signatures, Unknown location of signatures, Encrypted data in the packet etc. Primarily, the main objectives of the DPI regarded as : Deterministic behavior, dynamic update of signatures, Efficient memory utilization, scalability , signature support & additional functions.

Fig 1.1 Packet structure

We can also optimize the deep packet inspection process in different ways [2]. Complete Content Protection (CCP) & Fortigate Antivirus Firewall etc are some of them [2]. The DPI classifier associates TCP/UDP session to its corresponding signatures. After occurring a match with the patterns provided by the network system administrator, the session ID is inserted nto a session table and all the remaining packets having same session ID won't be delivered to DPI classifier .The DPI classifier can perform two types of inspections i.e. packet based , per-flow state (PBFS) & message based, per-flow state (MBFS).PBFS analyses data as packet by packet where as MBPS analyses the messages, the stream of data PBFS can also be used for DPI optimization.In this paper,we focus on string matching algorithms applicable to packet inspection process the remainder of the paper is organized as follows.In section II,related works about string matching are discussed.Section III accurately describes methods,Section IV presents the Statistical study on the previously specified techniques of pattern matching , and section V suggests a conclusion for the statistical study performed.Our goal is to find optimal technique that statistically quantifying the benefits of deep packet inspection.

2 Related works

The types of packet inspection are stateful packet inspection method & deep packet inspection .The former stateful packet inspection method process only the packet header as shown in Fig 2.1.If the payload contains any spam/virus/worm, it doesn't care about that and would be undetected. But in the case of DPI, both the header as well as the payload would be processed to find an attack hidden in the packet. Some attacks can transmit multiple packets for the same activity &those may have a weight in terms of thousands of kilobytes . So these packets should be fragmented into segments due to the restriction on packet size in transmission. DPI can also have the capability to reassemble all these packets& can find the attack distributed over a no. of packets.

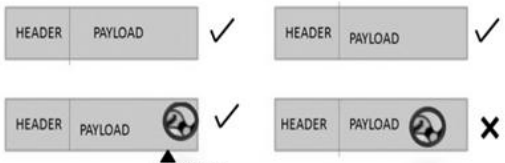


Fig 2. 1 Stateful inspection Vs DPI

There are a variety of approaches to string matching for DPI. Some of the major areas are Automaton-based, Heuristic-based and Filtering-based. The olden pattern matching process make use of the ordinary strings but as the network attacks grows along with the internet growth, it is made insufficient for tracking intrusions, traffic monitoring, detecting spam, viruses etc. Therefore the technology went for higher options i.e, regular expressions. The traditional methods for matching regular expressions are DFA & NFA. But both are not practical in the current situation in terms of memory & time. They are obsolete mechanisms in the field and are updated with its higher versions like D^2FA, δFA, $\delta^N FA$ etc.

2.1 Traditional methods: DFA & NFA

DFA is Deterministic Finite Automata which has only one transition for each individual input symbol specified in the set of input symbols. The tuple representation can be written as: A DFA in a 5-tuple –M= (Θ, Σ, δ, q_0, F) where M represents the machine, Θ represents the infinite set of states in the DFA, Σ represents the set of input symbols, δ represents the automata function which converts the input state to another state, q_0 indicates the initial state, & finally F holds the set of final states specified. The restriction factor behind the DFA is that it requires excessive amount of memory for the regex (Regular Expression) sets. While on the other side, it performs fast string matching.

NFA is Nondeterministic Finite Automata which has one or more transitions for each input symbol specified in the set of input symbols. The tuple representation can be written as: same as that of DFA and M = (Θ, Σ, δ, q_0, F) where Θ represents the infinite set of states in the NFA, Σ represents the set of input symbols, δ represents the automata function which converts the input state to another state, q_0 indicates the initial state, & finally F holds the set of final states specified. The con behind the NFA is that it it performs slow string matching process & requires more memory references for single character.

2.2 Aho-Corasik Algorithm [1]

Aho-Corasick Algorithm is one of the algorithm for the multiple string matching.

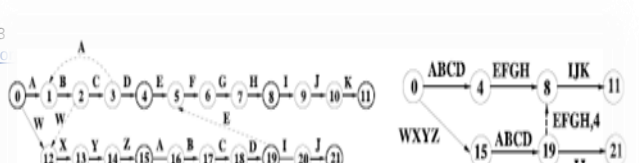
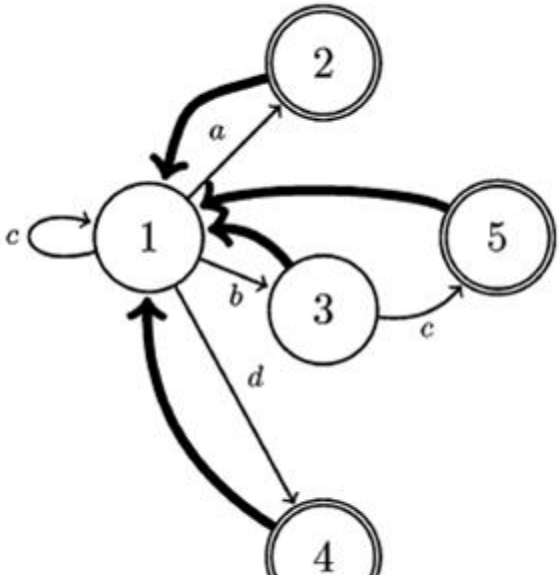


Fig 2.2 Compressed Ac For high speed DPI [1]



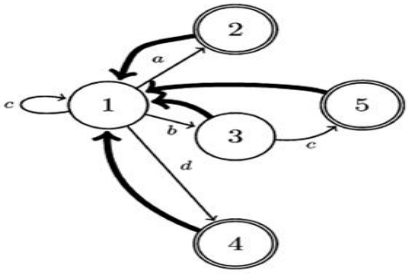


Fig 2. 3 D²FA

Here the transition set for the state 1 is fully defined. All the other states which deals with a transition about a non specified character should take the default transition. Exception to the above statement is that input character C on the state 3.state 3 defined an explicit transition about c because it defines a new transition compared to other states for the same input character. Thus it keeps the differences among state transitions.

In this point of view, this concept reduces the memory conception of DFA simultaneously it reduces the effectiveness of transitions. The effectiveness in the sense that it requires one additional node traversal.

3 TECHNIQUES

The techniques used for the string matching can be broadly classified into are Automaton-based, Heuristic-based and Filtering-based.

3.1 Automaton-based

The related work section described some of techniques related to automation –based approach. This section comprises of the software (BNF in SNORT)and hardware oriented(FPGA) methods.

3.1.1 SNORT

SNORT is an example of Intrusion Detection Systems [IDS] which perform light weight intrusion detection activities [1]. It is software oriented approach & is not cost effective also because it is General Public License [GNU89]. SNORT is based on set of rules. The design principles behind Snort's technology were performance, simplicity, and flexibility. The three primary sub systems inside Snort: the packet decoder, the Detection engine, the logging and alerting subsystem. Snort's detection engine performs the operation based on rule represented in terms of chain headers & options in two-dimensional linked list format. The library file inside the SNORT holds all the IP addresses for fast processing. The individual parts are placed inside the chain option parameter & common property parameters are placed inside of the chain headers. The decoding section provides support Ethernet, SLIP, and raw (PPP) datalink protocols [1]. By focusing on the rules, it is easy to write as well as it is very strong one. Here there are 3 types of rules importantly they are: pass, log, or alert The pass rule just drop out all the packets but log writes the full packet which was selected by the user while he is in the run time & alert gives notifications regarding the security details. As it can get a view about the packet content that have captured , it is very flexible to write rules for them.

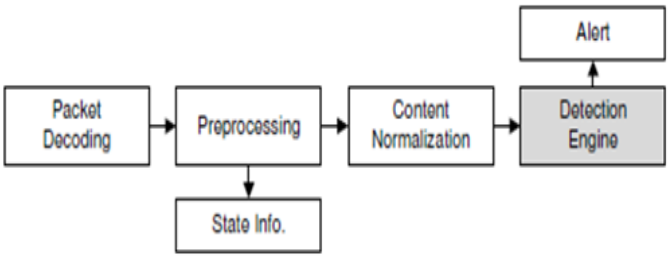


Fig 2.4 Snort Process Stages [1]

By summarizing [4] it is one of the cost effective system that can be applied for the commercial area

3.1.2 FPGA

It is one the hardware implementation on the system. The main highlight is the reconfigurability of the mechanism. The FPGA can be used for creating the logics for NFA. The main aim of the task is to minimize the mapping time between the states. The reduction in the mapping time directly points to

the configuration bits that would easiness the work. Although performing the above is difficult due to 2 reasons .First of all the FPGA bit stream formats are not commercialized second is the error occurred during the process may harm the entire hardware i.e. IC chip.

This paper [5] mainly analyses the implementation of FPGA on xlinix vertex architecture & it is found that NFA implementation is quite simple but the mapping time have some technical problems.

3.1.3 Heuristic Based

The heuristic based mechanism mainly details with the Wu-Manber algorithm which includes the shift operation[]. The shifting occurs with the help of shifting tables. The shifting table includes the shifting values for the bytes included in the packet If the shift value is set to zero then the pattern tries to a find a match with all the others that having same no of bytes. This process takes unlimited time.

3.1.4 Heuristic Based

The filter mechanism can be done in two ways i.e. multi-pattern search algorithm & single pattern algorithm. In single pattern, the substring is extracted from the regular expressions that require match. In Multiple patterns, possible suffixes are generated and searched [7]

4STATISTICAL ANALYSIS

Here we conduct the statistical study on some of the existing systems using some set of platforms [1]& the results are shown in fig 4.1

Fig 4.1 statistical study on different architectures

Algorithm / Component	Implementation Device	Throughput (Gbps)
Parallel Bloom Filters	FPGA XCV2000E	2.46
Aho-Corasick	FPGA	12.35
TCAM/FPGA	Xilinx Virtex2	10
FPGA	Virtex-4	10
NFA/(FPGA and IXP)	Xilinx Virtex2-6000&IXP 2400	1
Hash Function	Xilinx Vertex-II Pro XC2VP70	2
Hash Function and CRC	Xilinx Vertex2 2.712	4.560

5 CONCLUSIONS

The main aim of this paper is to conduct a statistical study on string matching mechanism for deep packet inspection. Here we covered the need for DPI, challenges etc. By considering all the techniques automata is the best one. From this point of view, Automata based techniques have some research topics still to be expanded & a lot of ongoing works are there for strengthening it. The statistical survey conducts on some of the discussed techniques. The analysis of the result indicates that, a lot works have to be done for satisfying the memory & time constraints specified in the challenges.

REFERENCES

[1] Tamer AbuHmed1, Abedelaziz Mohaisen2_, and DaeHun Nyang1 Tamer AbuHmed1, Abedelaziz Mohaisen2 and DaeHun, "*A Survey on Deep Packet Inspection for Intrusion Detection Systems*" Information Security Research Laboratory, Inha University1, Incheon 402-751, Korea 2Electronics and Telecommunication Research Institute, Daejeon 305-700, Korea.

[2] Niccol_o Cascarano _ Luigi Ciminiera_ Fulvio Risso,, "*Optimizing Deep Packet Inspection for High-Speed TrackAnalysis*" Anat Bremler-Barr IDC Herzliya, Israel Yotam Harcho, Israel Yotam Harchol⋆ , David Hay Hebrew University, Israel OWASP Israel 2011.

[3] Martin Roesch, Proceedings of LISA '99: 13th Systems Administration Conference Seattle, Washington, USA," *Space-Time Tradeoffs in Software-Based Deep Packet Inspection*" November 7–12, 1999.

[4] Roesch Proceedings of LISA '99: "*Snort — Light Weight Intrusion Detection for networks*", Martin 13th Systems Administration Conference Seattle, Washington, USA, November 7–12, 1999.

[5] Reetinder Sidhu sidhu@halcyon.usc.edu, Viktor K. Prasanna prasanna@ganges.usc.edu Department of EE-Systems, "*Fast Regular Expression Matching using FPGAs*" University of Southern Califonia ,Los Angles CA 90089.

[6] Kai Zheng,Member IEEE,Hongbin Lu, student member,IEEE, and Erich Nahum Mathieu Raffinot CNRS ,"*Scalable Pattern matching on Multicore Platform via Dynamic Differentiated Distributed Detection(D4)*",Poncelet Laboratory, Independent University of Moscow.

[7] Mathieu Raffinot CNRS, Poncelet Laboratory" *High-Level Comparative Genomics Habilitation dissertation* ",Mathieu Raffinot CNRS, Independent University of Moscow, 11 street Bolcho¨ı Vlassievski, 119 002 Moscow, Russia, mathieu@raffinot.netifting

[8] Sang Kil Cha, Iulian Moraru, Jiyong Jang, John Truelove, David Brumley, David Andersen Carnegie Mellon University, Pittsburgh, PA," *SplitScreen: Enabling Efficient, Distributed Malware Detection*", Carnegie Mellon University, Pittsburgh, PA.

[9] Domenico Ficara, Member, IEEE, Andrea Di Pietro, Student Member, IEEE, Stefano Giordano, Senior Member, IEEE, Gregorio Procissi, Member, IEEE, Fabio Vitucci, Member, IEEE, and Gianni Antichi, Member, IEEE "*Differential Encoding of DFAs for Fast Regular Expression Matching*". IEEE/ACM TRANSACTIONS ON NETWORKING, VOL. 19, NO. 3, JUNE 2011.