Data Leakage Detection in Cloud Computing Network

Utkarsh[1], Vaibhav Gupta[2], Yashwant Kumar Singh[3], Dr. Shashi Kant Singh[4]
[1,2,3]Student, CSE Department, Galgotias College of Engineering and Technology, Gr. Noida, India
[4]Professor, CSE Department, Galgotias College of Engineering and Technology, Gr. Noida, India

Abstract: *Giving security in a conveyed framework requires more than client verification with passwords or advanced testaments and classification in information transmission. Dispersed model of cloud makes it defenseless and inclined to modern disseminated interruption assaults like Distributed Denial of Service (DDOS) and Cross Site Scripting (XSS). To deal with extensive scale arrange get to movement and regulatory control of information and application in cloud, another multi-strung disseminated cloud IDS demonstrate has been proposed. Our proposed cloud IDS handles huge stream of information bundles, break down them and create reports effectively by incorporating learning and conduct investigation to recognize interruptions.*

Index Terms— *Cross Site Scripting, Cloud, DDOS, IDS.*

I. INTRODUCTION

The term cloud is analogical to "Web". The term distributed computing depends on cloud drawings utilized as a part of the past to speak to phone systems and later to portray web in.

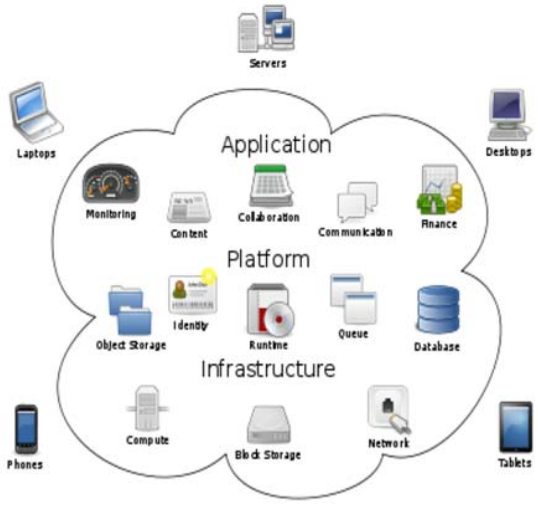

Figure 1 : Cloud Computing

Cloud processing is web based figuring where virtual shared servers give programming, framework, stage, gadgets and different assets and facilitating to client as an administration on pay-as you-utilize premise. Figure 1.demonstrates the idea [7]. All the data that a digitized framework brings to the table is given as an administration in the distributed computing model. Clients can get to these administrations accessible on the "web cloud" without having any past know-how on dealing with the assets included. Cloud clients don't claim the physical foundation; rather they lease the utilization from an outsider supplier. They devour assets as an administration and pay just for assets that they utilize. What they just need is a PC and web association. Distributed computing has reformed the IT world with its administrations provisioning framework, less support cost, information and administrations accessibility confirmation, quick openness and versatility. Distributed computing has three fundamental deliberation layers i.e. framework layer (which is a virtual machine reflection of a server), the stage layer (a virtualized working arrangement of a server) and application layer (that incorporates web applications) [1]. Equipment layer is excluded as it doesn't straightforwardly offer to clients. Distributed computing additionally has three administration models to be specific Platform as a Service (PaaS), Infrastructure as a Service (IaaS) and Software as a Service (SaaS) models. PaaS show encourages clients by giving stage on which applications can be created and run. IaaS convey administrations to clients by keeping up expansive frameworks like facilitating servers, overseeing systems and different assets for customers. SaaS demonstrate makes client straightforward of introducing and running programming administrations all alone machines. By and by, Salesforce.com, Google and Amazon are the main cloud specialist organizations who broaden their administrations for capacity, application and calculation on pay according to utilize premise. Information, application and administrations non-accessibility can be forced through Denial of Service (DOS) or Distributed Denial of Service (DDOS) assaults and both cloud specialist organization and clients progress toward becoming impediment to give or get cloud administrations [2]. For such sort of assaults Intrusion Detection System (IDS) can be emplaced as a solid guarded instrument. IDSs are host-based, arrange based and circulated IDSs. Have based IDS (HIDS) screens particular host machines, arrange based IDS (NIDS) recognizes interruptions on key system focuses and

circulated IDS (DIDS) works both on host and additionally organize. IDSs create cautions for the managers which depend on genuine positives or genuine alerts when really interruption happens and false positive or false cautions in the event of a wrong recognition by the framework. IDSs can distinguish interruption designs by basically examining the system bundles, applying marks (pre-characterized manages) and creating cautions for framework overseers. IDS utilizes two technique for recognition i.e. peculiarity discovery, that chips away at client conduct designs and suspicious conduct. Other technique is abuse discovery that can identify through prestigious assault designs and coordinating an arrangement of characterized guidelines or assault against framework vulnerabilities through port examining [3]. Since Cloud foundation has tremendous system activity, the customary IDSs are not sufficiently productive to deal with such a substantial information stream. Most known IDSs are single strung and because of rich dataset stream, there is a need of multi-strung IDS in Cloud figuring condition. In a conventional system, IDS screens, recognizes and alarm the authoritative client for system activity by conveying IDS on key system stifle focuses on client site. Yet, in Cloud arrange IDS must be set at Cloud server site and totally directed and oversaw by the specialist organization. In this situation, if an assailant figures out how to enter and harm or take user‟s information, the cloud client won't be advised specifically. The interruption information would just be imparted through the specialist co-op and client needs to depend on him. The cloud specialist co-op dislike to illuminate the client about the misfortune and can conceal the data for his picture and notoriety. In such a case, an unbiased outsider checking administration can guarantee sufficient observing and alarming for cloud client. In this report, we have proposed a productive multi-strung cloud IDS, managed and observed by an outsider ID checking administration, who can give ready reports to cloud client and master guidance for cloud specialist co-op. Keeping in mind the end goal to determine the issues which customary IDSs can not resolve, a productive and solid conveyed Cloud IDS model is proposed.

II. Literature Review

A. Analysis

In nowadays a solitary server handles the different solicitations from the client. Here the server needs to prepare the every one of the solicitations from the clients at the same time, so the handling time will be high. This may prompts loss of information and bundles might be deferred and undermined. On doing this the server can't prepare the question from the client in an appropriate way. So the preparing time gets expanded. It might prompts activity and clog. To defeat these issues we are going for the idea called distributed computing. In this distributed computing we will execute the Proxy server to evade these issues. Be that as it may, in this framework Data Efficiency is enhanced yet not the information security. At whatever point we talk about information productivity we ought to talk about information security likewise, in light of the fact that in the distributed computing we don‟t know from which cloud the information is coming, so in the current framework there is no framework to discover the information security. The framework in light of the new design has better adaptability and adaptation to non-critical failure. A bunch comprises of a solitary server and various intermediary servers and is gotten to by numerous customers. Intermediary servers stores information on neighborhood circles and read or compose information determined by a server. The server keeps up the record for all document put away in various intermediaries. At the point when a customer needs to download a few information, it will first send a demand to the Server and the Server then divert the demand to a comparing intermediary that have the required information and thus the information will be sent to the customer. With the blend of Cloud and Grid figuring ideas, the information demand can be proficiently overhauled in an opportune way. The real piece of the Project is Security, so previously mentioned stage talks about Cloud and Grid Technology, however not about security. The Security execution is accomplished by two stage, to be specific- Behavioral – Knowledge.

BehavioralAnalysis : Utilizing this strategy, we have to perceive expected conduct (honest to goodness utilize) or an extreme conduct deviation. The system must be accurately prepared to productively recognize interruptions. For a given interruption test set, the system figures out how to distinguish the interruptions. Be that as it may, we concentrate on distinguishing client behavioral examples and deviations from such examples. With this system, we can cover a more extensive scope of obscure assaults.

Knowledge Analysis : Utilizing a specialist framework, we can portray a pernicious conduct with a run the show. One preferred standpoint of utilizing this sort of interruption recognition is that we can include new standards without adjusting existing ones. Interruption recognition (ID) is a sort of security administration framework for PCs and systems. An ID framework assembles and breaks down data from different zones inside a PC or a system to recognize conceivable security ruptures, which incorporate both interruptions (assaults from outside the association) and abuse (assaults from inside the association). ID utilizes defenselessness appraisal (once in a while alluded to as

filtering), which is an innovation created to survey the security of a PC framework or system. Interruption discovery capacities include:

- Monitoring and analyzing both user and system activities.
- Analyzing system configurations and vulnerabilities.
- Assessing system and file integrity.

B. Related Existing Techniques

1.Intrusion detection for grid and cloud computing

Cloud and Grid registering are the most helpless focuses for intruder's assaults because of their dispersed condition. For such situations, Intrusion Detection System (IDS) can be utilized to improve the safety efforts by a precise examination of logs, designs and system activity. Conventional IDSs are not appropriate for cloud condition as system based IDSs (NIDS) can't identify encoded hub correspondence, likewise have based IDSs (HIDS) are not ready to locate the shrouded assault trail. Kleber, schulter et al. [5] have proposed an IDS benefit at cloud middleware layer, which has a review framework intended to cover assaults that NIDS and HIDS can't identify. The engineering of IDS administration incorporates the hub, benefit, occasion inspector and capacity. The hub contains assets that are gotten to through middleware which characterizes get to control arrangements. The administration encourages correspondence through middleware. The occasion inspector screens and catches the system information, additionally dissects which govern/strategy is broken. The capacity holds conduct based (correlation of late client activities to normal conduct) and learning based (known trails of past assaults) databases. The inspected information is sent to IDS benefit center, which examines the information and caution to be an interruption. The creators have tried their IDS model with the assistance of reproduction and discovered its execution palatable for constant usage in a cloud situation. In spite of the fact that they have not talked about the security approaches consistence check for cloud specialist co-op and their revealing techniques to cloud clients.

2.Intrusion detection in the cloud

Interruption discovery framework assumes an imperative part in the security and steadiness of dynamic protection framework against interloper unfriendly assaults for any business and IT association. IDS usage in distributed computing requires an effective, adaptable and virtualization-based approach. In distributed computing, client information and application is facilitated on cloud benefit provider's remote servers and cloud client has a restricted control over its information and assets. In such case, the organization of IDS in cloud turns into the obligation of cloud supplier. Despite the fact that the overseer of cloud IDS ought to be the client and not the supplier of cloud administrations. In the paper [1], Roschke and Cheng et al. have proposed a joining answer for focal IDS administration that can consolidate and coordinate different prestigious IDS sensors yield investigates a solitary interface. The interruption discovery message trade arrange (IDMEF) standard has been utilized for correspondence between various IDS sensors. The creators have proposed the arrangement of IDS sensors on particular cloud layers like application layer, framework layer and stage layer. Cautions created are sent to „Event Gatherer" program. Occasion gatherer gets and change over ready messages in IDMEF standard and stores in occasion information base archive with the assistance of Sender, Receiver and Handler modules. The examination segment breaks down complex assaults and displays it to client through IDS administration framework. The creators have proposed a viable cloud IDS administration design, which could be observed and managed by the cloud client. They have given a focal IDS administration framework in view of various sensors utilizing IDMEF standard for correspondence and checked by cloud client.

C. Security Issues in Cloud Computing:

Security threats can be categorized as follow [4];

1. Cloud data confidentiality issue

Secrecy of information over cloud is one of the glaring security concerns. Encryption of information should be possible with the customary methods. Be that as it may, encoded information can be secured from a pernicious client however the protection of information even from the chairman of information at administration provider's end couldn't be covered up. Seeking and ordering on scrambled information remains a state of worry all things considered. Previously mentioned cloud security issues are a couple and dynamicity of cloud engineering are confronting new difficulties with quick execution of new administration worldview.

2. Network and host based attacks on remote Server

Host and system interruption assaults on remote hypervisors are a noteworthy security worry, as cloud sellers utilize virtual machine innovation. DOS and DDOS assaults are propelled to refuse assistance accessibility to end clients.

3.Cloud security auditing

Cloud reviewing is a troublesome assignment to check consistence of all the security strategies by the seller. Cloud specialist co-op has the control of touchy client information and procedures, so a computerized or

outsider evaluating instrument for information uprightness check and legal investigation is required. Protection of information from outsider reviewer is another worry of cloud security.

4. Lack of data interoperability standards

It comes about into cloud client information secure state. On the off chance that a cloud client needs to move to other specialist organization because of specific reasons it would not have the capacity to do as such, as cloud user's information and application may not be perfect with different vendor's information stockpiling configuration or stage. Security and secrecy of information would be in the hands of cloud specialist co-op and cloud client would be reliant on a solitary specialist organization.

III. PROPOSED WORK

Distributedcomputing gives application and capacity benefits on remote servers. The customers don't need to stress over its support and programming or equipment up-degrees. Cloud show chips away at the "concept of virtualization" of assets, where a hypervisor server in cloud server farm has various customers on one physical machine. Conveying HIDS in hypervisor or host machine would permit the chairman to screen the hypervisor and virtual machines on that hypervisor. Yet, with the fast stream of high volume of information as in cloud display, there would be issues of execution like over-burdening of VM facilitating IDS and dropping of information parcels. Likewise if host is bargained by a culpable assault the HIDS utilized on that host would be killed. In such a situation, a system based IDS would be more appropriate for sending in cloud like foundation. NIDS would be put outside the VM servers on jug neck of system focuses, for example, switch, switch or door for system activity observing to have a worldwide perspective of the framework. Such NIDS would at present be confronting the issue of extensive measure of information through system get to rate in cloud condition. To make it safer for the user and organization which are using cloud network we can use following encryption techniques. In today's world there are so many threat while transferring a sensitive data over the internet. Along these lines, in our proposed work we will attempt to consolidate both the strategy for information spillage that is interruption discovery and interruption avoidance. There are many encryption and decryption algorithm used now-a-day like RC4 and Advanced Encryption Standard (AES).RC4 algorithm has a property of generating a **keystream.** It hasone of the demerit that is "One in each 256 keys can be a frail key. These keys are distinguished by cryptanalysis that can discover conditions under which one of more created bytes are emphatically related with a couple of bytes of the key."

The above demerit leads us to use AES algorithm for encryption and decryption of data.AES is an iterative rather than Feistel cipher. It is based on 'substitution–permutation network'. It comprises of a series of linked operations, some of which involve replacing inputs by specific outputs (substitutions) and others involve shuffling bits around (permutations).

Interestingly, AES performs all its computations on bytes rather than bits. Hence, AES treats the 128 bits of a plaintext block as 16 bytes. These 16 bytes are arranged in four columns and four rows for processing as a matrix.

Unlike DES, the number of rounds in AES is variable and depends on the length of the key. AES uses 10 rounds for 128-bit keys, 12 rounds for 192-bit keys and 14 rounds for 256-bit keys. Each of these rounds uses a different 128-bit round key, which is calculated from the original AES key.

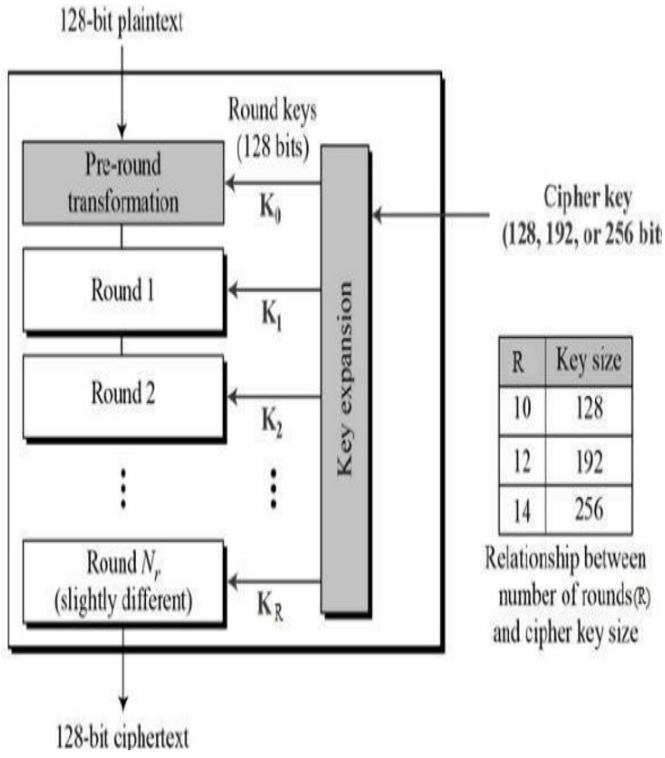

Figure2. Schematic of AES structure

Encryption using AES

Here, we restrict to description of a typical round of AES encryption. Each round comprise of four sub-processes. The first round process is depicted below −

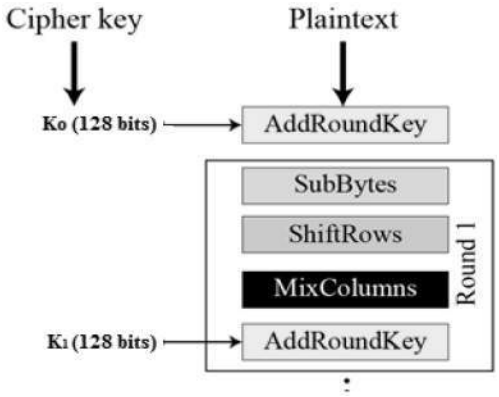

Figure3. AES Encryption

1.Byte Substitution (SubBytes)
The 16 input bytes are substituted by looking up a fixed table (S-box) given in design. The result is in a matrix of four rows and four columns.
2.Shiftrows
Each of the four rows of the matrix is shifted to the left. Any entries that 'fall off' are re-inserted on the right side of row. Shift is carried out as follows −

- First row is not shifted.
- Second row is shifted one (byte) position to the left.
- Third row is shifted two positions to the left.
- Fourth row is shifted three positions to the left.

The result is a new matrix consisting of the same 16 bytes but shifted with respect to each other.
3. MixColumns
Each column of four bytes is now transformed using a special mathematical function. This function takes as input the four bytes of one column and outputs four completely new bytes, which replace the original column. The result is another new matrix consisting of 16 new bytes. It should be noted that this step is not performed in the last round.
4. Addroundkey
The 16 bytes of the matrix are now considered as 128 bits and are XORed to the 128 bits of the round key. If this is the last round then the output is the ciphertext. Otherwise, the resulting 128 bits are interpreted as 16 bytes and we begin another similar round.

Decryption using AES
The process of decryption of an AES ciphertext is similar to the encryption process in the reverse order. Each round consists of the four processes conducted in the reverse order −

- Add round key
- Mix columns
- Shift rows
- Byte substitution

Since sub-processes in each round are in reverse manner, unlike for a Feistel Cipher, the encryption and decryption algorithm needs to be separately implemented, although they are very closely related.
Basically, the proposed model will be beneficial for an organization where everyone is connected over the same LAN. Our model will have 4 modules.
Module 1 is for registration of all the employees after which admin has approve all the registered employees along with their respective IP Address and Mac Address which would be stored in the database.
Module 2 is for uploading and sharing the file over the cloud server. The uploaded file needs to be encrypted before sharing or uploading over the cloud server.
Module 3 is for downloading the uploaded file where the user needs to decrypt the file before downloading it. If the IP Address and Mac Address of the user are different from the registered detail then an alert email would be sent to the admin.
Module 4 is for admin who can control all the process. This module consist of four sub parts which are - approved user, rejected user, intruders detail, list of files.

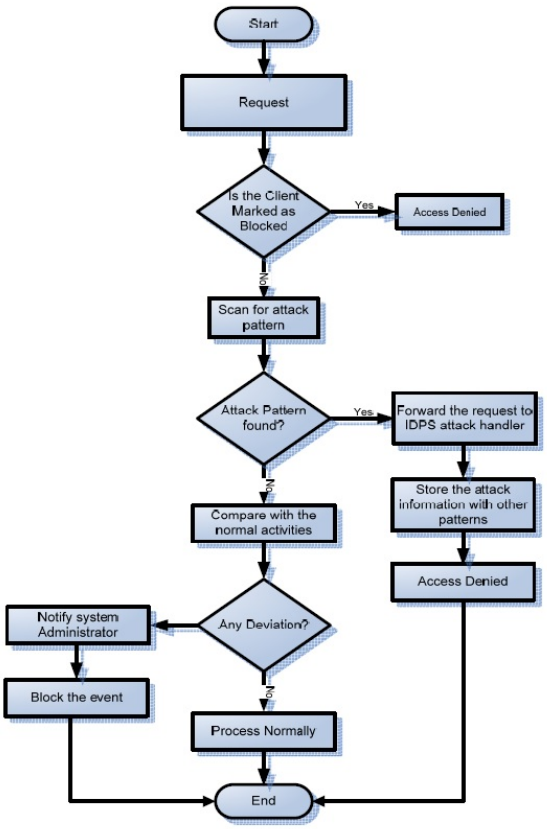

Figure4. IDS Framework

IV. CONCLUSION

Distributed computing is a "system of systems" over the Web, along these lines odds of interruption is more with

the knowledge of intruder's assaults. Distinctive IDS systems are
used to counter noxious assaults in conventional systems. For Cloud figuring, huge system get to rate, giving up the control of information and applications to benefit supplier and dispersed assaults weakness, a proficient, dependable and data straightforward IDS is required. In this report, the leakage of data would be prevented if any intruder try to intrude. In contrary, an alert email will be sent to the admin acknowledging him about the intruder and admin would take the appropriate action.

ACKNOWLEDGMENT

The making of the paper required co-operation and direction of various individuals. We in this manner think of it as our prime obligation to thank every one of the individuals who had helped us for making it effective. It is our massive delight to express our appreciation to Dr. Shashi Kant Singh (Professor - Computer Science and Engineering) as a guide who gave us useful and positive criticism amid the planning of this paper. To wrap things up, we are grateful to our companions and facilitator whose consolation and recommendation helped us to finish our venture and report. We are likewise grateful to our parents.

REFERENCES

[1] SEBASTIAN ROSCHKE, FENG CHENG, CHRISTOPH MEINEL," INTRUSION DETECTION IN THE CLOUD", EIGHTH IEEE INTERNATIONAL CONFERENCE ON DEPENDABLE, AUTONOMIC AND SECURE COMPUTING, 2009.

[2] CHI-CHUN LO, CHUN-CHIEH HUANG, JOY KU, "A COOPERATIVE INTRUSION DETECTION SYSTEM FRAMEWORK FOR CLOUD COMPUTING NETWORKS", 39TH INTERNATIONAL CONFERENCE ON PARALLEL PROCESSING WORKSHOPS, 2010.

[3] ANDREAS HAEBERLEN," AN EFFICIENT INTRUSION DETECTION MODEL BASED ON FAST INDUCTIVE LEARNING", SIXTH INTERNATIONAL CONFERENCE ON MACHINE LEARNING AND CYBERNETICS, HONG KONG, 19-22 AUGUST 2007.

[4] RICHARD CHOW, PHILIPPE GOLLE, MARKUS JAKOBSSON, "CONTROLLING DATA IN THE CLOUD: OUTSOURCING COMPUTATION WITHOUT OUTSOURCING CONTROL", ACM COMPUTER AND COMMUNICATIONS SECURITY WORKSHOP, CCSW 09, NOVEMBER 13, 2009.

[5] KLEBER, SCHULTER, "INTRUSION DETECTION FOR GRID AND CLOUD COMPUTING", IEEE JOURNAL: IT PROFESSIONAL, 19 JULY 2010.

[6] IRFAN GUL, M. HUSSAIN, "DISTRIBUTED CLOUD INTRUSION DETECTION MODEL", INTERNATIONAL JOURNAL OF ADVANCED SCIENCE AND TECHNOLOGY VOL. 34, SEPTEMBER, 2011.