Data Protection Analysis in Cloud Computing

Vallabhaneni Pranay, Vahiduddin Shariff, J. Swapna Priya, Chandrasekhar Rao Chennupati

Dept. of CSE, Sir C R Reddy College of Engineering, Eluru, AP, India
Dept. of IT, Vignan’s Lara Institute of Technology, Vadlamudi, AP, India
Dept. of MCA, Bapatla Engineering College, Bapatla, AP, India

Abstract
Cloud computing is basically an Internet-based network made up of large numbers of servers - mostly based on open standards, modular and inexpensive. Clouds contain vast amounts of information and provide a variety of services to large numbers of peoples. The benefits of cloud computing are Reduced Data Leakage, Decrease evidence acquisition time, they eliminate or reduce service downtime, they Forensic readiness, they Decrease evidence transfer time the main factor to be discussed is security of cloud computing, which is a risk factor involved in major computing fields. This paper also attempts to explain the drawbacks in conventional system designs, which results in low performance due to network congestion and less data efficiency. We consider cloud and grid computing systems to improve the performance of the system. Cloud systems are characterized by a main server and other connected servers which provide certain services. Cloud systems, especially public cloud systems are prone to intrusions and care must be taken to secure the system. The emphasis in this thesis is to make cloud systems secure using intrusion detection system. Intrusion detection can be performed using either behaviour based or knowledge based techniques or both.

Keywords
Cloud Computing, Data Protection Analysis, Intrusion Detection, Cloud Architecture

I. Introduction
Cloud Computing is becoming one of the next industry buzz words. Cloud computing builds upon advance of research in virtualization, distributed computing, grid computing and utility computing. Cloud computing is a collection of all sources to enable resource sharing in terms of scalable infrastructures, middleware and application development platforms, and value-added business applications. In the past-three decades, the world of computation has changed from centralized (client-server not web-based) to distributed systems.

A. What is the Cloud?
Cloud computing is receiving a great deal of attention, both in publications and among users, from individuals at home to the U.S. government. Yet it is not always clearly defined.1 Cloud computing is a subscription-based service where you can obtain networked storage space and computer resources. One way to think of cloud computing is to consider your web browser, go to the email client, and log in. The most important part of the equation is having internet access. Your email is not housed on your physical computer; you access it through an internet connection, and you can access it anywhere. If you are on a trip, at work, or down the street getting coffee, you can check your email as long as you have access to the internet. Your email is different than software installed on your computer, such as a word processing program. When you create a document using word processing software, that document stays on the device you used to make it unless you physically move it.

B. Cloud Computing Shares Characteristics
Autonomic computing — Computer systems capable of self-management.
Client–server model — Client–server computing refers broadly to any distributed application that distinguishes between service providers (servers) and service requesters (clients).
Grid computing — “A form of distributed and parallel computing, whereby a ‘super and virtual computer’ is composed of a cluster of networked, loosely coupled computers acting in concert to perform very large tasks.”
Mainframe computer — Powerful computers used mainly by large organizations for critical applications, typically bulk data processing such as census, industry and consumer statistics, enterprise resource planning, and financial transaction processing.
Utility computing — The ‘packaging of computing resources, such as computation and storage, as a metered service similar to a traditional public utility, such as electricity.
Peer-to-peer — Distributed architecture without the need for central coordination, with participants being at the same time both suppliers and consumers of resources (in contrast to the traditional client–server model).

C. Types of Clouds
There are different types of clouds that you can subscribe to depending on your needs. As a home user or small business owner, you will most likely use public cloud services.

1. Public Cloud - A public cloud can be accessed by any subscriber with an internet connection and access to the cloud space.
2. Private Cloud - A private cloud is established for a specific group or organization and limits access to just that group.
3. Community Cloud - A community cloud is shared among two or
more organizations that have similar cloud requirements.

4. **Hybrid Cloud** - A hybrid cloud is essentially a combination of at least two clouds, where the clouds included are a mixture of public, private, or community.

**D. Services**

There are three types of cloud providers that you can subscribe to: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). These three types differ in the amount of control that you have over your information, and conversely, how much you can expect your provider to do for you. Briefly, here is what you can expect from each type.

1. **Software as a Service**

A SaaS provider gives subscribers access to both resources and applications. SaaS makes it unnecessary for you to have a physical copy of software to install on your devices. SaaS also makes it easier to have the same software on all of your devices at once by accessing it on the cloud. In a SaaS agreement, you have the least control over the cloud.

2. **Platform as a Service**

A PaaS system goes a level above the Software as a Service setup. A PaaS provider gives subscribers access to the components that they require to develop and operate applications over the internet.

3. **Infrastructure as a Service**

An IaaS agreement, as the name states, deals primarily with computational infrastructure. In an IaaS agreement, the subscriber completely outsources the storage and resources, such as hardware and software that they need.

**II. Architecture for the Created System**

**A. Problems**

Suppose that two users, Client1 and Client2 use the “conventional client-server” system and both are simultaneously sending multiple requests to the same server. In this case, acknowledgment time is increasing from the server to users, due to heavy traffic in the network. If the acknowledgment time is increasing, perhaps data will be lost. The second problem is, multiple users are requesting the same data at the same time in server. So server might not give equal performance for all the systems. The third problem is, there is no security mechanism. Security mechanism is the main topic in this thesis.

**B. Solution**

The main cloud server is connected to all the proxy servers and it maintains the index of the data information by all the proxy servers. At first, the user sends the request to the main cloud server about the “hi.doc” file. Then the main cloud server will forward the request to the exact proxy. The proxy server gives back the information to the main cloud server. Finally, the proxy server will send the file to the user. In this case the main cloud server can handle multiple user requests. The proxy server helps to decrease the acknowledgment time and gives the reply to the users in a proper manner. A proxy server can be the other way around: a server which receives and queues requests, makes format changes and packet divisions etc for information from just one single information carrying server. The network administrator cannot monitor all the clients, so it will depend on the IDS which gives the alerts to the remaining “cloud server”. In this case we can hide the data from the intruders. In cloud computing, there is more possibility to hack the data. So for preventing data from hackers. We IDs are auditing engines, so models of auditing system can describe their architecture. The director or analysis engine, may be centralized or distributed or may be hierarchical or fragmented. Information may be gathered from hosts, from network, from both or from other directors. When an intrusion occurs some response is appropriate. If the intrusion attempt is detected before the attack is successful, the system can take the action to prevent the attack from succeeding. An IDS inspects all inbound and outbound network activities and identifies suspicious patterns that may indicate a network or system attacks from someone attempting to break into or compromise a system. Processes make the difference in the realm of computation. On one hand, an individual has full control on data and processes in his/her computer. On the other hand, we have the cloud computing wherein, the service and data maintenance is provided by some vendor which leaves the client/customer unaware of where the processes are running or where the data is stored. So, logically speaking, the client has no control over it. The cloud computing uses the internet as the communication media. The Cloud computing system can be easily threatened by various attacks, because most of the cloud computing systems provide service to so many people who are not proven to be trustworthy. Due to their distributed nature, cloud computing environment are easy targets for intruders looking for possible vulnerabilities to exploit. Cloud computing have two approaches i.e. Knowledge-based IDS and Behavior-Based IDS to detect intrusions in cloud computing.

Behavior-based intrusion detection techniques assume that an intrusion can be detected by observing a deviation from normal or expected behavior of the system or the users. The model of normal or valid behavior is extracted from reference information collected by various means. The intrusion detection system later compares this model with the current activity. When a deviation is observed, an alarm is generated. In other words, anything that does not correspond to a previously learned behavior is considered intrusive. Therefore, the intrusion detection system might be complete (i.e. all attacks should be caught), but its accuracy is a difficult issue (i.e. you get a lot of false alarms). Knowledge-based intrusion detection techniques apply the knowledge accumulated about specific attacks and system vulnerabilities. The intrusion detection system contains information about these vulnerabilities and looks for attempts to exploit these vulnerabilities. When such an attempt is detected, an alarm is triggered. In other words, any action that is not explicitly recognized as an attack is considered acceptable. In this paper we proposed a system to detect intrusions in cloud computing environment. Intrusions are detected with the
help of Knowledge-based and behavior-based approaches where knowledge-based approach uses signatures (evidences of attacks) and behavior-base approach uses set of normal activities to check abnormality providing a security mechanism called IDS. An IDS is placed in the proxy server.

III. Cloud Computing Architecture

![Cloud Computing Architecture Diagram](image)

The client sends a request to the main server and it forwards the request to all the proxy servers to process. All the proxy servers give acknowledgement to the main server and it replies to the client. This is working procedure for grid computing. Let us consider an example; initially the client triggers a request at time “n”, the server accepts the request at time “n+1” and it equally divides the task (15000 iterations) and assign them to proxy servers. Let us consider in this case there are three proxy servers and each will process 5000 iterations parallel and get back to the main server at time “n+10”. After getting processed output from all the proxy server’s main server provides output to the client at the time “n+12”. If the same experiment is carried out in conventional “client-server” system the client will trigger the request at time “n+10”. Afterwards main server starts processing the task (15000 iterations) and it get backs client at the time “n+31”. Hence, grid system has processed the task (15000 iterations) at time n+12 seconds, where as conventional “client-server” system has processed the task (15000 iterations) at n+31 seconds. Therefore, for comparing these two systems we can say, the grid computing has less computation time than the conventional “client-server”.

IV. Intrusion Detection System Architecture

Before reading this chapter we have to clarify one thing that, we are intending to protect the “cloud system”. Refer to the “Cloud core technologies” chapter for information about vulnerabilities in cloud computing and how IDS can prevent data from intruders in the cloud system. The below explanation is all about the IDS parts and IDS functions. The elements that are participating in the architecture are nodes, service, event auditor and storage devices. Anode has resources which are accessed homogeneously through middleware. The middleware sets the access-control policies and supports a service-oriented environment. In the environment a service provides its functionalities via middleware, which facilitates communication. In the architecture, the event auditor plays an important role in the system. Initially, it captures the information from different sources such as the log system, nodes, and services. After capturing the data the IDS service analyzes the data based on the intrusion detection techniques such as knowledge based and behavior based. If an intrusion is detected in the system then the IDS service uses middleware communication mechanisms for sending alerts to the other nodes. The middleware synchronizes the knowledge based and behavior based databases.

The storage service holds data that should be analyzed by the IDS service. Because all nodes have to access the same data in the environment. So, the middleware must transparently create a virtualization in the homogenous environment.

![Architecture of grid and cloud computing intrusion detection](image)

The client sends a request to the server for getting a service. In the “conventional client server” system, the client communicates with the end server directly, due to which traffic congestion or data loss etc might take place. So to overcome this issue we have implemented a proxy server, which extends the functionality of the main cloud server and is the mediator between your web browser and the end server. Initially, your web browser sends a request to the proxy server, after which the proxy server forwards the request to the end server. The end server then gives acknowledgment to the proxy server. Finally, the proxy server replies to the browser. Therefore, there is no direct communication between the user and the end server. So, HTTP request is originated from the intermediate proxy server. As a result the client computer’s IP address will be in hidden state and illegitimate users cannot access the client computer’s IP address. This type of proxy server is also known as anonymous proxy server.

Let us assume that user wants to get a file from the main cloud server. Initially, the user sends a service request to the proxy server. Here, the proxy server checks the service request by using filtering rules based on the traffic such as IP address or protocols. If the request is validated by the filter then it forwards the service request to the main cloud server.

The optional choice for the proxy servers are that it can change the request to all the servers equally and it transparently shares the traffic. The proxy server then forwards the request to the main cloud server.

The client requests the server for getting a file from the main cloud server. The client sends a request to the server for getting a service. In the “conventional client server” system, the client communicates with the end server directly, due to which traffic congestion or data loss etc might take place. So to overcome this issue we have implemented a proxy server, which extends the functionality of the main cloud server and is the mediator between your web browser and the end server. Initially, your web browser sends a request to the proxy server, after which the proxy server forwards the request to the end server. The end server then gives acknowledgment to the proxy server. Finally, the proxy server replies to the browser. Therefore, there is no direct communication between the user and the end server. So, HTTP request is originated from the intermediate proxy server. As a result the client computer’s IP address will be in hidden state and illegitimate users cannot access the client computer’s IP address. This type of proxy server is also known as anonymous proxy server.

Let us assume that user wants to get a file from the main cloud server. Initially, the user sends a service request to the proxy server. Here, the proxy server checks the service request by using filtering rules based on the traffic such as IP address or protocols. If the request is validated by the filter then it forwards the service request to the main cloud server.

The optional choice for the proxy servers are that it can change the request to all the servers equally and it transparently shares the traffic. The proxy server then forwards the request to the main cloud server.

A client communicates to the end server by using internet. So in this case a proxy server can be installed in several points anywhere between the client and the end server.

Reverse proxies: Client sends a request to the reverse proxy by using internet and then the reverse proxy forwards the request to the web server in an internal network.

Load-balancing: The reverse proxy server shares the traffic to
various web servers. In this case every web server has its individual application area. As a result, the reverse proxy server may have to alter the URL’s in each web page. By that we can achieve load balancing in the network. Intrusion detection system service helps to increase the security in the cloud system by using two methods i.e. behavior based and knowledge based service. We refer to the previous section for further information. The audited data is sent to the IDS service core, which analyzes the behavior using artificial intelligence to detect deviations. It has two sub systems namely analyzer and alert systems. Analyzer system the analyzer gets audit data and examines whether a heuristic in the database is being broken, after which it sends the outcome to the IDS service. For these outcomes, IDS estimates the attack probability and if probability ratio is high then it alerts the other nodes. Alert system in the cloud system if any one node is harmed by the intruder then alert system will alert the remaining nodes in the network regarding the attack. Storage service is a database. When the node receives a request or an acknowledgment, then the analyzer system will compare the node information in the storage service. Storage service has two services namely behavior and knowledge service. Knowledge service we used audit information for the communication system and the logging system for evaluating the knowledge service. Moreover, we are free to delete and modify rules at will in a knowledge service. Behavior service it compares recent user actions to the usual behavior. It is divided into two types i.e. user behavior and node behavior. User behavior is nothing but analyzing the user’s behavior. By using this method we can identify expected behavior or a severe behavior deviation. Event auditor has two components for detecting an intrusion in the network as data is exchanged between the nodes and environment states. In the first component, when the data is exchanged between the nodes, audit information about the communication between the nodes is being captured. Therefore, audit data captures only the node information, but not network data.

A. Features
- Highly virtualization and standardized infrastructures
- Massive scalability
- Fault tolerant & high reliable
- Intra and inter cloud load balance
- Instant application deployment

B. Benefits
- No need to install or update any software and hardware
- Unlimited use
- Always on
- Access from any where
- Many services to choose from

Disadvantages of cloud computing:

1. Features might be limited:
This situation is bound to change, but today many web-based applications simply are not as full-featured as their desktop-based applications. For example, you can do a lot more with Microsoft PowerPoint than with Google Presentation’s web-based offering.

2. Requires a Constant Internet Connection
Computing is impossible if you cannot connect to the Internet. Since you use the Internet to connect to both your applications and documents, if you do not have an Internet connection you cannot access anything, even your own documents.

3. Can Be Slow
Even with a fast connection, web-based applications can sometimes be slower than accessing a similar software program on your desktop PC. Everything about the program, from the interface to the current document, has to be sent back and forth from your computer to the computers in the cloud.

4. Does Not Work Well With Low-Speed Connections
Similarly, a low-speed Internet connection, such as that found with dial-up services, makes cloud computing painful at best and often impossible. Web-based applications require a lot of bandwidth to download, as do large documents. If you are laboring with a low-speed dial-up connection, it might take seemingly forever just to change from page to page in a document, let alone to launch a feature-rich cloud service. In other words, cloud computing is not for the broadband-impaired!

V. Future Enchancement
In the future, we’ll implement our IDS, helping to improve green (energy-efficient), white (using wireless networks), and cognitive (using cognitive networks) cloud computing environments. We also intend to research and improve cloud computing security.

A. Security
The information housed on the cloud is often seen as valuable to individuals with malicious intent. There is a lot of personal information and potentially secure data that people store on their computers, and this information is now being transferred to the cloud. This makes it critical for you to understand the security measures that your cloud provider has in place, and it is equally important to take personal precautions to secure your data. The first thing you must look into is the security measures that your cloud provider already has in place. These vary from provider to provider and among the various types of clouds. What encryption methods do the providers have in place? What methods of protection do they have in place for the actual hardware that your data will be stored on? Will they have backups of my data? Do they have firewalls set up? If you have a community cloud, what barriers are in place to keep your information separate from other companies? Many cloud providers have standard terms and conditions that may answer these questions, but the home user will probably have little negotiation room in their cloud contract. A small business user may have slightly more room to discuss the terms of their contract with the provider and will be able to ask these questions during that time. There are many questions that you can ask, but it is important to choose a cloud provider that considers the security of your data as a major concern. No matter how careful you are with your personal data, by subscribing to the cloud you will be giving up some control to an external source. This distance between you and the physical location of your data creates a barrier. It may also create more space for a third party to access your information. However, to take advantage of the benefits of the cloud, you will have to knowingly give up direct control of your data. On the contrary, keep in mind that most cloud providers will have a great deal of knowledge on how to keep your data safe. A provider likely has more resources and expertise than the average user to secure their computers and networks.
technique to overcome the drawbacks in the existing cloud and grid system. The individual analysis performed in each node reduces the complexity and the volume of data in comparison to previous solutions, where the audit data is concentrated in single points. To summarize, the cloud provides many options for the everyday computer user as well as large and small businesses. It opens up the world of computing to a broader range of uses and increases the ease of use by giving access through any internet connection. However, with this increased ease also come drawbacks. You have less control over who has access to your information and little to no knowledge of where it is stored. You also must be aware of the security risks of having data stored on the cloud. The cloud is a big target for malicious

References


Vallabhaneni Pranav completed his M.Tech in CSE in University college of Engineering, JNTUK. He is currently working as an Asst.Professor in CSE department, Sir C R Reddy college of engineering, Eluru, West Godavari district. He is having 5 months of teaching experience.

Vahiduddin Shariff completed his M.Tech in CSE in Vignan University. He is working as Asst. Professor in CSE Department, Sir C R Reddy College of Engineering, Eluru, West Godavari district. He is having 2 year 6 months of teaching Experience.

J. Swapna Priya completed her M.Tech in CS in JNTU Ananthapur. She is working as Asst. Professor in IT Department, Vignan’ Lara Institute of Technology, Vadlamudi. She is having 4 years of teaching Experience.

Chennupati ChandraSekhar Rao completed his M.Tech in CSE in Vignan University. He is working as lecturer in MCA Department, Bapatla Engineering college, Bapatla. He is having 1 year 8 months of teaching Experience.