ISSN: 2277-9655
Impact Factor: 3.785

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

AVOID DATA MINING BASED ATTACKS IN RAIN-CLOUD

D.Sudhagar*

* Assistant Professor, Department of Information Technology, Jerusalem Collage of Engineering, Chennai, India.

DOI: 10.5281/zenodo.51435

ABSTRACT

Cloud computing is an umbrella term used to refer to Internet based development and services. It offers users the ability to connect to computing resources and access IT managed services with a previously unknown level of ease. Due to this greater level of flexibility, the cloud has become the breeding ground of a new generation of products and services. However, the flexibility of cloud-based services comes with the risk of the security and privacy of users' data. Thus, security concerns among users of the cloud have become a major barrier to the widespread growth of cloud computing. One of the security concerns of cloud is data mining based privacy attacks that involve analyzing data over a long period to extract valuable information. In this paper, we first identify the data mining based privacy risks on cloud data and propose a distributed architecture to eliminate the risks in Rain-cloud environment.

KEYWORDS: Cloud computing, Security, Data mining based attacks, Rain-cloud environment.

INTRODUCTION

Cloud computing is a computing paradigm, where a large pool of systems are connected in private or public networks, to provide dynamically scalable infrastructure for application, data and file storage. With the advent of this technology, the cost of computation, application hosting, content storage and delivery is reduced significantly.

Cloud computing is a practical approach to experience direct cost benefits and it has the potential to transform a data center from a capital-intensive set up to a variable priced environment[1].

The idea of cloud computing is based on a very fundamental principal of reusability of IT capabilities. The difference that cloud computing brings compared to traditional concepts of "grid computing", "distributed computing", "utility computing", or "autonomic computing" is to broaden horizons across organizational boundaries.

Cloud Computing Components:

The cloud computing model consists of five characteristics, three delivery models, and four deployment models. The five key characteristics of cloud computing are: location-independent resource pooling, on-demand self-service, rapid elasticity, broad network access, and measured service [2]. These five characteristics represent the first layer in the cloud environment architecture. (Fig.1)

Five Characteristics:

1) On-demand self service provides automatic computing capability management to systems, without requiring human interaction.

2) Broad network access allows heterogeneous clients. Such as mobile phones, laptops to connect to Cloud systems over the network

3) Resources pooling in Cloud systems is available as pooling resources for multiple consumer which is able to dynamically assign and reassign according to consumer demand.

ISSN: 2277-9655
Impact Factor: 3.785

4) Rapid elasticity offers rapidly and elastically provision of capabilities. We can grow and shrink our capacity very quickly in minutes or hours.
5) Measure service provides monitoring, controlling & reporting of resources usage.

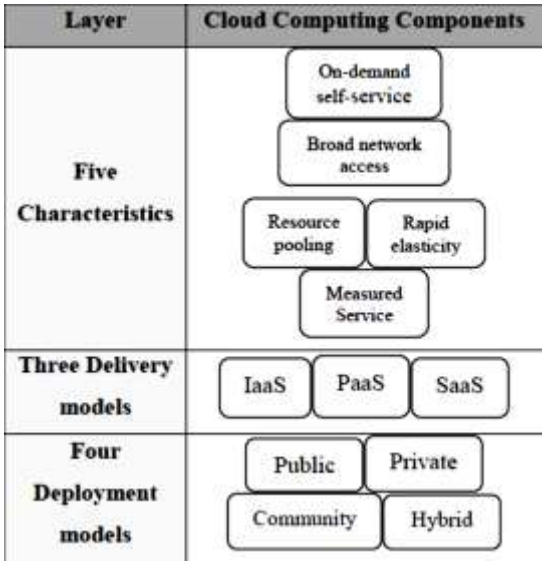

Fig.1 Cloud Environment Architecture.

Cloud delivery models:
Cloud Providers offer services that can be grouped into three categories [3][4].

Infrastructure as a Service (IaaS)
Infrastructure as a Service is a provision model in which an organization outsources the equipment used to support operations, including storage, hardware, servers and networking components. The service provider owns the equipment and is responsible for housing, running and maintaining it. The client typically pays on a per-use basis. Some common examples are Amazon, GoGrid, 3 Tera, etc.

Platform as a Service (PaaS)
Platform as a Service (PaaS) is a way to rent hardware, operating systems, storage and network capacity over the Internet. The service delivery model allows the customer to rent virtualized servers and associated services for running existing applications or developing and testing new ones. PaaS providers offer a predefined combination of OS and application servers, such as LAMP platform (Linux, Apache, MySql and PHP), restricted J2EE, Ruby etc. Google's App Engine, Force.com, etc are some of the popular PaaS examples.

Software as a Service (SaaS)
Software as a Service (SaaS) is a software distribution model in which applications are hosted by a vendor or service provider and made available to customers over a network, typically the Internet.SaaS is becoming an increasingly prevalent delivery model as underlying technologies that support Web services and service-oriented architecture (SOA). Today SaaS is offered by companies such as Google, Salesforce, Microsoft, Zoho, etc.

Cloud deployment models:
In cloud computing there are three types of cloud deployment take place – public cloud, private cloud, hybrid cloud and hybrid cloud.

1. Public cloud or External cloud: It allows users access to the cloud via interfaces using web browsers. Users need to pay for the time duration they use service, i.e., pay-per-use.
2. Private cloud or Internal cloud: Its operation is within an organization's internal enterprise data center. The main advantage here is that it is easier to manage security, maintenance and upgrades and also provides more control over the deployment and use.

3. Hybrid cloud or Mixed cloud: It is a combination of public cloud and private cloud. In this model a private cloud is linked to one or more external cloud services.
4. Community cloud or Group cloud: When many organizations jointly construct and share a cloud infrastructure, their requirements and policies then such a cloud model is called as a community cloud.

LITERATURE REVIEW

Rain Cloud

The Rain Cloud System is a collection of several clouds (Fig.2). These clouds collectively form a group of excess resources like rain drops and reduce the drought or lack of resources in a network [13]. There could be several problems occurred in private clouds such as lacking of hardware and software resources, network, congestion of packets, data become bottleneck etc.

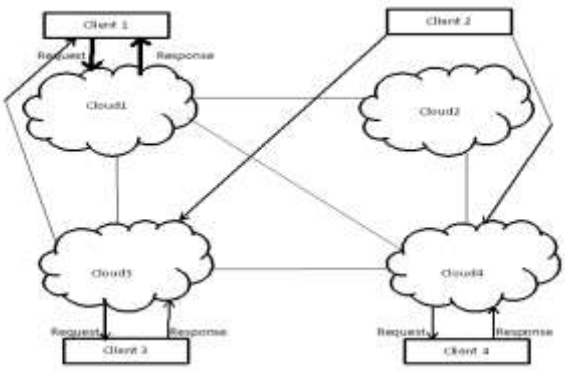

Fig.2 Rain Cloud System

These problems always either slow down the network, loss of packets or information. To resolve such kind of problems "multi-clouds" or "rain clouds" are introduced. These clouds provide access to resources in emergency situation like pouring rain drops to clients or users. Characteristics of Rain-Clouds are

- Rapid Elasticity
- Broad network Access
- Rapid Connectivity within clouds.

The rain cloud architecture always try to maintain the network congestion free because in small organizations the load of data traffic rapidly increases that causes severe problems.

DATA MINING ON CLOUD

Data mining is one of the fastest growing fields in computer industry that deals with discovering patterns from large data sets. It is a part of knowledge discovery process and is used to extract human understandable information. Mining is preferably used for a large amount of data and related algorithms often require large data sets to create quality models [6][7].

The relationship between data mining and cloud is worth to discuss. Cloud providers use data mining to provide clients a better service. If clients are unaware of the information being collected, ethical issues like privacy and individuality are violated [9][10]. This can be a serious data privacy issue if the cloud providers misuse the information. Again attackers outside cloud providers having unauthorized access to the cloud, also have the opportunity to mine cloud data.

In both cases, attackers can use cheap and raw computing power provided by cloud computing to mine data and thus acquire useful information from data. As cloud is a massive source of centralized data, data mining gives attackers a great advantage in extracting valuable information and thus violating clients' data privacy [8].

Data mining based attacks:

Cloud computing is becoming one of the most enticing technologies, because of its cost-efficiency and flexibility. Various security issues in the cloud are impeding the vision of cloud computing as a new IT procurement model.

Our research focus is to provide a solution for the threats that are the major issue for anyone when they want to adopt cloud services for their work. For this purpose, a framework should be designed for execution of data and information securely in cloud environment [9]. It will protect users' data, messages, information against various attacks. Some of the most common attacks are as follows.

- Tampering: Attacker may alter the information stored in the database.
- Eavesdropping: Attacker gains access over the data path.
- Viruses and Worms: These are piece of code that decreases the performance of hardware and application.

RESULTS AND DISCUSSION

System Architecture

In this section we discuss our proposed system architecture that avoid data mining based privacy attacks on the cloud [13]. Our system (Fig.3) consists of two major components: Cloud Data Distributor and Cloud Providers.

The Cloud Data Distributor receives data in the form of files from clients, splits each file into chunks and distributes these chunks among cloud providers. Cloud Providers store chunks and responds to chunk requests by providing the chunks.

In the rain-cloud system single data distributer cannot provide the complex data and data mining based information.

The above system architecture is that a single data distributor can create a bottleneck in the system as it can be the single point of failure.

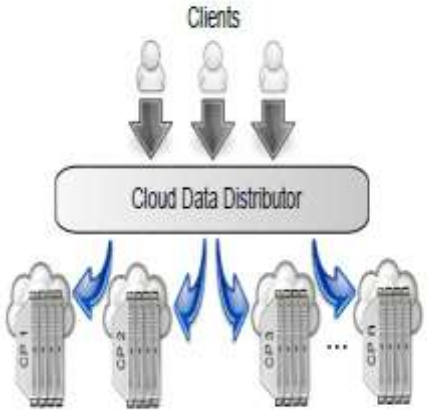

Fig.3 System Architecture

To eliminate this, multiple distributors of cloud data can be introduced. In case of multiple data distributors, for each client, a specific distributor will act as the primary distributor that will upload data, whereas other distributors will act as secondary distributors who can perform the data retrieval operations.

The following figureshows the extended system architecture with multiple distributors of data in rain-cloud environment.

ISSN: 2277-9655
Impact Factor: 3.785

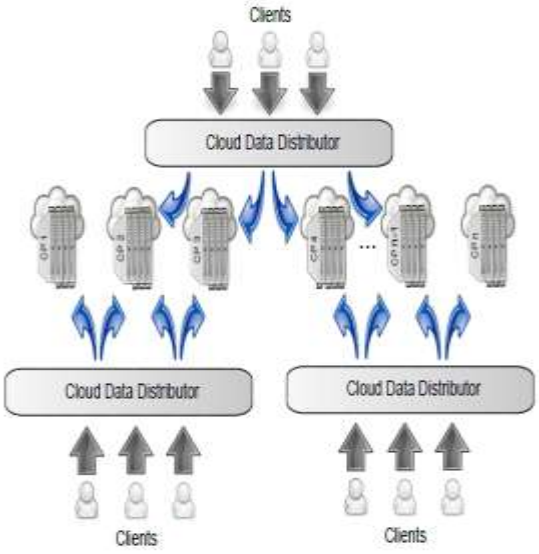

Fig.4 Extended System Architecture

This proposed architecture is to use the multiple data distributers and multiple data providers used in rain cloud environment (Fig.4). This method is used to avoid the data mining based attacks in rain-cloud environment.

CONCLUSION

Ensuring security of cloud data is still a challenging problem. Cloud service providers as well as other third parties use different data mining techniques to acquire valuable information from user data hosted on the cloud. In this paper, we have discussed the impact of data mining on cloud and have proposed a distributed structure to eliminate mining based privacy threat on cloud data. Our approach combining categorization, fragmentation and distribution, avoids data mining by maintaining privacy levels, splitting data into chunks and storing these chunks of data to appropriate cloud providers.

Although the proposed system provides an effective way to protect privacy from mining based attacks, it introduces performance overhead when client needs to access all data frequently, e.g. client needs to perform a global data analysis on all data. The analysis may have to access data from multiple locations, with a degraded performance. In future, we look forward to improve our system by reducing such overhead.

REFERENCES

1. Bharath K.Samanthula, Yousef Elmehdwi, Gerry Howser, Sanjay Madrian,"A secure data sharing and query processing framework via federation of cloud computing", Department of Computer Science, Missouri University of Science and Technology, 500 West 15th Street, Rolla, MO65401, United States, 2013
2. Mark D. Ryan, "Cloud computing security: The scientific challenge, and a survey of solutions", The Journal of Systems and Software 86 (2013) 2263– 2268
3. C. Clifton and D. Marks. Security and privacy implications of data mining. In ACM SIGMOD Workshop, pages 15–19, 1996.
4. M. Bramer. Principles of Data Mining. Springer, 2007.
5. M. Kantardzic. Data Mining: Concepts, Models, Methods and Algorithms. John Wiley & Sons, Inc., 2002.
6. Cloud Security Alliance, “Top threats to cloud computing”, *Cloud Security Alliance*, March 2010.
7. I. Kotenko, M. Stepashkin, and E. Doynikova, “Security analysis of information systems taking into account social engineering attacks”, *IEEE 19th International Eurimicro Conference on Parallel, Distributed, and Network-Based Processing*, 2011.
8. Clavister, "Security in the cloud", Clavister WhitePaper, 2008.

9. G. Brunette and R. Mogull (eds), "Security guidance for critical areas of focus in cloudcomputing", CloudSecurityAlliance, 2009.
10. S. Subashini and V. Kavitha, "A survey on securityissues in service delivery models of cloudcomputing", Journal of Network and ComputerApplications, 34(1), 2011, pp 1-11.
11. H. Takabi, J.B.D. Joshi and G.-J. Ahn, "Security and Privacy Challenges in Cloud ComputingEnvironments", IEEE Security & Privacy, 8(6),2010, pp. 24-31.
12. Sangdo Lee, Hyoungyill Park, Yongtae Shin, "Cloud computing availability: multi-clouds for big data service". Convergence and Hybrid Information Technology Communications in Computer and Information Science Volume 310, 2012, pp 799-806.