Chapter 21
A Framework for Protecting Users' Privacy in Cloud

Adesina S. Sodiya
Federal University of Agriculture, Nigeria

Adegbuyi B.
Federal University of Agriculture, Nigeria

ABSTRACT

Data and document privacy concerns are increasingly important in the online world. In Cloud Computing, the story is the same, as the secure processing of personal data represents a huge challenge. The main focus is to preserve and protect personally identifiable information (PII) of individuals, customers, businesses, governments and organisations. The current use of anonymization techniques is not quite efficient because of its failure to use the structure of the datasets under consideration and inability to use a metric that balances the usefulness of information with privacy preservation. In this work, an adaptive lossy decomposition algorithm was developed for preserving privacy in cloud computing. The algorithm uses the foreign key associations to determine the generalizations possible for any attribute in the database. It generates penalties for each obscured attribute when sharing and proposes an optimal decomposition of the relation. Postgraduate database of Federal University of Agriculture, Abeokuta, Nigeria and Adult database provided at the UCIrvine Machine Learning Repository were used for the evaluation. The result shows a system that could be used to improve privacy in cloud computing.

1. INTRODUCTION

Cloud computing is a model for enabling ubiquitous (independent of device and location), convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction (NIST, 2011).

Cloud computing provides massive computation power and storage capacity which enable users to deploy applications without infrastructure investment. Many privacy-sensitive applications like health services are built on cloud for economic benefits and operational convenience. Correspondingly, privacy

DOI: 10.4018/978-1-5225-7113-1.ch021

protection has become a critical issue and one of most concerning sharing of information by an individual, a business, a government agency, or other entity via the cloud, lead to issues of confidentiality.

Problems associated with Cloud Computing stem from loss of control, lack of trust and multi-tenancy (Ranchal et al., 2010). These problems exist mainly in third party management models. The customer's loss of control is due to the fact that cloud providers solely host data, applications and resources.

User identity data, access control rules and security policies are stored, managed and enforced by cloud providers. Consumers rely on the providers to ensure data security and privacy. Trusting a third party requires taking the risk of assuming that the trusted third party will act as it is expected (which may not be true all the time). Tenants (cloud service users) share a pool of resources and may have opposing goals. If tenants cannot be trusted, they need to be isolated with some level of guarantee. In a third party managed model, service providers (e.g., Google and Amazon) manage and control various aspects of the cloud.

Privacy concerns exist wherever personally identifiable information (information that can be used to uniquely identify, contact, or locate a single person) is collected and stored – in digital form or otherwise. Improper or non-existent disclosure control can be the root cause for privacy issues. Data privacy issues can arise in response to information from a wide range of sources, such as:

- Healthcare records
- Criminal justice investigations and proceedings
- Financial institutions and transactions
- Biological traits, such as genetic material
- Residence and geographic records
- Ethnicity
- Privacy breach

The challenge in data privacy is to share data while protecting Personally Identifiable Information (PII). The fields of data security and information security design and utilize software, hardware and human resources to address this issue.

Quite a number of applications try to eliminate the issue of disclosure by deleting fields or attributes that uniquely identify individuals. Unfortunately, this measure is not sufficient as a combination of a number of attributes that can be used to identify a group of persons if not a particular individual. These attributes are known as quasi-identifiers.

Its definition encompasses any information about an individual who can be identified directly from the information, or whose identity can be reasonably ascertained by reference to other information. Information does not have to be true, written down, sensitive or 'important' to be personal information. A consumer has to disclose his PII to use a cloud service. It becomes even more complex when cloud service providers use services from other providers to provide a service. Since there may be a chain, tracking the distribution of PII may not be simple task. The concept of transferring sensitive data to another company is also a concern. Privacy and security can only be as good as its weakest link. Cloud computing can increase the risk that a security breach may occur. Consequently the majority of cloud providers – including Amazon's Simple Storage Service (S3), the Google Compute Engine and the Citrix Cloud Platform - do not guarantee specific levels of security and privacy in their service level agreements (SLAs) as part of the contractual terms and conditions between cloud providers and consumers (Gholami and Laure, 2015). This shows that there is still the need to develop strategies, methods and technologies for enhancing security and privacy in the cloud.