ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 4, April 2014

Mitigating Data Mining Attack in Cloud

A.Raja Rajeswari, R.Sakkaravarthi

PG student, Department of Computer Science and Engineering, Muthayammal Engineering College, Namakkal, India

Professor, Department of Computer Science and Engineering, Muthayammal Engineering College, Namakkal, India

ABSTRACT: Cloud computing refers to the delivery of computing resources that is used in excess of the network. As an alternative of maintaining personal data on the own hard drive or updating important applications for user needs, user can use a service over the network, to a different location, to store user information and / or use its applications. This also provides flexibility so it is very useful in a new generation of services and products. One of the main security problems in cloud is data mining based privacy attacks that involve analyzing data over a long period to extract valuable information. It gives the outside attackers and providers having unconstitutional access to the cloud and a prospect of analyze the client information over a extensive period of time to extract the sensitive information that causes privacy violation of clients. This is a big problem in many clients of cloud. In this paper, we identify the data mining based privacy risks on cloud data and propose a distributed architecture to remove the privacy risks.

KEYWORDS: Cloud computing; Security; Data mining

I. INTRODUCTION

Cloud computing is a model for well-located, on-demand network access to a shared pool of configurable computing resources (e.g., applications, servers, networks, services and storage) that can be rapidly provisioned and released with least service provider or management effort interaction. These cloud models promote availability and is composed of three service models, five essential characteristics and four deployment models [1].

Cloud computing is the deliverance of computing services over the Internet. A cloud service allows an individuals and businesses to use software and hardware that are managed by third party at remote locations. The cloud computing model allows users to access computer resources and information from anywhere that a network connection is available. An example of cloud service includes social networking sites, online file storage, webmail, and online business applications. Cloud computing is also provides a common pool of resources and data, includes data & information storage room, computer system processing power and dedicated corporate system and user needed applications.

Cloud services are typically made available via a private, public, community clouds or hybrid clouds. Cloud services are popular because all persons can access their own Emails, social networks (google+, Facebook) or photo editing service from anywhere anytime in the world, at minimum charge or no charges.

Some cloud service providers may provide the individual information of users for advertising purpose and to learn more about the users for other reasons. Individuals should pay careful attention to whether and how the cloud company protects their own individual information. Also users are supposed to care for their own personal information or data by using any privacy settings that the service may offer. Due to the strong commercial nature of cloud services like technical, non-technical and industrial aspects are involved in cloud provisions. Since all the above areas still contain major spaces or gaps so, the recommendations are not restricted to purely technical issues, but also cover non- technical aspects related in particular to the economical and legalistic side of cloud systems [2].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 4, April 2014

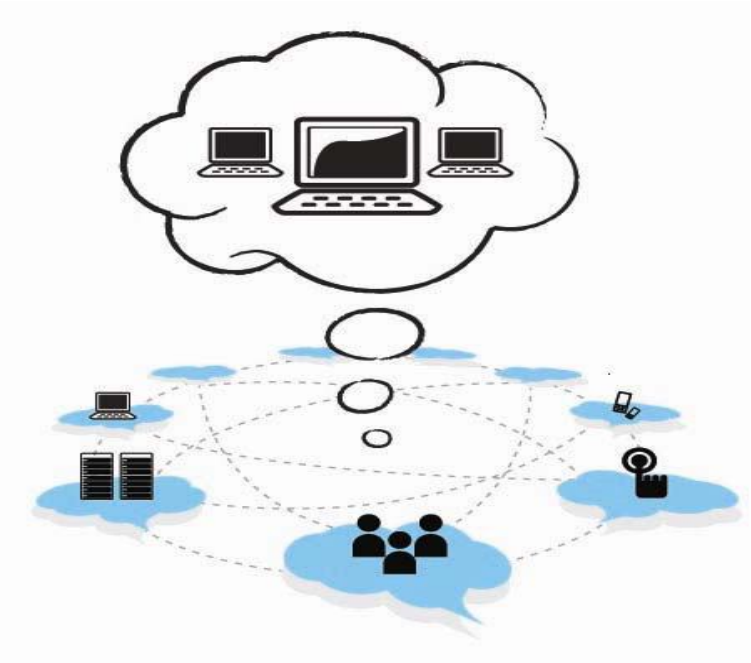

Fig.1. Cloud computing structural design

A. *Advantages of using Cloud:*

- Hardware costs is low
- Data Security
- Less power consumption
- Ease of replacement or repair
- Noise is also very less

II. BACKGROUND OF DATA MINING ACCESS IN CLOUD

Data mining is one of the fastest growing fields in computer industry [3] that deals with discovering patterns from large data sets [4]. It is a part of knowledge discovery process and is used to extract human understandable information [5]. Data mining is a relatively young field with many issues that still need to be researched in depth, any off-the-shelf data mining system products and domain specific data mining application software's are available. As a young research field, data mining has made broad and significant progress since its early beginnings in the 1980s. Today, data mining is used in a vast array of areas, and numerous commercial data mining systems are available.

III. EXISTING APPROACH

The existing cloud system is vulnerable when data remains under a single cloud provider. This provides advantages to the attackers as they have fixed targets in the forms of cloud providers. In network outage the data can loss for the cloud provider like malware attack etc. If the attacker be able to attack the particular client means, he is aims to a fixed cloud service providers tries to access to the user's data and analyzes the user's data. This is very easiest job for the attackers. Given that the entire data belong to a client stay under a single cloud provider, both inside and outside attackers gets the profit of using data mining to a large level. Inside attackers in this context refers to malicious workers at a cloud provider. Data mining model often want large number of observations and single provider structural design is a great benefit suiting the case as all the samples remain under the provider. Thus single provider architecture is the biggest security threat regarding in cloud based data mining.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 4, April 2014

In single cloud provider can having the following main issues:

- Less Security
- Loss of data
- No privacy
- Cost of maintenance is high

IV. PROPOSED WORK

To reduce or remove the above all disadvantages of data storage of a client to the single provider, data can be split into chunks and distributed among multiple cloud providers. This distributed system can be providing visualized system an attacker chooses a specific client. But the distribution of data obliges him to target for multiple cloud providers, making his job increasingly difficult. Data mining based attacks on cloud involves attackers on two categories. First one is malicious employees inside provider and another one is outside attackers.

Distribution of data chunks among multiple providers is stored in one table called chunk table. The entries of this table contain information regarding data chunks. These information include the virtual id, privacy level (PL), Cloud Provider Table index of the current cloud provider storing the chunk (CP), Cloud Provider Table index of the snapshot provider (SP) (if any), set of positions of misleading data bytes (M) (if any) for all chunks [6]. This distribution of data chunks among multiple providers restricts a cloud provider from accessing all chunks of a client. Even if the cloud provider performs mining on chunks provided to the provider, the extracted information leftovers unfinished. Once more, mining of data in distributed system is challenging [7]. Specially correlating data from various sources is unwieldy [8] and often leads to unsuccessful and ineffective mining. Aggarwal et al. [9] demonstrates partitioning of data over multiple databases in such a way to ensure that the exposure of the contents of some database does not result in a violation of privacy. The distributed architecture for cloud redefines the partitioning of data in terms of preserving privacy from mining based attacks.

A. *Network Construction in distributed system:*

We construct a network to overcome the Security problem. In the cloud, nodes are exchange data or services directly to each and every node. All nodes are connected to all other available nodes. Path details are also maintained in the cloud servers. All node information's are maintained in the cloud system. Any one of the node wants to connect to this network signup is needed. Node details and path details are maintained in the server system.

B. *File fragmentation:*

We are splitting and store the files in cloud network i.e., different system. A file lookup is guaranteed to succeed if and only if the file is present in the system. A file lookup terminates in a small and bounded number of hops. The files are uniformly distributed among all active nodes. The system handles dynamic node joins and leaves. Store the file in different system in confidential manner for secure purpose, i.e. unauthorized user did not view the file.

C. *Location, Routing & Inference guards:*

We used three different types of guards for maintaining the files. There are Location guard, Routing guard and Inference guards. Location guard is a location key, which is used as a key to the location of a file in the network this key is used to retrieve the corresponding file from the network. The routing guard is secure to locate a file in the network given its location key such that neither the key nor the location is revealed to an adversary. Inference guard is to protect the system from traffic analysis-based inference attacks, such as file replica, user IP Address, file size inference attack and lookup frequency inference attack.

V. CONCLUSION

Security of cloud data is the main problem for maintaining important information's. Cloud computing offers some benefits for individuals and organizations. There are also works based on privacy and security concerns. Cloud service providers and other third parties use different data mining techniques to acquire valuable information from user data hosted on the cloud. In this paper, we proposed a distributed structure approach that combining distribution, fragmentation and prevent data mining by maintaining privacy levels, splitting data into chunks and storing these chunks of data to appropriate cloud providers.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 4, April 2014

VI. ACKNOWLEDGEMENT

I am, A.RAJA RAJESWARI (rajarajeswaribalaji@gmail.com) acknowledged that this work has been done as a part of my thesis (project) work in M.E., CSE from Muthayammal Engineering College, Tamilnadu, India and it is supported by my guide Dr. R. SAKKARAVARTHI, Ph.D., Professor & Project coordinator of CSE, Muthayammal Engineering College, Tamilnadu, India.

REFERENCES

1. NIST Cloud's definition, 15th version- http://csrc.nist.gov/groups/SNS/cloudcomputing/
2. Keith Jeffery and Burkhard Neidecker-Lutz, "The Future of Cloud Computing Opportunities for European Cloud Computing Beyond 2010," European Communities, Information Society and Media, 2010.
3. M.Kantardzic, "Data Mining - Concepts, Models, Methods and Algorithms," John Wiley & Sons, Inc., 2002.
4. M. J. Shaw, C. Subramaniam, G. W. Tan, and M. E. Welge, "Knowledge management and data mining for marketing," Decision Support System, 31(1), pp. 127–137, 2001.
5. M. Bramer, "Principles of Data Mining," Springer, 2007.
6. Himel Dev, Tanmoy Sen, Madhusudan Basak and Mohammed Eunus Ali, "An Approach to Protect the Privacy of Cloud Data from Data Mining Based Attacks," Data Cloud -2012.
7. G. M. Weiss, "Data mining in the real world - Experiences, challenges and recommendations," In DMIN, pp. 124–130, 2009.
8. Q. Yang and X. Wu, "Challenging problems in data mining research," International Journal of Information Technology and Decision Making, 5(4), pp.597–604, 2006.
9. G. Aggarwal, M. Bawa, P. Ganesan, H. Garcia-molina, K. Kenthapadi, R. Motwani, U. Srivastava, D. Thomas, and Y. Xu, "Two can keep a secret : A distributed architecture for secure database services," In Proc. CIDR, 2005.