ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 15, March 2016

A NOVEL APPROACH TO INTRUSION DETECTION IN ENCRYPTED ENVIRONMENTS

Princy P,Scaria Alex,Ambikadevi Amma T

JCET College, Palakkad, Kerala

ABSTRACT: Nowadays, the Internet is established in numerous areas of everyday life. In recent years the Internet has evolved in to a critical communication infrastructure that is omnipresent in almost all aspects of daily life. This dependence of modern societies on the Internet has also resulted in more criminals using the Internet for their purposes, causing a steady increase of attacks, both in terms of quantity as well as quality. 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. Because the IDSes inspect the contents of a packet, it is difficult to find attacks by the current IDS. This project presents a novel approach to intrusion detection for encrypted environment. This approach applies encrypted traffic analysis to intrusion detection, which analyzes 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. The system does not extract private information or require enormous pre-operation beforehand, which are needed in conventional encrypted traffic analysis.Although research on the detection of attacks has been performed for several decades, today's systems are not able to cope with modern attack vectors. One of the reasons is the increasing use of encrypted communication that strongly limits the detection of malicious activities. While encryption provides a number of significant advantages for the end user like, for example, an increased level of privacy, many classical approaches of intrusion detection fail. Since it is typically not possible to decrypt the traffic, performing analysis w.r.t. the presence of certain

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 15, March 2016

patterns is almost impossible. 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 with a high degree of accuracy.

Keywords: IDS, SQL, DOS, Brute-force.

INTRODUCTION

In recent years the Internet has evolved into a critical communication infrastructure that is omnipresent in almost all aspects of our daily life. This dependence of modern societies on the Internet has also resulted in more criminals using the Internet for their purposes, causing a steady increase of attacks, both in terms of quantity as well as quality. 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, for example, to exfiltrate classified documents. Also, the detection of malicious actions done by an authorized person needs to be addressed.

DETECTION APPROACHES

An Intrusion Detection Systems (IDSs) can be divided into two basic classes: knowledge-based and behavior-based

Knowledge-Based Techniques

Knowledge- based detection applies knowledge accumulated about specific attacks and system vulnerabilities. The IDS contains information about these vulnerabilities and looks for attempts to exploit them. Some popular IDSs like Snort, Suricata, or Cisco's Secure IDS use so called signatures. Therefore, knowledge-based IDSs are often referred to as signature-based IDSs. These signatures contain rules (=knowledge) that are applied to the traffic, for example, by using string-matching techniques. In addition, next to popular signature-based IDSs, there are other approaches in the field of knowledge-based intrusion detection that are not in common use lately, for example, expert

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 15, March 2016

systems or state-modeling techniques like state-transition and petrinets. Other techniques besides pattern-matching concepts can also be used. Furthermore, next to packet-based techniques, knowledge-based approaches can also be applied to flows. Knowledge-based techniques are reactive from nature and not able to detect new and unknown threats. Even more, because of the time required to create signatures, they are often available shortly before the patches can be obtained as well

Behavior-Based Techniques

Behavior-based techniques build models of benign activities in a network. These models are used to predict the expected state of the network, which is then compared to the current, measured state. If the predicted and the measured value differs more than a specific threshold, an alarm will be raised because of the significant deviation. For the creation of the models (respectively baseline), these techniques typically need a comprehensive learning phase, sometimes called training period, in order to develop the model by monitoring the characteristics of typical activities over a period of time. For example, let us assume that the Web activity of an enterprise comprises an average of 13 percent of network bandwidth at the gateway to the Internet during typical workday hours, measured and averaged over several weeks. When Web activity comprises significantly more bandwidth than expected, an alarm can be raised, for example, informing an administrator. However, creating accurate models is a challenging task. If for instance a backup is performed incrementally on a daily basis, but only once a month fully, the network load will intentionally rise once a month. Maybe this will not be observed during the training, if the learning period only covers a week. Thus, behavior-based approaches often suffer from high false alarm rates, because of unknown benign behavior.

Intrusion detection is an important area of research. Traditionally, the approach taken to find attacks is to inspect the contents of every packet. However, packet inspection cannot easily be performed at high-speeds. Therefore, researchers and operators started investigating alternative approaches, such as flow-based intrusion

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 15, March 2016

detection. "A flow is defined as a set of IP packets passing an observation point in the network during a certain time interval. All packets belonging to a particular flow have a set of common properties."

PREVIOUS METHOD

A HTTPS-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.

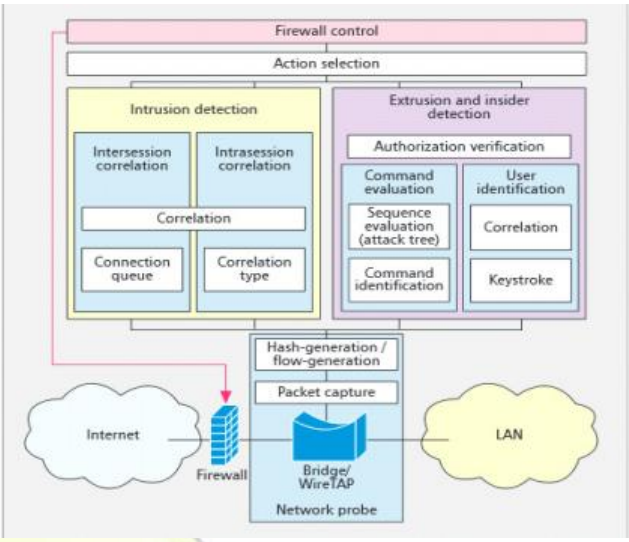

Fig: Architecture of intrusion detection system

NETWORK PROBE

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.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 15, March 2016

The Packet Capturemodule 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 of Fowler, Noll, and Vo; FNV). 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.

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, e.g. winnertakes-all. Also, the differences between the corresponding values of two series can be weighted differently. 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 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, 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

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 15, March 2016

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 Correlationconsiders 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 Correlationis 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 oneuser 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 analyzes the connection and forwards the information needed for the calculation of the similarities to the Correlationmodule.

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 analyzes 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

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 15, March 2016

clusters where a cluster consists of the encrypted packets of one command and the corresponding response of the server. The beginning and end of each cluster are identified by the transport direction, the timings, and the sizes of the payloads. Within the next step the similarity of these clusters to different commands (generated in advance) is calculated and the corresponding probabilities are saved for each cluster.For our work, attack trees are designed and used to identify the probability that a sequence of commands is used for executing an attack. A part of an attack tree describing possible actions for a privilege escalation. Three different approaches are analysed: the use of exploits, searching for configuration errors, and brute-force related attacks. Every branch of the tree can then be split into further paths, where every path can consist of several commands which in turn can be used to execute a potential malicious action. The concept of attack trees is used in conjunction with the knowledge about commands used during an attack to construct prototypical (hacking as well as benign) sessions. The resulting statistical data of the packet streams generated by these sessions can be stored and used afterwards for the similarity analysis. The attack trees are also required to evaluate if a sequence of commands can be used for an attack. This is realized by the Sequence Evaluation. If the commands are able to accomplish a path in an attack tree a (possible) malicious behavior will be identified and an alarm can be raised.

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 Identificationin our architecture. The Keystrokemodule receives the timings of the encrypted packets (and therefore of the keystrokes) of a connection.

PROPOSED SYSTEM (IMPLEMENTATION)

FTP encrypted server and FTP clients are created by using FileCOPA tool. 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.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 15, March 2016

The sql injection is created by FTP clients. Due to user's loging in number of times, the DOS attack is created. After generating the attacks the network packets are captured by using the tools pcap library or 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 andextrusion 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. The performances metric such as accuracy, false alarm rate, false negative rates, precision, recall everything will be calculated and compared with the existing system.

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. The design enables attack detection independently from (i) the exact communication between server and client, (ii) ports and IP addresses used, (iii) the encryption algorithm in use, (iv) the configuration of the server and (v) the speed as well as (vi) the kind of attacks executed. Therefore, the architecture can be used widely and without any specific configuration. All information required is

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 15, March 2016

gathered in real-time from current connections by the use of similarity measurements. Because some of the parameters used by the architecture are preset, e.g., the number of connections to correlate with, further evaluation will be done to test if this value can be optimized w.r.t. the target network. For intersession correlations, a minimum number of concurrent users is required. Therefore, we will analyze the influence of benign users in more detail, considering aspects like the validity period of data temporarily saved in the Connection Queue. The detection capability of the Extrusion and Insider Detection is limited by the commands known by the Command Identification and the attack trees in the Sequence Evaluation. At the moment, the Command Evaluation is done by two modules, testing the similarity to known commands on the one hand and using prototypical sessions on the other hand. For the first approach, the best set of implemented commands has to be investigated in detail while for the second one, the optimal construction of prototypical sessions has to be examined. After that, the capabilities of the two modules have to be evaluated and compared again.