ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

A Novel Approach to Attacks Detection for Encrypted Web Applications

Princy P[1], Scaria Alex[2], Ambikadevi Amma [3]

P.G. Student, Department of Computer Engineering, JCET Engineering College, Palakkad, Kerala, India[1]

Assistant Professor, Department of Computer Engineering, JCET Engineering College, Palakkad, Kerala, India[2]

Professor, Department of Computer Engineering, JCET Engineering College, Palakkad, Kerala, India[3]

ABSTRACT: Web-based applications are becoming common; attacks against these applications pose a serious problem. An Intrusion Detection System (IDS) is one way of dealing with such attacks. An Intrusion Detection Systems (IDS) is located beside the web server and monitors the users' activities by protocol analysis and pattern matching. In other words, IDSes reconstruct HTTP headers and payload from captured packets, and identify attacks by comparing traffic to signatures of attacks. Thus the process requires the privilege of watching the entire payload of packets. Because the IDSes inspect the contents of a packet, it is difficult to find attacks by the current IDS. This approach applies encrypted traffic analysis to intrusion detection, which analyses contents of encrypted traffic using only data size and timing without decryption. First, the system extracts information from encrypted traffic, which is a set comprising data size and timing or each web client. Second, the accesses are distinguished based on similarity of the information and access frequencies are calculated. Finally, malicious activities are detected according to rules generated from the frequency. One of the reasons is the increasing use of encrypted communication that strongly limits the detection of malicious activities. To overcome this shortcoming here present a new behavior-based detection architecture that uses similarity measurements to detect intrusions as well as insider activities like data exfiltration in encrypted environments. Similarity based intrusion and extrusion detection show that the system detects various attacks like SQL injection, DOS, Brute force Attacks with a high degree of accuracy.

KEYWORDS: IDS, Encrypted Environment, SQL injection, DOS, Brute force Attacks

I. INTRODUCTION

Web applications are becoming an essential part of everyday lives, with many of different activities of humans now dependent on the functionality and security of these applications. Attacks against web applications constitute a serious problem. Intrusion Detection Systems (IDSes) are one solution, however, these systems do not work effectively when the accesses are encrypted by protocols. Web applications are ubiquitous, perform mission critical tasks, and handle sensitive user data.

World Wide Web has become an ultimate source of information. Traditional services such as banking, education, medicine, defence, and transportation are being presented by web applications. Whenever the users make use of any web application, all the activities of the users get automatically get appended into the web log files. However, one of the main problems with sending data over the internet is the security threat it poses i.e. the personal or confidential data can be stolen or hacked in many ways. Therefore it becomes very important to take data security into consideration, Encryption and thus cryptography was invented to protect communication for various reasons. In particular, the increased desire for privacy (confidentiality) and protection against manipulations of a packet on its way from sender to recipient (integrity) needs to be pointed out. Encryption technology can also be used maliciously.

The web log file data helps the website owners in number of ways such as customization of web content, pre-fetching and caching, E-commerce, etc. However, the web log file data is exposed to number of attacks. In the current trends of the world, the technologies have advanced so much that most of the individuals prefer using the internet as the primary medium to transfer data from one end to another across the world. There are many possible ways to transmit data using the internet: via emails, chats. The data transition is made very simple, fast and accurate using the internet.

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

II. RELATED WORK

A Secured Web shop was generated and the Tsung benchmarking tool was used for the creation and simulation of users and their interactions with the website, for example, searching for products, buying goods, or reading descriptions and blogs. During the interaction of the benign users, brute-force login attempts and SQL injections had been executed against the Web shop. After generating the attacks the Network Probe copies the network packets and generates the necessary structures for further analysis. After that, the data is sent to the modules for Intrusion Detection and Extrusion and Insider Detection. Network Probe [1] the data packets of the network link are copied by the Network Probe, which is running on a transparent bridge/WireTAP. Therefore it can be placed everywhere in the network, typically behind a firewall or near an edge router. The probe is homemade because besides generating flow data, sequences of payload sizes related to the different connections have to be saved temporarily. Also, different approaches of data extraction and hash generation have been examined and routines for performance measurements and debugging are implemented within the probe. The Packet Capture module uses the pcaplibrary to investigate source and destination IP addresses and ports, timestamps, sizes of the payloads, and TCP-flags of all packets. During the analysis the payload sizes and timings are saved as sequences, which can be accessed by hash keys. Different hash values are created for every new data connection (Hash-Generation) based on the combination of IP addresses and ports (for instance, by using the hashing algorithm. Based on these hashes, new packets can be assigned to connections already identified and the statistical data of each connection can be accessed fast and efficiently. After that, the data is sent to the modules for Intrusion Detection and Extrusion and Insider Detection. Fig 1 shows the architecture of intrusion detection system.

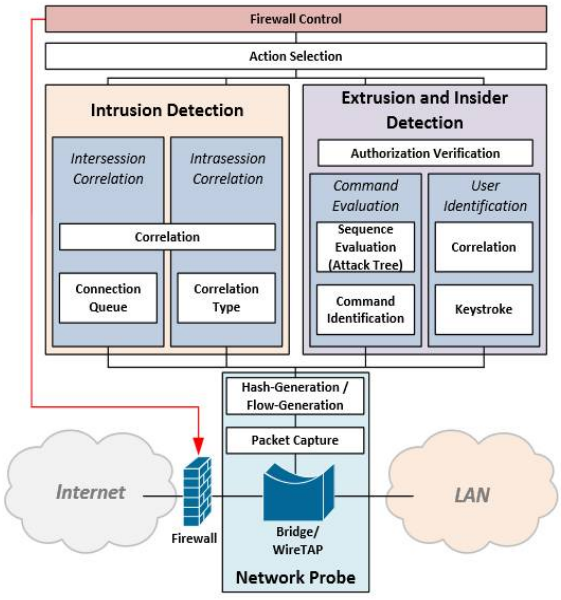

Fig: 1: Architecture of Intrusion Detection System

Intrusion Detection Correlation techniques can be used to determine the degree of similarity of two functions. Different types of correlation measurements exist, for example, for nominal or metric scaled data, numerical data, or ordinal scaled data, and different strategies can be applied. The evaluated similarities can be used for various aspects of intrusion and extrusion detection. For example, a distinction between benign and malicious sessions to a Web

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

server can be achieved because the statistical data of the benign sessions are quite similar to a specific degree, while malicious packet series are more different. For instance, using a search function or looking for products and placing an order have other typical sequences of packet sizes and timings than a SQL injection [2], a dump of the database, or brute-forcing an account. An intersession correlation is as follows. Typically a higher amount of all connections will be benign while only a small portion will be malicious. If every connection is correlated with a minimum number of other connections, the correlation results will be relatively high for the benign connections. However, when a malicious connection is correlated with benign ones, the resulting similarity will be low. Consequently, this technique does not need a configuration or knowledge of the benign data in advance. The knowledge is generated in real-time based on the characteristics of the connections. In addition, effects like flash crowds can be detected and evaluated correctly. An attacker could try to generate a majority of malicious connections to avoid detection. However, this effect can be recognized. During taking over the majority of connections, the calculated correlations of benign connections are dropping. This happens constantly or quite fast, depending on the frequency of new malign connections. By monitoring the tendency and mean values of correlations, this development can be detected, showing a likely attack. The module Intersession Correlation considers multiple connections of different users to a service, for example, a Web shop, and temporarily stores them in the Connection Queue. Next the Correlation module calculates the similarity of the connections based on the payload sizes and packet timings. Intra session Correlation is used if intersession correlation is not feasible. This will be mainly the case if too few parallel connections exist, for example in the case of remote shells like SSH. The system selects the kind of correlation to be used based on the service. For intra session correlations, single sessions consisting of one or a few connections of one user are evaluated. Several correlation constellations are possible, for example, multiple sessions initiated in parallel or only one session used multiple times to brute-force different passwords. This is taken into consideration by the module Correlation Type which analyses the connection and forwards the information needed for the calculation of the similarities to the Correlation module.

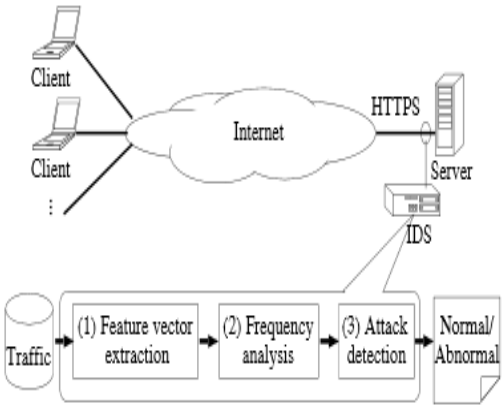

Fig 2: attacks detection

Extrusion and Insider Detection performing extrusion and insider detection by using the concepts of intra- and intersession correlations can sometimes be impossible due to the low number of users. There, the resulting similarity values cannot be clearly separated. To overcome this, other knowledge about characteristics of a service or data connection has to be used to construct. Within the Command Evaluation, first the Command Identification analyses the statistical data of encrypted connections to identify which commands are entered by a user. For this purpose the encrypted network packets are categorized with the use of clusters where a cluster consists of the encrypted packets of one command and the corresponding response of the server. Again, similarity measurements are used to detect the degree of correlation of the data stream with profiles of employees to identify the user of a connection. Therefore we have integrated a module User Identification in the architecture.

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

III. PROPOSED WORK

Encrypted server and clients are created. The brute-force login attempts, DOS attack and sql injections had been executed against the Web site. The brute force attack is created by using hydra tool. It can perform rapid dictionary attacks against web site. Intrusion Detection the Dos and brute force attacks are detected by using information theoretic measure. It uses information based metric Kullback libeler divergence or skews divergence. The information theoretic measures are probability based measures which is applied for any type of data type. Information distance (or divergence) is a measure of the difference between different probability distributions. However, the K-L divergence does not give a probability of similarity between the two distributions but a relative value. So, we have to define which value of K-L divergence indicates an attack. Kullback–Leibler distance metrics for the intrusion detection in terms of early detection, lower false positive rates, and stabilities.

It can also define the Kullback–Leibler distance as

$$D_1(P,Q) = D_1(P\|Q) + D_1(Q\|P)$$

$$D_1(P,Q) = \sum_{i=1}^{n}((p_i - q_i)\log_2(\frac{p_i}{q_i}))$$

Since this means that P and Q have to be absolutely continuous probability distributions. Here denoted as another client flow. The features of clients are extracted from the network monitoring tool. After that compute KL divergence, if it is more than the threshold, and then the system detects the attacks. KL divergence is used extensively for identifying anomalies. In addition, effects like flash crowds can be detected and evaluated correctly. The information theoretical metrics have proved to be suitable also for distinguishing between Dos attacks and flash-crowds. The sql injection is created by clients. Due to user's logging in number of times, the DOS attack is created. After generating the attacks the network packets are captured by using the tools wire shark and generate the necessary data's and structures for further analysis. After that, the data is sent to the modules for intrusion detection and extrusion detection. In DOS attack and brute-force login attempts detection, Kullback libeler divergence method is used. At first, the features are extracted from the network monitoring tool. Then compute continuous probability distributions. The computed KL divergence is more than the threshold, and then the system detects the attacks.

SQL Injection Attack: 1. Query result size estimation, The Sql injection attacks occur when developers combine hard-coded strings with user-provided input to create queries. If input sources are not properly validated, attackers may be able to change the developer's intended SQL query by inserting new SQL keywords through specially crafted input strings. Our technique dynamically analyses the developer-intended query result size for any input, and detects attacks by comparing this against the result of the actual query. In SQL, the cardinality of the attributes and the selectivity are used to estimate the query result sizes. Therefore, we also assume that we can obtain the cardinality and selectivity information of each table. Two statistics are typically important in query optimization. One is the number of tuples contained in a single table. This value is known as the table cardinality, and is denoted by ||T||. The second statistic is known as the column cardinality, and is denoted by dx, wherein x is the column. The selectivity is basically a measure of how much variety there is in the values of a given column in relation to the total number of rows in a given table. The selectivity of a database is calculated using a specific formula

Selectivity = selected number of tuples /total number of tuples

Equivalence and Largest Selectivity algorithm is used for estimates the query result sizes. The ELS algorithm consists of two important phases. The first is a preliminary phase, which is performed before any query result sizes are computed. In this phase, the predicates implied by the join selectivity's are calculated. The second phase incrementally computes the query result sizes.

2. Regular Expressions for SQL Injection, A regular expression is a notation to specify a set of strings. A regular expression is a sequence of the following items: A literal character, a matching character, character set, or character class, A repetition quantifier, An alternation clause, A sub pattern grouped with parentheses. An important point to keep in mind while choosing your regular expression(s) for detecting SQL Injection attacks[2] is that an attacker can inject SQL into input taken from a form.

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

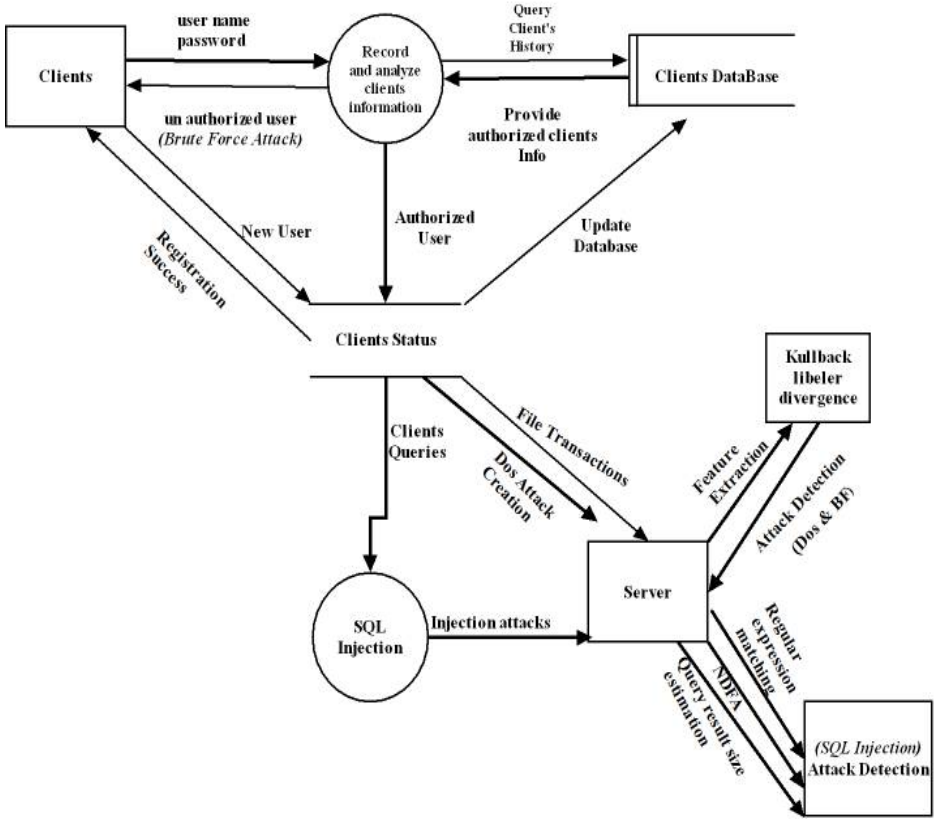

Fig 3: Block Diagram

3. Regular expression matching using NFA [3], A regular expression consists of characters and meta characters. Our implementation accepts the following meta characters: '*' (repetition of more than zero character); '?' (Zero or one character); '.' (An arbitrary character); '+' (more than one repetition of character); '()' (specify the priority of the operation); '|' (logical OR). Fig 3 describes the block diagram of proposed work.

NFA: NFAs are like DFAs with two additions: Allow δ(q, a) to specify more than one successor state, Add ε-transitions. Formally, an NFA is a 5-tuple (Q, Σ, δ, q 0, F), Where, Q is a finite set of states, Σ is a finite set (alphabet) of input symbols, δ: $Q \times \Sigma \rightarrow P(Q)$ is the transition function ,$q \in Q$, is the start state, $F \subseteq Q$ is the set of accepting, or final states, ε- Denotes an empty character. In the NFA, the empty (ε) input is permitted. Thus, a state for the NFA can transit to multiple states. The state transition with ε input denotes a ε transition. At first the regular expressions is converted into NFAs [7], where 'ε' denotes a ε-transition.

A SQL injection attack is detected by using three methods. They are Regular expression matching, Query result size estimation and regular expression matching using Non Deterministic Finite Automata. The performances metric such as accuracy, false alarm rate, false negative rates, precision, recall everything will be calculated and compared with the existing system.

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

IV. EXPERIMENTAL RESULTS

As per the experimental results fig 4 presents the detection results for client connections. Having one percent of attacksof all connections, the PD is 91.70 percent and the ACC of the system is 84.32 percent. With anincreasing ratio of attacks, these values decrease;having 2.7 percent of attacks the PD is 86.34percent and the ACC is 82.24 percent (2.7 per-cent was the next step, because the user trafficwas generated on a randomized time base resulting in non-integral ratios).For the classification of a connection, multiplesingle connection evaluationsare used to increasethe accurateness. Every connection is correlatedwith 11 other connections as this value was identified as the optimal number of correlation partners by empirical test runs. Adding or reducingone connection has a slight influence of about0.1 percent and, of course, rises if further connections are added/reduced. Based on that, theconnections are classified as benign or malicious.Having one percent of malicious connections,the final detection rate is about 99.3 percent.With an increasing portion of malicious connections, the detection rate drops because of thepossibility to examine multiple malicious onesagainst each other (resulting in higher similarityvalues). Having 2.7 percent of attacks, the detection rate is still 97.3 percent.

	Accuracy [%]	Prob. of detection[%]	False alarm rate[%]	False negative rate [%]
Intersession correlation				
1% malicious traffic	84.32	91.70	15.92	10.32
2.7% malicious traffic	82.24	86.34	17.80	24.26
Command evaluation				
Malicious session	86.00	96.37	45.56	4.26
Benign session	75.00	83.63	35.33	15.32

Fig 4:Detection capabilities

If a command series complieswith an attack path, an alarm will be raised (single nodes can be left unset if a command wasmarked unknown). By that, the identification ofall malicious sessions was possible. Therefore anidentification of attacks is possible withoutknowledge of the target service, the transporteddata, or the need for a system configuration. Comparing with the previous method this innovative method gives more accurate detection of attacks in the encrypted web applications.

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

V. CONCLUSION

Research in intrusion detection has been done for several decades. Although numerous systems have been proposed, current systems are not able to detect intrusion hidden in an encrypted payload with non-intrusive means. While knowledge-based systems are not able to detect encrypted attacks, behavior-based systems are suffering from high FARs. To overcome the current shortcomings, propose a new architecture based on the use of inherent knowledge of data connections by the calculation of their similarities. In contrast to existing approaches, this architecture does not need a learning phase nor a complex configuration or knowledge about the service to protect. Therefore, the architecture can be used widely and without any specific configuration. All information required is gathered in real-time from current connections by the use of similarity measurements. The brute-force login attempts, DOS attack and sql injections had been executed against the FTP. The brute force attack is created by using hydra tool. It can perform rapid dictionary attacks against FTP. The sql injection is created by FTP clients. Due to users couldn't get the needed resource due to conjunction, the DOS attack is created. After generating the attacks the network packets are captured by using the tool wire shark and generate the necessary data's and structures for further analysis. After that, the data is sent to the modules for intrusion detection and extrusion detection. In DOS attack and brute-force login attempts detection, Kullback libeler divergence method is used. At first, the features are extracted from the network monitoring tool. Then compute continuous probability distributions. The computed KL divergence is more than the threshold, and then the system detects the attacks. A Sql injection attack is detected by using three methods. They are Regular expression matching, Query result size estimation and regular expression matching using Non Deterministic Finite Automata. Similarity based intrusion and extrusion detection show that the system detects various attacks with a high degree of accuracy.

REFERENCES

[1] Koch .R,GollingM,Rodosek G.D, "Behavior-Based Intrusion Detection in Encrypted Environments", IEEE Communications Magazine, vol.12,pp.232,2014.
[2] C.V. Wrightetal, J. Mach. Learn, "On Inferring Application Protocol Behaviors in Encrypted Network Traffic", IEEE Transactions, vol. 7, pp. 2745–69, Dec. 2006.
[3] Young-Su Jang, Jin-Young Choi, "Detecting SQL injection attacks using query result size '', IEEE Communications,vol.11, pp. 23–26, 2014.
[4] Liu Yang a, RezwanaKarim a, VinodGanapathy a, Randy Smith , "Fast, memory-efficient regular expression matching with NFA-OBDDs", vol. 13,2013.
[5] Monowar H. Bhuyana,, D. K. Bhattacharyyab, J. K. Kalitac, "An empirical evaluation of information metrics for low-rate and high-rate DDoS attack detection'', vol. 21 , 2014.
[6] V. Paxson, "Bro: A system for detecting network intruders in real-time", IEEE Transactions on Computer Networks, vol. 31, pp. 2435–2463, Dec. 2013.
[7] JaspreetKaur , Rupinder Singh and PawandeepKaur, "Prevention of DDoS and Brute Force Attacks on Web Log Files using Combination Neural Network'', International Journal of Computer Applications ,Volume 120 – No.23, June 2015.
[8] H.Nakahara, T. Sasao, "A Regular Expression Matching Circuit Based on a Modular Non-Deterministic Finite Automaton with Multi Character Transition",IEEE Transactions on Computer Networks, vol 21,june 2014 .
[9] Ming-Yang Su "Real-time anomaly detection systems for Denial-of-Service attacks by weighted k-nearest-neighbor classifiers",IEEE Transactions,vol. 65,pp.45-56,2014