Deep Packet Analyzer by using Proposed Aho-Corasick Algorithm

Sushant Nalawade[1] Reshma Walunj[2] Chandan Bhatte[3] Prof. D. R. Ingle[4]
[1,2,3,4]B. E. Student
[1, 2, 3, 4] Computer Engineering
[1,2,3,4]Bharati Vidyapeeth College of Engineering

Abstract---Deep Packet Analyzer (DPA) lies at the core of Intrusion Detection/Prevention Systems. This type of packet analyzer aims to identify various malware (including spam and viruses) by inspecting both the header and the payload of each packet and comparing it to a known set of patterns. This known set of pattern is stored in the signature database. Despite extensive research effort, ordinary anomaly detection systems still suffer from serious drawbacks such as high false alarm rates due to the enormous variety of network traffic. The intrusion detection system (IDS) proposed in this paper is operates on network flows rather than on entire network packets. Incoming traffic is analyzed using Aho-Corasick Algorithm and comparing with signature database. In short, the proposed IDS receive network traffic and analyze them with a Aho-Corasick Algorithm.

Keywords: Sniffer, Detection, Intrusion, Prevention, Signature, Aho-corasick algorithm.

I. INTRODUCTION

In today's world of network technology security plays a very important part and providing security to that technology is becoming a necessity for an organization. Most of the organization's work is dependent on the internet i.e.to communicate with the people to provide them news, online shopping, email facilities, credit card details and many more. In this fast growing, rapidly developing technology and widespread use of the Internet, a lot of problems have been faced to secure the system's critical information within or across the networks. There are many people attempting to attack on systems to extract critical information. A huge number of attacks have been observed in the last few years. Intrusion Detection and Prevention Systems (IDPS) play a vital role against those attacks by protecting the system's critical information. As firewalls and anti viruses are not enough to provide full protection to the system, organizations have to implement the IDPS to protect their critical information against various types of attacks. We have made in a prototype using Aho-Corasick Algorithm which not only helps in detecting the malicious packet but also helps in detecting it at a faster rate as the time complexity of the algorithm is very low for large amount of data. This paper is organized as follow. Section II introduces what is IDS and terminology. Sections III describe the working of Aho-Corasick Algorithm. Section IV describes Modus-operandi of the IDS. Section V is future Scope and section VI provides a conclusion.

II. INTRUSION DETECTION SYSTEMS (IDS)

Intrusion means to interrupt someone without permission. Intrusion is an act of using computer system resources without any rights and permission, which causes incidental damage. Intrusion Detection means any mechanism which detects the intrusive behavior. Intrusion Detection System (IDS) monitors network traffic and its suspicious behavior against security. If it detects any threat then it alerts the system or network administrator. The objective of IDS is to detect and inform about intrusions. IDS is a set of techniques and methods that are used to detect suspicious activities both at the network and host level. There are two main types of Intrusion Detection System, Host Based Intrusion Detection Systems (HIDS) and Network Based Intrusion Detection Systems (NIDS).

A. Intrusion Prevention System (IPS)

IPS is an advance combination of IDS, personal firewalls and anti-viruses etc. The purpose of an Intrusion Prevention System (IPS) is not only to detect an attack that is trying to interrupt, but also to stop it by responding automatically. The response includes logging off the user, shutting down the system, stopping the process and disabling the connection etc. Similar to IDS, IPS can be divided into two types, i.e. Host-Based Intrusion Prevention Systems and Network-Based Intrusion Prevention Systems.

B. Signature

Signature is pattern which identifies the known set of malicious code. Signatures may be present in different parts of a data packet depending upon the nature of the attack. For example, you can find signatures in the IP header, transport layer header (TCP or UDP header) and/or application layer header or payload. IDS depends upon signatures to find out about intruder activity [5].

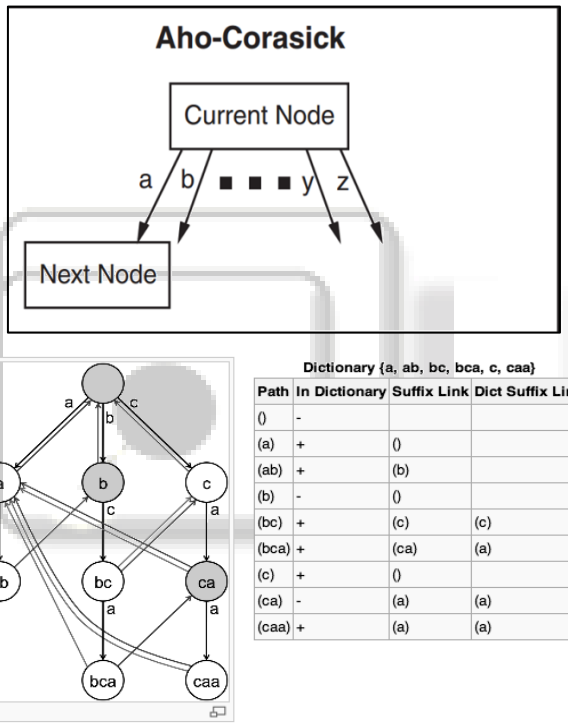
III. AHO-CORASICK ALGORITHM

String pattern matching problem is defined as "to find all occurrences of pattern in text. Earlier String matching algorithms can be single pattern matching or multiple pattern matching. Single pattern string search example is algorithm proposed by Boyer-Moore

1) For intrusion detection system we need multiple pattern string matching algorithm as number of signatures are large. The multiple string matching algorithms unlike the single string matching algorithms merge all the signatures to form a kind of fast searchable single structure which improves performance of algorithm. Example of one such algorithm is Aho-Corasick algorithm
2) This algorithm matches all the strings of given strings simultaneously with the given input text. Algorithm is able to match strings in worst-case time linear in the size of the input
3) The complexities of constructing a pattern matching machine and scanning a text are linear to the total

length of given patterns and the length of a text, respectively

4) Aho-Corasick works by constructing a state machine from the strings to be matched. Algorithm starts with a root node and constructs a pattern tree for all input patterns. The constructed tree structure has end nodes which are markers for signaling a match. When the searching mechanism reaches to such node it's the detection trigger. A pattern matching machine consists of go to function, failure function and output function
5) The search tree is constructed of in the initialization stage of algorithm after initial construction the pattern matching can continue. This is suitable for intrusion detection systems. The mechanism can be explains by following diagram. It shows the formed pattern tree for dictionary (a,ab,bc,bca,c,caa) also shows the generation of failure transitions and final nodes. The failure transitions allow for faster search and eliminate the task of starting the search all over again.



Dictionary {a, ab, bc, bca, c, caa}

Path	In Dictionary	Suffix Link	Dict Suffix Link
()	-		
(a)	+	()	
(ab)	+	(b)	
(b)	-	()	
(bc)	+	(c)	(c)
(bca)	+	(ca)	(a)
(c)	+	()	
(ca)	-	(a)	(a)
(caa)	+	(a)	(a)

Fig. 1:Aho-corasick Automata.

Implementation of Aho-Corasick algorithm can be explained in following steps.

A. Construction of pattern tree

In the first phase of the tree building, signatures are added to the tree. The tree structure is implemented using a data structure. The structure contains a Boolean flag representing whether node is final node also contains pointers to other transitions.

B. Fail function

fail function generate back links which improves the efficiency of algorithm and simulates automaton. Queue structure and breadth first search is used for this.

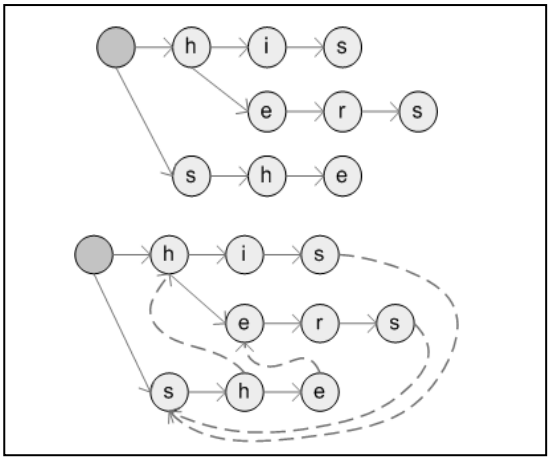


Fig. 2: Aho-Corasick Working.

C. Actual Pattern Matching

Actual pattern matching is simple the matching starts at root and goes transitions as per input characters. Failure to reach to end condition indicates the no match otherwise there is a match.

IV. MODUS OPERANDI

The proposed system operates by first sniffing the incoming packet. These packets's data is then passed to the Aho-Corasick algorithm for analysis. String matching takes place between contains of the packet and the signature database. Corresponding result is generated.

A. Sniffer

It captures each packet as it transmitted over network and decodes the packet's raw data showing the values of various fields in the packet. When the traffic is captured the entire contents of packets can be recorded. After sniffing means capturing the packet it analyzes the packet that is packet source port, destination, destination port, packet size and protocol etc. The following snapshot displays the packet accordingly in the console description window with all the details.

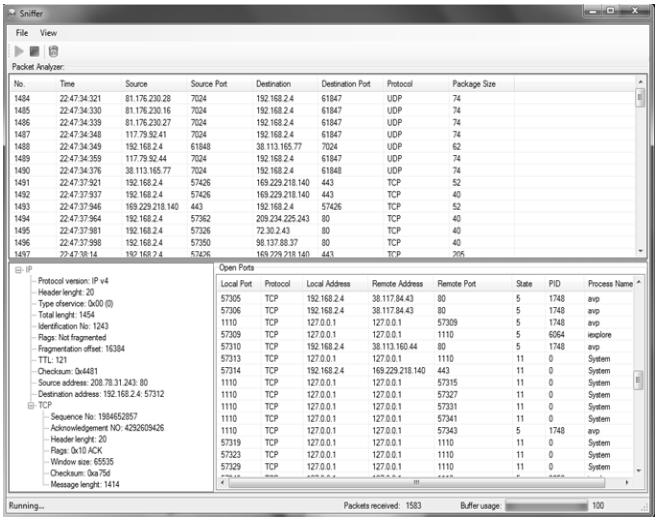


Fig. 3: Packet Analyzer

B. Windows Native API

Packet analyzer is built by implementing windows live API in visual studio 2010. Incoming / outgoing IP packets on the host machine which is running windows operating system can be captured using this API [5].

C. String Matching

In the third part string matching takes place in which incoming flows of the packets from various protocols are sniffed and then given to the pattern matching Aho-Corasick algorithm. This Aho-corasick algorithm compares contains of the packet with signature database bit by bit. Signature database consist of the virus database file. When an IDS detects an intruder, it has to inform security administrator about this using alerts. Alerts may be in the form of pop-up windows, logging to a console, sending e-mail and so on. Alerts are also stored in log files. The log messages are usually saved in file. Log messages can be saved either in text or binary format. The following figure shows architecture of Network Packet Analyzer using Aho-corasick algorithm.

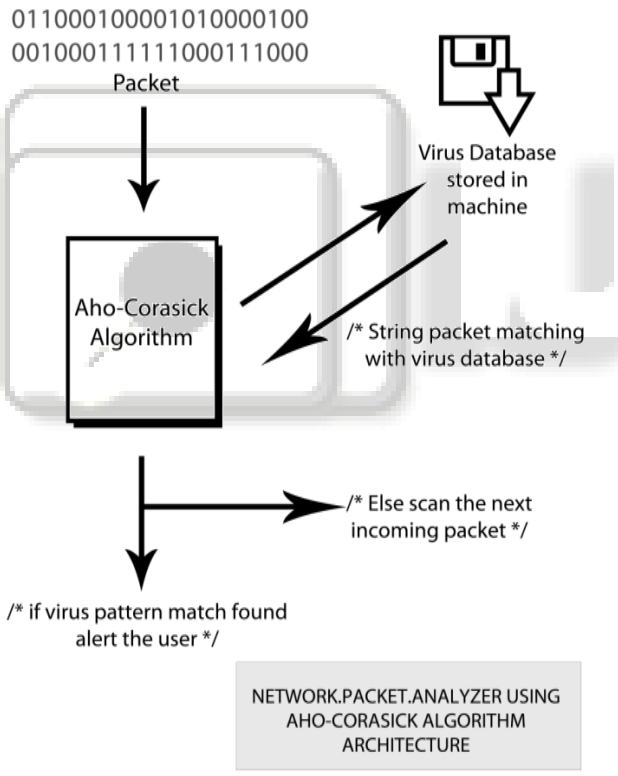


Fig. 4: Network Packet Analyzer Architecture using Aho-corasick Algorithm.

V. FUTURE SCOPE

It's possible to further increase the performance of the Aho-corasick algorithm using techniques discussed in [4].[4] Discusses various compression techniques which reduce the size of the built search tree in main memory which in turn reduces memory footprint of the algorithm. It's also possible to add data mining techniques to make smarter detections. Also we have to keep in mind this is signature based IDS and has its own disadvantages [15]such as 0 day attacks hence this analyzer can be combined with a anomaly based IDS to minimize them.

VI. CONCLUSION

This paper discusses techniques to support the security of an host against threats or attacks. IDPS provides the facility to detect and prevent from attacks by inheriting multiple approaches. Signature based IDS uses various pattern matching algorithms. This paper provides description of the Aho-corasick algorithm used for pattern matching. Aho-Corasick pattern matching algorithm provides better performance in terms of speed, space complexity & algorithm complexity.

REFERENCES

[1] R. S. Boyer and J. S. Moore. A fast string searching algorithm. Comm. ACM, 20(10):762-772,1977.
[2] S. Antonatos, K. G. Anagnostakis, and E. P. Markatos. Generat-ing realistic workloads for network intrusion detection systems.InACM Workshop on Software and Performance, 2004.
[3] Deterministic Memory-Efficient String Matching Algorithms for Intrusion Detection: Nathan TuckTimothy SherwoodBrad Calder,George Varghese IEEE INFOCOM 2004
[4] T. Nishimura, S. Fukamachi, T. Shinohara, "Speed-up of Aho-Corasick Pattern Matching Machines by Rearranging States," spire, pp.0175, Eighth Symposium on String Processing and Information Retrieval (SPIRE'01), 2001
[5] Miao Wang , Cheng Zhang , Jingjing Yu , Native Api Based Windows Anomaly Intrusion Detection Method Using SVM, June 2006.
[6] Jianming Yu and Yibo Xue, Robust Quick String Matching Algorithm for Network Security, Journal of Computer Science & Network Security, pp180~184, Vol.6, No.7B, July 2006.
[7] Zongwei Zhou, Yibo Xue, Junda Liu, Wei Zhang and Jun Li, MDH: A High Speed Multi-phase Dynamic Hash String Matching Algorithm for Large-Scale Pattern Set, ICICS 4861, pp. 201-215, 2007.
[8] S. Wu, U. Manber," A fast algorithm for multi-pattern searching," Tech. R. TR-94-17, Dept. of Comp. Science, Univ of Arizona, 1994.
[9] J.S.Wang, H.K.Kwak, Y.J.Jung, H.U.Kwon, C.G.Kim and K.S.Chung, A Fast and Scalable string matching algorithm using contents correction signature hashing for network IDS, IEICE Electronic Press, vol 5, no 22, 949-953, 2008.
[10] Jianming, Y., Yibo, X., and Jun, L. 2006. Memory efficient string matching algorithm for network Intrusion Management System, In Proceedings of Global Telecommunications Conference, San Francisco, California, USA, pp. 1—5
[11] Ahmed Patel, Qais Qassim, Christopher Wills. A survey of intrusion detection and prevention systems, InformationManagement & Computer Security Journal (2010).
[12] J.van Lunteren, "High-performance pattern-matching for intrusion detection,"in IEEE INFOCOM, 2006.

[13] Khalid Alsubhi_, Yassir Alhazmi†, Nizar Bouabdallah‡, Raouf Boutaba.. Rule Mode Selection in Intrusion Detection and Prevention Systems .
[14] Moses Garuba, Chunmei Liu, and Duane Fraites Intrusion Techniques: Comparative Study of Network Intrusion Detection Systems.
[15] Teng, Henry S., Chen, Kaihu, and Lu, Stephen C-Y, "Adaptive Real-time Anomaly Detection Using Inductively Generated Sequential Patterns," 1990 IEEE Symposium on Security and Privacy.

