Novelty Journals

ISSN 2394-7314

International Journal of Novel Research in Computer Science and Software Engineering
Vol. 2, Issue 3, pp: (38-46), Month: September-December 2015, Available at: www.noveltyjournals.com

A Review on Improved Storage Security and Performance Using Container Deduplication

[1]Revati Rane, [2]Priyanka Gholap, [3]Smita Bari, [4]Kalyani Patil, [5]Prof. S.A. Ahirrao

[1, 2,3,4,5] Department of Computer Engineering, Sandip Institute of Engineering And Management,
Savitribai Phule Pune University, Nashik, India

Abstract: **Web services have moved outsourced to a multi-level design, in which the Web server, the input logic and back-end applications and performance data are on a database server or file. About a year last web services and applications server to attack the attacker had increased. In this paper we present a double guard, an IDS system that models the behaviour of network user sessions in the two Web server front-end and back-end database data. By controlling both web and database subsequent data requests, we are able to discover attacks that an independent IDS would not be able to identify. By using the allocation request and consult IDS security system for web servers and databases can provide. We implement double guard with an Apache web server with MySQL and easy to virtualization. Moreover, in the environment of cloud computing it is performed using the communication processing large files and therefore it is very important and important to provide efficient data security and file processing approach. Deduplication is a technique to remove copies of the data and are widely used in cloud storage to reduce storage space and bandwidth load. Convergent encryption widely adopted for safe de-duplication; a critical point of convergence Encryption is managing a large number of convergent key. Initially where each user maintains a baseline approach, an independent master key to encrypt the convergent and outsourcing to the cloud key. However, as a key management system reference creates an enormous number of keys with the growing number of users and requires that users spend protect the master key, which is an unreliable inefficient. To this end, we propose dekey, a new building in which users do not have to manage all the keys on their own, but certainly on multiple servers to distribute the key elements converged. Dekey is implemented using the RSA algorithm.**

Keywords: **IDS, virtualization, de-duplication, convergent. encryption.**

1. INTRODUCTION

In the world today there are many computer use specifically for web applications. Most people do their business through the web application. So chances are that personal data must then be hacked, more security for Web servers and database servers. To protect the web services of various levels, the intrusion detection system have been widely used to detect known attacks by merging abuses of traffic patterns to protect web services of various levels. A class of IDS [10] that uses machine learning to detect unknown attacks by identifying abnormal net-work traffic that deviates from the "normal" behaviour. Individually, the web IDS and IDS database abnormal network traffic to detect a ball of them. However, we found that these cases IDS detected in normal traffic is used to attack the Web server and database server. For example, if an attacker without administrator rights can to a normal web server with user access credentials logs on, he / she can find a way to a privileged database query data by exploiting vulnerabilities in the web edition. Neither the website nor the IDS database can see this type of attack, because web IDS would only typical user traffic and IDS database would only normal traffic of a privileged user. This type of attack can be easily detected, if the database identifiers can identify a request privilege from the Web server is not connected to the user privileged access. Unfortunately, the architecture of multi-threaded server present, it is impossible to detect or profiles as causal association between Internet traffic and DB-server

traffic, since traffic is not unique user sessions are attributed. In this paper we present a double guard [10] at the IDS system used to detect attacks on web services of various levels. In this system, the double guard model create normal user sessions isolated in both the front-end web, such as HTTP and backend as a file or SQL included for network transaction. Double Guard we are a bit virtualization technology for assigning web session for each user a dedicated container that provides isolation of the virtual environment. So let's be each Web request consultations with subsequent databases, the employee with the exact container ID have to take. Doubleguard will be borne by the web server and database traffic for mapping profile in the correct and accurate accounting. Therefore, double guard profile causal association both to build the accounting server and DB traffic. We have our container architecture OpenVZ double guard realized [1]. The Web container-based architecture not only promote the profile of the causal association, but also provides insulation which prevents future session-hijacking attacks. As ephemeral containers can be instantiated and destroyed easily, every client session we assigned a task included, it, so that even if an attacker is able to provide a unique meeting that commitment, damage to the session at risk is limited; other user sessions are not affected.

The advent of cloud storage that encourages businesses and organizations to outsource data storage cloud third. A key challenge for storage services in the cloud today is to manage increasing volumes of data. To make data management, deduplication [11] is a method known to reduce storage space and bandwidth to climb cloud storage. Instead of maintaining multiple copies of data with the same content, deduplication eliminates redundant data. Only a natural and other redundant data referring to this copy data deduplication occurs at the file level and block level. Duplicate the same file level deduplication file .To eliminate duplication block level, it eliminates duplicate copies of data blocks that occur in non-identical files. Traditional encryption requires different users to encrypt your data with your own key. So take copies identical data of different users in different ciphertexts for duplication impossible. Convergent encryption provides a viable option to enforce data confidentiality while performing deduplication. Decrypts encrypted data copy / one copy of convergent key by calculating the cryptographic hash of the contents of the data itself is derived.

Understanding how convergent encryption can be implemented, we believe it is a layered approach. That is, the original copy of the data is first with a key derived from the encrypted converged data itself copy, and the key is then converged by a master key that are performed locally and securely from any encrypted user. The encrypted key is stored then convergent, copied along with the corresponding encrypted data storage in the cloud. The master key can be used to recover the encrypted keys, and therefore encrypted files. Thus, each user will only have to maintain the master key and metadata about the exported data. However, the focus of the baseline suffers from two critical questions 1) there is a huge number of keys with the growing number of users.2) generate the accidental loss of master keys.

Therefore, to overcome this problem, it proposes a new design called offering efficiency dekey and guaranteed for key management convergence on users and sites cloud storage reliability. Our idea is to apply deduplication to convergent key. In particular, we are building secret keys and convergent actions to distribute them through several independent server keys. Only the first user to load data needed to calculate and distribute such covert actions, while all subsequent users who have the same data do not have to calculate and store again, copy these populations. The convergent actions of a secret key will be accessible only by authorized users who have the appropriate data backup. This reduces the memory requirements of the converged key and makes key management reliably against failures and attacks. To prevent unauthorized access, safe PoW (Proof of ownership) is also required to confirm that the user is in fact the same file, if it is a duplicate found. After confirming participants resulting link to the server is available with the same file without loading the related file will be provided. A user can select the encrypted file to the server pointer that can only be deciphered by the owners of the equivalent data with key converged download. So convergent encryption will enable the cloud for deduplication in the ciphertext and proof of ownership of shares (POW) does not prevent unauthorized users from accessing the file.

2. LITERATURE SURVEY

Vishwesh.N, et.al.[2]. This article is about the intrusion alerts correlation provides a collection of components, sensor alerts intrusion detection intrusion concise reports to reduce the number of positive replicated false positive transformation warnings and irrelevant. In addition, warnings multilevel connects one attack, described the goal of producing an overview of the activities related to network security. Rather, double Guard works in multiple feeds of network traffic using a single IDS, which sees more than one session to produce an alarm uncorrelated or summary of alerts generated by other independent IDS.

Novelty Journals

ISSN 2394-7314

International Journal of Novel Research in Computer Science and Software Engineering
Vol. 2, Issue 3, pp: (38-46), Month: September-December 2015, Available at: www.noveltyjournals.com

Christopher Kruegel and Giovanni Vigna [3], this work is anomaly detection Web-based attacks. , This paper presents an intrusion detection system that uses a number of different anomaly detection techniques to detect attacks on Web servers and Web-based applications. The system correlates server programs referred to client requests with the parameters contained in these consultations. The specific properties of the application of the parameters, the system analyses focused and generate a small number of false positives.

Bryan Parno, et.al [4], this paper is CLAMP; an architecture to prevent data leakage, even in the presence of commitment web server or SQL injection attacks. CLAMP protects sensitive data by implementing strong access control and user data by isolating code running on behalf of different users. By contrast, it focuses on modelling Double pattern among HTTP requests and queries database to detect malicious users sessions.

S.Shalini, S.Usha [5], this document describes the attack prevention cross-site scripting (XSS) vulnerabilities in web applications on the client side. Scripting attacks (XSS) Cross site they occur when access to information from trusted sites. Client-side solution acts as a web proxy for cross-site scripting attacks, manually generated rules to mitigate cross-site scripting attempts to alleviate. This is orthogonal to the two-pronged approach, to use input validation as an additional defence. Symmetric encryption uses a shared secret key to encrypt and decrypt information. A symmetric encryption Scheme consists of three basic functions generation:

1. Key 2.Encryption 3.Decryption.

Aparna Ajit Patil,et.al [6] .This paper describes the block-level deduplication at the storage system in the hybrid cloud. When a user uploads a file to the cloud, the file is divided into a series of blocks. Block are coded key share through convergent and then a token is generated for the token generation algorithm. After encrypting the data with key converged users hold the key before sending the ciphertext to the cloud. Each block is then compared with the database of the cloud. After comparison, if a match is found in the database cloud, only metadata is stored in DB block.

Shweta Pochhi, Vanita Babanne [7]. This paper reports on deduplication safe and authorized data. To overcome the attacks, the end of the proof of ownership (POWs) has introduced what a customer can test competently to a server, the client has a file.

N.O.Agrawal et.al [8]. This paper reports on insurance deduplication and data security CEKM efficient and reliable. The basic idea of this work is to eliminate copies of data storage and limit the damage from the stolen data, if we decrease the value of the stolen information to the attacker. In this paper, it is implemented user behavioural profiling technology and lures. User Profiles and lures, serve two purposes: First, the validation that is authorized access to the data, if access to information is detected abnormal, and the second is to confuse the attacker with false information. We postulate that the combination of these security features will provide an unprecedented level of security for deduplication and sundry attacker.

Ankita Mahajan and ratnaraj Kumar [9] .In this paper describes about data deduplication and approved in dynamic cloud computing. Presented data deduplication difference authorized to protecting data security privileges of users in the duplicate check. Several new sub-structures to support deduplication authorized duplicate check. Furthermore, the data in the dynamic cloud is another important area, which is considered. The framework to support dynamic data on the operation of various data such as block modification, insertion and deletion occurs.

3. PROPOSED SYSTEM

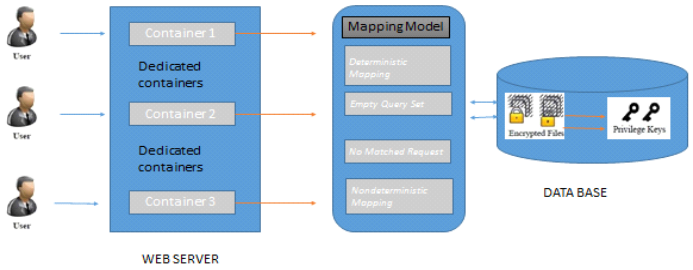

Fig 1: System architecture

Novelty Journals

ISSN 2394-7314

International Journal of Novel Research in Computer Science and Software Engineering
Vol. 2, Issue 3, pp: (38-46), Month: September-December 2015, Available at: www.noveltyjournals.com

The system architecture of the project is shown in Fig 1, the system architecture shows three modules or three layers of the user, the server and the database layer.

User uses applications that Server provides for him / her. Each user ID has a separate web application web container for processing.

As a container initialized easily destroyed and lasts only a short time, which would be capable of providing a single container for each user. We can initialize thousands of containers in a single system, so that these virtualized containers can be discarded or reset quickly re-initialized in order to serve new sessions. In Double new packaging and used packaging recycling they are approaching dynamically generated. Each session is assigned a dedicated web server and separate from other sessions. In this system, we represent and sign for different users around the same Web server to the application, we identify the behaviour of both the session and the use of user and if we are, or to detect abnormal behaviour in a meeting, we will handle all the network traffic in the session, as infected. If the attacker attacks the container then it remains in this container, without the knowledge of the presence or existence of another communication session. Our container-based server is a simple matter to identify causal pairs of Web applications and SQL queries resulting in a given session. Moreover, as the traffic can be easily separated by session, we may be able to compare and analyse the request and queries through various meetings. We put sensors on both sides of the server. On the Web server, our sensors are provided on the host system and cannot be attacked directly, since only the virtual containers are exposed to attackers. Our sensors are not attacked in the database server either as it is assumed that the attacker complete control of the database server.

Once the mapping model construct that can be used to detect abnormal behaviour. Both the Web application and database queries per session must be in Accor-dance with the model. If any request or query that violates the model of normality within a session is to treat the meeting as a possible attack.

A. Attack Scenarios:

Doubleguard System Intrusion Detection is effective in Record the following types of attacks:

1. Privilege Escalation Attack:

The fig.2 shows how regular users use management consultations to reach inside information.

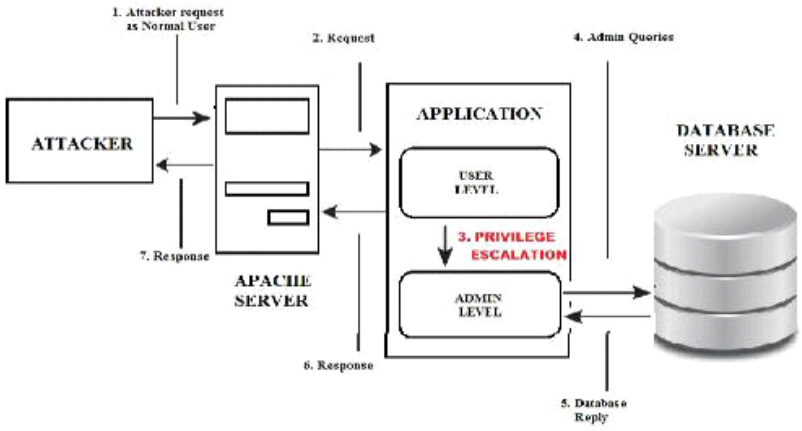

Fig 2: Privilege Escalation Attack

Suppose the site of both regular users and administrators. Regular users of a web application can be activated with the set of SQL queries when an administrator initiates a Web request to all administrator-level visits.

Suppose an attacker's Web server as a normal change or update their/its details and trying to get a data management by enabling a user query management. This type of attack can never been detected by IDS, which is either Web server or database because both IDS IDS allow requests and inquiries. But according to our allocation model, query database does not correspond to the request, so that can detect such attacks.

Novelty Journals

ISSN 2394-7314

International Journal of Novel Research in Computer Science and Software Engineering
Vol. 2, Issue 3, pp: (38-46), Month: September-December 2015, Available at: www.noveltyjournals.com

2. Hijack Future Session Attack:

Fig.3 is said that a web server can damage all future sessions kidnapping by not generating any query databases for the needs of general users.

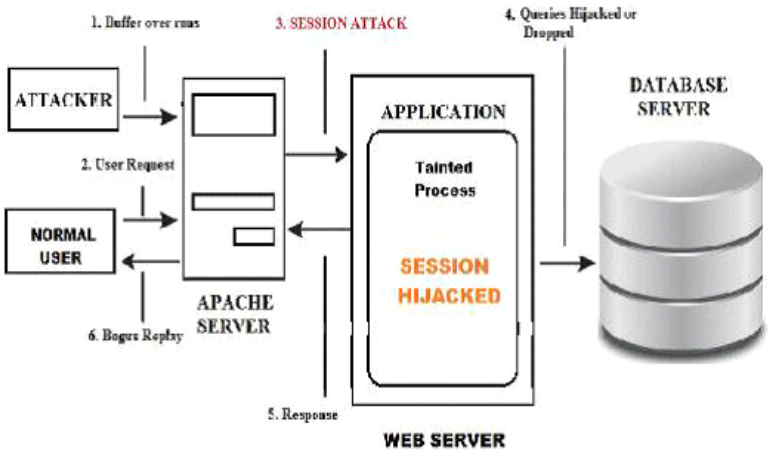

Fig 3: Hijack Future Session Attack

This type of attack has happened, especially in web server. An attacker takes over the Web server and kidnapped all sessions allowed users to launch attacks. An attacker could hear, send false answers and drop user request for hijacking sessions users. We can say that a man-in-the-middle attack A denial-of-service attack or a replay attack are the categories of session hijacking attack.

3. Injection Attack:

As shown in fig.4 this type of attack, an attacker can use existing exposure in the web server logic to inject the data or string content which contains the achievements and then use the web server to control these achievements to attack the backend database. The SQL injection attack changes the structure of SQL queries and it generates SQL queries in different structure, even if the injected data were to go through web server side.

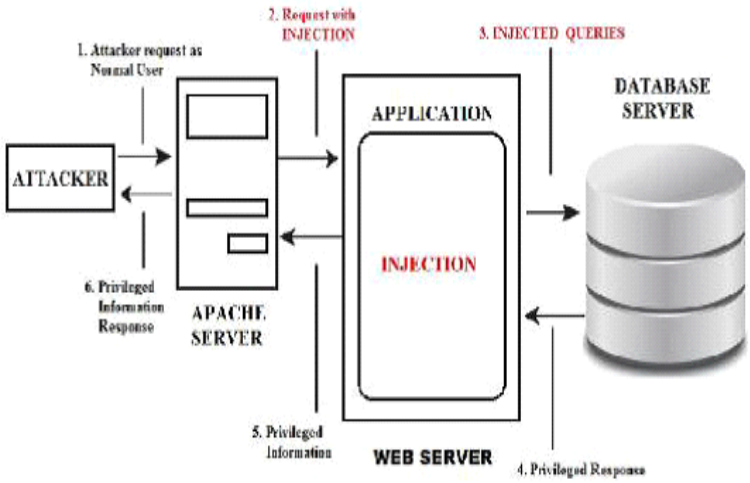

Fig 4: Injection Attack

4. Direct DB Attack:

As shown in fig.5 it is possible for an attacker, the web server or bypass firewalls and directly to the database, which is also shown already taken on the Web server.

Novelty Journals

ISSN 2394-7314

International Journal of Novel Research in Computer Science and Software Engineering
Vol. 2, Issue 3, pp: (38-46), Month: September-December 2015, Available at: www.noveltyjournals.com

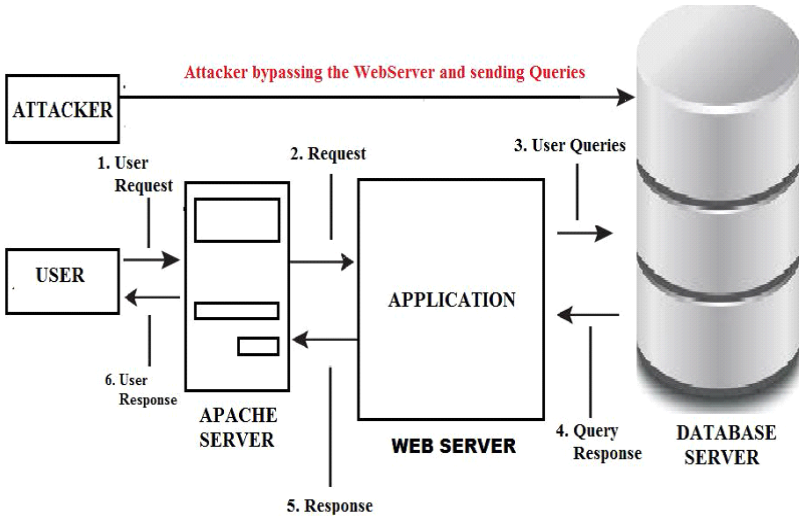

Fig 5: Direct DB Attack

An attacker and the submission of such requests from the web server without sending Web -requests. Without coordinated web requests for such consultations, the IDS Web server to recognize any. Also, if this DB enables data queries in the set of queries, the database identifiers would also not be recognized. However, this type of attack to capture our approach because we cannot match all Web requests on these issues.

B. Modeling Deterministic Mapping And Pattern:

We build an exact model of the mapping relationships between web requests and database queries as the links are static. Instead of one-to-one mapping, that is, one web request to the webserver usually invokes a number of SQL queries. It may happen that some requests will just retrieve data from the webserver ,that is, no queries will be generated by such web requests. Whereas, in some cases, one request will invoke a number of database queries. Finally, in some cases, the webserver will have some periodical tasks that trigger database queries without any web requests driving them. We categorize the four mapping patterns as follows. As the request is at the origin of the data flow, we consider each request as the mapping source. We can say that the mappings in the model are always in the form of one request to a query set rm !Qn. The four mapping patterns are illustrated as follows.

1. Deterministic Mapping:

This type of mapping is the most common and perfectly matched pattern.Web request rm appears in all traffic with the SQL queries set Qn. If in any session in the testing phase with the request rm, there is absence of a query set Qn matching the request, it indicates a possible intrusion.

2. Empty Query Set:

In some cases, the SQL query set can be the empty set. That means, the web request neither caused nor generated any database queries. For instance, when a web request for retrieving an image GIF file from the same webserver is done, a mapping relationship does not exist because only the web requests are noticed.

3. No Matched Request:

In this case, these queries can't match up with any web requests, and we place these unmatched queries in a set NMR. During the testing phase, any query within set NMR is observed as legitimate. The size of NMR depends on webserver logic, but it is particularly small.

4. Nondeterministic Mapping:

In this case, each time that the same type of web request arrives, it always matches up with one (and only one) of the query sets in the pool. It is quite difficult to identify traffic that matches this pattern. This happens only within dynamic websites.

For data deduplication, three units established in our system, that is, users, private cloud provider (S-CSP) secure cloud services in the public cloud. The S-CSP reach duplication by checking whether the contents of the two files are the same, and keep only one of them. The right of access to a file to describe a number of privileges (token) is based. Each file is Related some files denoting tokens on the specified permissions. A user calculates and sends test signals to the public cloud duplicate check for unauthorized duplication.

Users have access to the private cloud and public cloud servers. The S-CSP provides outsourcing services data and stores the data on behalf of users. To reduce storage costs, the S-CSP eliminate redundant storage through deduplication data and keeps only unique data. A user is a person who outsource data storage, want the S-CSP and access the data later. Each file is protected with the convergent fundamental key and privilege to perform deduplication approved differential charges. The private cloud managed private key to the privileges responding to requests for tokens of user files. Which provides the user interface to authorize private cloud files and requests.

When the user at this time means to upload and download files from cloud storage user first request to the Web server to load the file only authorized users will be presented in the Web server for this purpose, use the test algorithm Send property, if you load the file is divided into blocks, i.e., the block size of 4 KB by default. After file size block occurs. Each block contains ciphertext own, unique identification chips and a private key. After deduplication Detection occurs. The data storage server that contains all uploaded files and DB Profiler store all file metadata.

Case 1: When all the data file F1 and F2 are different store database in different blocks it. Case 2: If the F1 = F2 file, only one file is stored in the database to avoid duplication of data. Case 3: When F1 € F2 then compare the blocks of data memory and only different memory blocks will be both file database.

C. Workflow for File Upload /Download:

Authorized users can download the file from the cloud storage. Only privileged user access have the authority to uploading and downloading the file from the cloud. It contains the actual flow file with multiple operations.

System Workflow:

Get the Token.

Upload / Download the File [after token generation].

Forward the request for upload to the web service API.

Web service API will connect to the Private cloud and validate request.

Response from the private cloud, for verification [Boolean response].

Data will give to the Security Component which will responsible for encryption / description. It will have its own private data base for storing keys and other information.

Actual Encrypted data will be store onto the repository.

D. Token Generation:

The token generation algorithm uses the token to generate enemy clearly identify blocks and keep going in the right file locking sequence at the time of downloading the file.

Input: File as input

1. Web browser client request token to private cloud

2. Web services validate token

3. Return token to web server

4. Web client got token

Output: Generate token

Novelty Journals

ISSN 2394-7314

International Journal of Novel Research in Computer Science and Software Engineering
Vol. 2, Issue 3, pp: (38-46), Month: September-December 2015, Available at: www.noveltyjournals.com

4. MATHEMATICAL MODEL

Fig.6 shows mathematical model of our system.

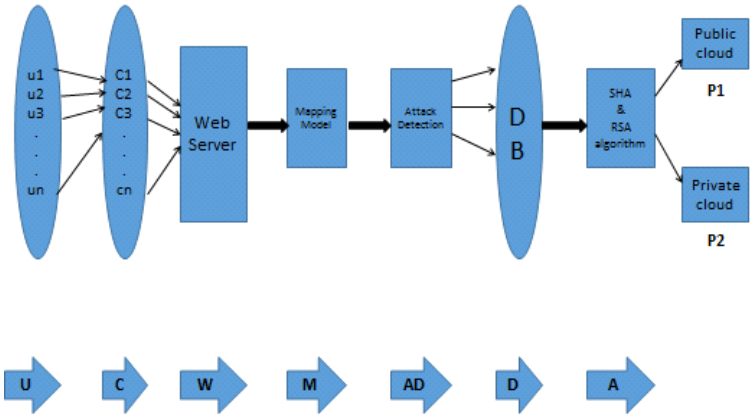

Fig 6: Mathematical Model

Let S be a system such that, S={U,C,W,M,AD,A,P1,P2 }

where,

U = user i.e. U={u1,u2,u3.....un}

C = container provided to each user i.e. C={c1,c2,c3....cn}

W =Web Server to which each session having container is assigned .

M = mapping model used to detect the abnormal behaviour

AD = Attack detected by mapping model

D = Database

A = algorithms for encryption and decryption(SHA and RSA)

P1 = Public cloud for request and response for files.

P2 = Private cloud for key management.

5. CONCUSION

In this paper, we have proposed efficient IDS system that models the network behaviour in multi-tiered web application and builds casual mapping model for identifying various types of attacks and minimize the false positives in web application. We achieved this with the help of double guard with lightweight virtualization (isolated session using session ID) and enhances the security in web application. We presented an intrusion detection system that builds models of normal behaviour for multitier web applications from both front-end web requests and back-end database queries. Unlike previous approaches that correlated or summarized alerts generated by independent IDSs, Double guard detects the intruder into multitier web application. It also proposed secure deduplication with the help of token generation and secure upload download we can assure the user about high data security and also avoid data deduplication.

REFERENCES

[1] Openvz, http://wiki.openvz.org, 2011.

[2] Vishwesh.N, Rakesh Kumar.D, Anil Kumar.M, Mamatha.G, “Generation of Meta Alerts by Aggregating Intrusion Alerts”,IOSR Journal of Computer Engineering (IOSR-JCE),Volume 8, Issue 5 (Jan. - Feb. 2013), PP 63-69.

[3] Christopher Kruegel, Giovanni Vigna, **"**Anomaly Detection of Webbased Attacks**",** October 27–31, 2003.

[4] Bryan Parno, Jonathan M. McCune, Dan Wendlandt, David G. Andersen, Adrian Perrig, "CLAMP: Practical Prevention of Large-Scale Data Leaks."

[5] S.SHALIN, S.USHA,"Prevention Of Cross-Site Scripting Attacks (XSS) On Web Applications In The Client Side",IJCSI International Journal of Computer Science Issues , Vol. 8, Issue 4, No 1, July 2011.

[6] Aparna Ajit Patil, Asst. Prof. Dhanashree Kulkarni, "Block Level Data Duplication on Hybrid Cloud Storage System",International Journal of Advanced Research in Computer Science and Software Engineering , Volume 5, Issue 8, August 2015.

[7] Shweta Pochhi, Vanita Babanne,A Survey on Secure and Authorized Data Deduplication,International Journal of Science and Research (IJSR) 2012.

[8] N.O.Agrawal, Prof Mr. S.S.Kulkarni, "Secure Deduplication And Data Security With Efficient And Reliable CEKM"International Journal of Application or Innovation in Engineering & Management (IJAIEM), Volume 3, Issue 11, November 2014.

[9] Ankita Mahajan 1, Ratnaraj Kumar, "Secure Method For Authorized Deduplication And Data Dynamics In Cloud Computing",International Journal of Computer Engineering and Applications , Volume IX, Issue VII, July 2015.

[10] Meixing Le, Angelos Stavrou, Member, IEEE, and Brent ByungHoon Kang, Member, IEEE, "DoubleGuard: Detecting Intrusions in Multitier Web Applications"ieee transactions on dependable and secure computing, vol. 9, no. 4, march 2014.

[11] Jin Li, Xiaofeng Chen, Mingqiang Li, Jingwei Li, Patrick P.C. Lee, and Wenjing Lou, "Secure Deduplication with Efficient and Reliable Convergent Key Management"ieee transactions on parallel and distributed systems, vol. 25, no. 6, june 2014.