Volume 3, Issue 12, December 2013 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

A Review of Intrusions Detection in Multi-Tier web Applications

Patil Asha
Student, Department of Computer Engg.
PVPIT, Pune. Pune University, India

Prof. V.M.Lomte
Professor Department of Computer Engg.
PVPIT Pune, Pune University, India

Abstract: ***A web application is an application that is accessed over a network such as the Internet. They are increasingly used for critical services, In order to adopt with increase in demand and data complexity web application are moved to multitier Design. As web servers must be publicly available around the clock the server is an easy target for outside intruders. Thus web applications are become a popular and valuable target for security attacks. These attacks have recently become more diverse and attention of an attacker have been shifted from attacking the front-end and exploiting vulnerabilities of the web applications in order to corrupt the back-end database system. In order to penetrate their targets, attackers may exploit well known service vulnerabilities. To protect multitier web applications several intrusion detection systems has been proposed. We survey several methods those are meant for intrusion detection. Some of them use known Priori prepared patterns also called signatures of known attack such system are grouped under the category of misuse detection. While some Methods deal with profiling user behavior. In other words, they define a certain model of a user normal activity. Any deviation from this model is regarded as anomalous such methods are termed as Anomaly detection methods.***

Keywords: vulnerabilities, intrusion detection system, anomaly-based, multi-tiered web application, container based approach.

1. INTRODUCTION

Intrusion detection plays one of the key roles in computer system security techniques. An intrusion detection system (IDS) is a device or software application that monitors network or system activities for malicious activities or policy violations and produces alerts. There are two general approaches to intrusion detection: anomaly detection and misuse detection. A signature based IDS works similar to anti-virus software. It employs a signature database of well-known attacks, and a successful match with current input raises an alert. Similarly to anti-virus software, which fails to identify unknown viruses a signature-based IDS fails to detect unknown attacks. To overcome this limitation, researchers have been developing anomaly-based IDSs. An Anomaly-Based Intrusion Detection System is a system for detecting computer intrusions and misuse by monitoring system activity and classifying it as either normal or anomalous. It works by building a model of normal data/usage patterns, and then it compares the current input with the model. A significant difference is marked as an anomaly.

Fig.1 Three tier Architecture

In multi-tier web architecture often referred to as n-tier architecture. The back-end database server are kept protected behind a firewall and web application made it possible for user to access set of services from web servers which are remotely accessible over the Internet .The current ids system installed at web server and at database server are unable to detect intrusions where a normal traffic is used for attacking back end database [6]. We found that these IDS cannot detect cases wherein normal traffic is used to attack the web server and the database server. Though they are protected from direct remote attacks, the back-end systems are susceptible to attacks that use web requests as a means to exploit the back-end. Existing prevention systems are often insufficient to protect this class of applications, because the security mechanisms provided are either not well-understood or simply disabled by the web developers ``to get the job done." Therefore, prevention mechanisms should be complemented by intrusion detection systems, which are able to identify attacks and provide early warning about suspicious activities.

2. RELATED WORK

There are two types of systems that are called anomaly detectors those based upon a set of rules which are further used for what is regarded as good or normal behavior, and others that learn the behavior of a system under normal operation. We summarize different methods used by intrusion detection systems to represent knowledge on a system and analyze

audit Information in order to detect an intrusion. Behavior models are built by using rule-based approaches to specify behavior patterns or by performing a statistical analysis on historical data. Signature based detection systems should work side-to-side with anomaly detection systems.

Rule Based Systems

Martin Roesch [1] proposed a new open source intrusion detection and prevention system which is based on use of handcrafted rules to identify known attacks. A human studies an attack and identifies the characteristics (e.g., behavior and/or content) that distinguish it from normal data or traffic. The combination of these characteristics is known as the signature, and it becomes part of a database of attack signatures. When the IDS encounters data matching the signature it raises an alarm. The remarkable thing about this approach is that combining the benefits of signature, protocol, and anomaly-based inspection. That is actually the basic difference in using rule-based expert systems for anomaly and misuse detection. In the first case, the rules are generated using some other techniques. In the second case, the rules are given to the system in advance. The behavior rule based intrusion detection usually depends on packet anomalies present in protocol header parts.

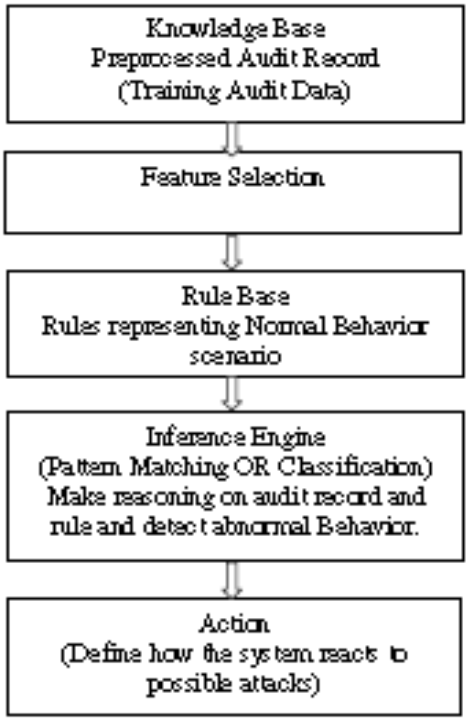
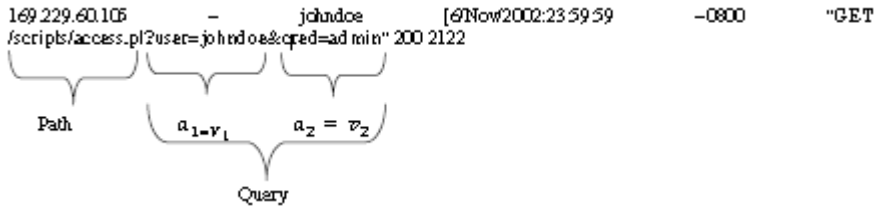


Fig.2 Rule based Intrusion Detection Approach

Behavior is communication procedures of software. Since all software runs and communicates in accordance with predefined programs, there is a communication pattern which consists of sent and received character strings, destinations, communication protocols, sending and receiving intervals and so on. Scenario-based intrusion detection method consists of not only sequential events but also random order events. And certain scenarios are described by correlations between communications. Therefore consider 2 types of correlations: Asynchronous event sequence: Event sequences are necessary to detect a communication behavior of software and for detection event order is important. Consider asynchronous event sequence can deal with scenario that includes three events (E1 → E2) → E3 that means observing E3 after both of E1 and E2. We have to prepare 2 event sequence E1 → E2 → E3 and E2→ E1 → E3 to describe this event sequence by generic state transition machines. Data pattern Matching: Two or more communications have correlation with respect to IP addresses, port numbers, domain names, URLs and so on. Or correlation in sending/receiving commands, queries and so on. These correlations can be verified by comparing some packet headers/payload data. One of the major disadvantages of this method is that Rule based anomaly detection techniques learn rules that captures the normal behavior of a system. A test instance that is not covered by any such rule is considered as an anomaly.

Multimodal Approach

Christopher Kruegel and Giovanni Vigna proposed [2] first anomaly detection system especially designed for the detection of web-based attacks. Attack is detected by applying simple pattern-matching techniques to the contents of HTTP requests. During the detection phase we analyze the all HTTP request logged by most common web server like apache web server. System takes input the web server log files and analyze Common Log Format and produces an anomaly score for each web



Input to the detection system is U= {µ1, µ2, µ3….µm} (Set of URI's those are extracted from successful GET request) the analysis process focuses on the association between programs, parameters, and their values. Consider Only GET requests with no header Request. For query q, Sq = {a1, a2} Detection process uses a number of different models to identify anomalous entries within a set of input requests associated with a program r. A model is used to assign a probability value to either a query or one of the query's attributes. This probability value indicates the occurrence of given feature with regard to an established profile. The feature value with sufficiently low probability value indicates potential future attack. Model can operate in one of the two modes as follows Training: Training phase is required to characterize the behavior of specific model and allow models to learn the characteristics of normal events try to set anomaly score threshold values in order to distinguish between normal and anomalous inputs. Detection: In this phase only anomaly score are calculated and anomalous queries are reported. Anomaly score is Probability value returned by corresponding model that are associated with the query or one of the attribute. A value close to 0 indicates anomalous event i.e. a value of pm close to 1 indicates anomalous event.

$$\text{Anomaly score} = \sum_{m \in \text{Models}} w_m * (1 - p_m)$$

Where,

Wm = Weight associated with model m.
Pm = Probability value returned by model m

If the weighted score is greater than the detection threshold determined during the learning phase for that parameter, the anomaly detector considers the entire request anomalous and raises an alert. Main advantage of this technique is attacker cannot hide single malicious attribute within query with many normal attribute.

Webstat

Giovanni Vigna, William Robertson, Vishal Kher, Richard A. Kemmerer [3] proposed an integrated approach which is based on stateful analysis of multiple event streams. In this Approach Intrusion is defined as sequence of intruder actions that bring system from normal state to compromised state trough a number of intermediate states. State Transition Analysis method then analyzes a sequence of actions that an intruder performs in order to break into a system and such sequence of actions is called signature actions. Thus signature actions means minimum possible set of actions needed to perform successful attack. Those states, Transitions, actions are represented by State transition diagram thus it is easy to model an behavior of multistage and complex attack by using state transition diagram. It uses Language extension module that defines web-specific events .An event provider that parses web server logs and generates the corresponding events and collects events from external environment. A STATL description of an attack scenario used by intrusion detection system to analyze a stream of events and detect possible ongoing intrusions. A number of STATL scenarios were developed to detect attacks against web servers. STATL scenario uses variables to record just those parts of the system state that are needed to define an attack signature. These attacks are depending on one or more event streams. The event provider reads the events stored in the server application log file as they are generated. Event provider Create events as defined in Language Extension Modules and inserts events into the event queue of the STAT Core. The STAT Core extracts the events from the event queue and passes them to active attack scenarios for analysis. WebSTAT consider multiple event streams and thus it is able to correlate both network-level and operating system-level events with entries contained in server logs. Advantage of this technique is that threat scenario is represented in a visual form and very easy to read. The key point in this detection approach is signatures actions must be accurate for the formulation of intrusions. Since the list of attribute changes to be recorded for a system is comprehensive, but all the attributes cannot be recorded. This may not give all the possible set of actions needed to formulate intrusions it ends up as fewer transaction states. This results in inappropriate signature actions.

Swaddler

Marco Cova, Davide Balzarotti, Viktoria Felmetsger, and Giovanni Vigna [4] proposed an novel approach which is based upon detailed characterization of the internal state of a web application, by means of a number of anomaly models. Web application internal state is defined as information that survives single client and server interaction or simply the information associated with single user session. The minimum state information is passed as a cookie to a browser. Minimum context information such as a session ID must be passed between the browser and the server to identify the rest of the state information. The key point here is it is easy to model out typical intrusion scenario by keeping track of all states in which that intrusion is normally executed.System operates in two modes Training and Detection. During training

phase the profiles for the application blocks are formed using the events generated by sensor. And during detection phase these profiles are used to identify the anomalous application states. Main Advantage is that there are Attacks that cannot be detected only by observing the external behavior of web application. This approach detects attacks that attempt to bring an application in an inconsistent anomalous state.

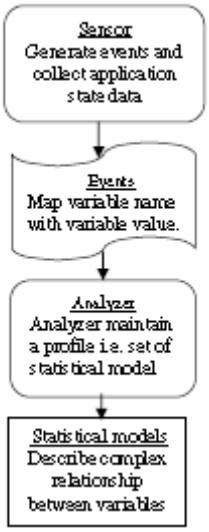


Fig 3 Profile creation phase

Combined Approach for Analysis Web Request and Database Request

Giovanni Vigna, Fredrik Valeur, Davide Balzarotti, William Robertson, Christopher Kruegel,Engin Kirda [5]proposed a system for anomaly detection which is composed of web-based anomaly detector, a reverse HTTP proxy, and a SQL query anomaly detector. The key approach on which system is designed is as follows.

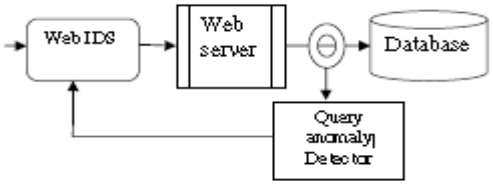


Fig. 4 Combined analysis web request and database request

In this approach intrusion detection system is implemented at both levels at webserver and database server because of addition of detection system at database level the analysis of query will allows system to detect malicious web request which are mistakenly considered as normal. And when such type of anomalous query is detected in association with the normal web request a description of anomaly is sent over feedback channel to the webserver anomaly detection system in order to update the model accordingly and prevent future attack. As Approach applies the serial composition of anomaly detector it suffers from increase in false positive in order to deal with this problem work is supplemented by a novel techniques data compartmentalization and reverse proxying. The key idea here is to replicate the website on two or more webserver with different levels of privilege such severs are called as sibling web server. Anomaly score obtained by web based anomaly detection is used to drive a reverse HTTP proxy which is application installed at sibling web server. The job of reverse proxy is to intercept HTTP request which is destined for webserver and depending upon the individual web request anomaly score forwards request to sibling webserver with appropriate level of privilege

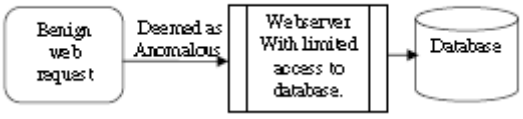


Fig.5 HTTP Reverse Proxy

The web request which is mistakenly treated as anomalous is simply send to the webserver having limited access to the database instead of being dropped and if request does not need to access the sensitive information that request will be served correctly. Thus system provides reduced level service in false positive. The disadvantage of above approach is it cannot detect attack where normal web requests are used as means to exploit back end database. These two independent

IDS installed at webserver and database server fails detect intrusion cases wherein normal traffic is used to attack the web server and the database server.

FPGA-Based Intrusion Detection

Abhishek Das, David Nguyen, Joseph Zambreno, Gokhan Memik, and Alok Choudhary [6] proposed a Field Programmable Gate Arrays based architecture for anomaly detection for network intrusion detection. A field-programmable gate array (FPGA) is an integrated circuit designed to be configured by a customer or a designer after manufacturing. It comprises new feature extraction module which is meant for collecting network characteristics feature and PCA as detection method.

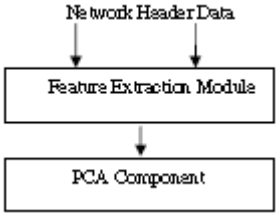


Fig .6 FPGA-based intrusion detection

Collected network data is fed to the FEM and FEM then analyzes Network behavior on temporal basis or on specified interval of connection. The key idea used in this approach is to model out an anomalous behavior associated with two general types of intrusions first is time based and second is connection based. The FEM consist of number of components such as feature sketch (FS) which is an application of sketches used for data-stream modeling.Feature controller (FC) which controls input to hash function by using flag. Second is the Hash functions (HF) each row in the FS is addressed by a different HF, and a data aggregate (DA) component takes H values and estimates the actual value for a query. With the help of all these components network characteristics can be monitored and tracked in real time. Once all features are extracted the resulting values are fed into an outlier detection scheme in order to capture the attacks. PCA is used to reduce dimensions of data without much loss of information.PCA that transfers the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate i.e. first principal component. And the second greatest variance lies on the second coordinate i.e. second principal component. It projects a new set of axes which best suit the data. These sets of axes model the normal connection data. During detection phase mapping of live network data onto these "normal "axes is done and distance between axis is calculated. If the distance is greater than a certain threshold, then the connection is classified as an abnormal. As discussed above the technique is purely anomaly based and we need to adopt combination of signature based and anomaly based detection technique in order to have sound and highly efficient intrusion detection.

Semi-Markov Model for Anomaly Detection on User Browsing Behavior

Yi Xie and Shun-Zheng Yu [7] proposed a new approach which is entirely based upon hidden semi Markov model which is used to briefly describe the browsing behaviors of web user and detect online application associated DDOS attack.
User's normality is judged on Entropy of the user's HTTP request sequence.HsMM is an extension of the hidden Markov model (HMM) with explicit state duration. It is a stochastic finite state machine which is best described by (S, Π , A, P.)
Where:

S is discrete set of Hidden state with cardinality N.
Π is the probability distribution for the initial state.
A is the state transition matrix with probabilities.
P is the state duration matrix with probabilities.

HTTP request sequence received by the webserver is (r1, r2, r3, r4, r5, r6, r7, r8, r9,r10).When the observed request sequence is inputted to the HsMM the algorithm may group them into three clusters as follows (r1,r2,r3,r4).(r5,r6,r7),(r8,r9,r10,r11)and denote them as state sequence (req1,req2,req3).HsMM trains the model from a set of request sequences made by a lot of normal users and characterizes the normal users browsing behaviors during detection. We simply measure the deviation of an observed request sequence made by a user with the mean entropy of training data. In this scheme user's normality depends on the entropy of his/her HTTP sequence fitting to the model. Thus we need to set a threshold on the sequence's length to decide whether the sequence is normal or not. Threshold is "decision length" which considers the total no of the request in the sequence "sequence decision length" and the time factor for HTTP sequence, "time decision length". Thus real-time response time and precision of our detection system is depending upon the decision length.

Histogram-Based Traffic Anomaly Detection

Andreas Kind, Marc Ph. Stoecklin, and Xenofontas Dimitropoulos [8] described a new approach in which we simulate different traffic feature with the help of histogram. Model out histogram patterns, and identifies deviations from the created models. The key idea on which entire approach is based upon is as follows. The constructed histograms which simulate network feature follows regular patterns and model the normal behavior of a network. Network anomalies may disturb the shape of normal patterns of one or more features. In this approach real world network traffic is collected and

depending upon number of traffic feature anomalies' are detected. Patterns of common Behavior are identified by quantifying how similar two histograms are. A number of different approaches can be used to quantify how similar two Histograms are. Clustering is needed for identifying and modeling patterns of normal behavior. After performing clustering we need to distinguish the clusters that correspond to the normal and anomalous behavior. Set of clusters that model the normal behavior of a network we keep it as baseline. During detection phase we measure how the observed network behavior differs from baseline. For each feature the anomaly detection system computes a vector that encodes the online behavior of the network. If the vector falls within the scope of baseline clusters, then the online behavior is considered normal. Otherwise the behavior of the network is considered abnormal.

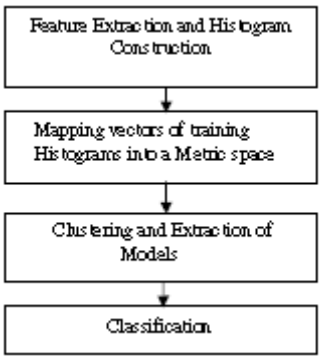


Fig. 7.Histogram-based traffic anomaly detection

DOUBLEUARD

The limitations of all above discussed approach are considerably removed by this approach where normal traffic is used as means to exploit back end database systems.

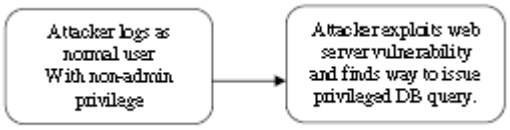


Fig.8.Threat scenario

In such cases neither the web server IDS nor the database server IDS would able to detect attack. Because the web IDS would assume the traffic of normal user login and Database IDS would assume traffic of privileged user. Meixing Le, Angelos Stavrou, Brent ByungHoon Kang [9] proposed a new approach called Double guard to detect intrusions in multitier web applications. This approach assumes that there is causal mapping of web requests and resulting SQL queries in a given session.And above modeled attack can be readily detected if the database IDS can determine that a privileged request from the web server is not associated with user-privileged access. And the entire approach of Doubleguard is based upon the mapping model which maps the web request along with set of resultant query invoked by that request within an individual session.The mapping model it can be used to detect abnormal behaviors. Both the web request and the database Queries within each session should be in accordance with the model. If there exists any request or query that violates the normality model within a session, then the session will be treated as a possible attack.

Advantages of Doubleguard are:

1.All the other approach use intrusion alerts aggregations and alerts correlation where alerts are classified into some meaningful groups or simply Group alerts into attack threads one thread contains all alerts related to one attack. Such type of alert aggregation and correlation is not required in Doubleguard approach.

2. As Doubleguard does not classifies event on time basis as it uses container based and session separated so it uses container id to casually map related events.

3. Doubleguard approach also does not require us to analyze the source code or know the application logic for intrusion detection.

4. Doubleguard can detect SQL injection attacks by taking the structures of web requests and database queries without looking into the values of input parameters.

3. CONCLUSION

In this way we surveyed few techniques which are meant for intrusion detection against multitier web applications. Some of the technique use single IDS to detect and prevent webserver from malicious request while some approach use combined approach to detect intrusions at both web and database level. Apart from all above discussed approach the last approach is having some additional detection capability to detect attack where normal traffic is used as means to launch database attack. Because of container based and session separated approach of Doubleguard use multiple input streams to

produce alerts. Such correlation of different data streams provides a better characterization of the system for Anomaly detection because the intrusion sensor has a more Precise normality model that detects a wider range of threats. This approach is more advantageous in sense that monitoring both web and subsequent database requests, we are able to detect attacks that an independent IDS would not be able to identify

REFERENCES

[1] http://www.snort.org.
[2] C. Kruegel and G. Vigna "Anomaly detection of web-based attacks" In Proceedings of the 10th ACM Conference on Computer and Communication Security (CCS '03), Washington,DC, Oct. 2003. ACM Press.
[3] G. Vigna, W. K. Robertson, V. Kher, and R. A. Kemmerer. A stateful intrusion detection system for world-wide web servers. In ACSAC 2003.IEEE Computer Society.
[4] M. Cova, D. Balzarotti, V. Felmetsger, and G. Vigna. Swaddler: An Approach for the Anomaly-based Detection of State Violations in WebApplications. In RAID 2007.
[5] G. Vigna, F. Valeur, D. Balzarotti, W. K. Robertson, C. Kruegel, and E. Kirda. Reducing errors in the anomaly-based detection of web-based attacks through the combined analysis of web requests and SQL queries. Journal of Computer Security, 2009.
[6] Abhishek Das, David Nguyen, Joseph Zambreno, Gokhan Memik, and Alok Choudhary An FPGA-Based Network Intrusion Detection Architecture IEEE transactions on information forensics and security, vol. 3, no. 1, march 2008
[7] Yi Xie and Shun-Zheng Yu A Large-Scale Hidden Semi-Markov Model for Anomaly Detection on User Browsing Behaviors IEEE/ACM transactions on networking, vol. 17, no.1, February 2009.
[8] Andreas Kind, Marc Ph. Stoecklin, and Xenofontas Dimitropoulos Histogram-Based Traffic Anomaly Detection IEEE transactions on network service management, vol. 6, no. 2,June 2009.
[9] Meixing Le, Angelos Stavrou, Brent Byung Hoon Kang, "DoubleGuard: Detecting Intrusions in Multi-tier Web Applications" 2012

