

Chapter 34

Classification of File Data Based on Confidentiality in Cloud Computing Using K–NN Classifier

Munwar Ali Zardari

Universiti Teknologi Petronas, Malaysia

Low Tang Jung

Universiti Teknologi Petronas, Malaysia

ABSTRACT

Cloud computing is a new paradigm model that offers different services to its customers. The increasing number of users for cloud services i.e. software, platform or infrastructure is one of the major reasons for security threats for customers' data. Some major security issues are highlighted in data storage service in the literature. Data of thousands of users are stored on a single centralized place where the possibility of data threat is high. There are many techniques discussed in the literature to keep data secure in the cloud, such as data encryption, private cloud and multiple clouds concepts. Data encryption is used to encrypt the data or change the format of the data into the unreadable format that unauthorized users could not understand even if they succeed to get access of the data. Data encryption is very expensive technique, it takes time to encrypt and decrypt the data. Deciding the security approach for data security without understanding the security needs of the data is a technically not a valid approach. It is a basic requirement that one should understand the security level of data before applying data encryption security approach. To discover the data security level of the data, the authors used machine learning approach in the cloud. In this paper, a data classification approach is proposed for the cloud and is implemented in a virtual machine named as Master Virtual Machine (Vmm). Other Vms are the slave virtual machines which will receive from Vmm the classified information for further processing in cloud. In this study the authors used three (3) virtual machines, one master Vmm and two slaves Vms. The master Vmm is responsible for finding the classes of the data based on its confidentiality level. The data is classified into two classes, confidential (sensitive) and non-confidential (non-sensitive/public) data using K-NN

DOI: 10.4018/978-1-5225-8176-5.ch034

algorithm. After classification phase, the security phase (encryption phase) shall encrypt only the confidential (sensitive) data. The confidentiality based data classification is using K-NN in cloud virtual environment as the method to encrypt efficiently the only confidential data. The proposed approach is efficient and memory space friendly and these are the major findings of this work.

INTRODUCTION

Cloud Computing is an internet based distributed virtual environment. All computational operations are performed on cloud through the Internet (Rawat 2012). It consists of a set of resources and services offered through the internet. Cloud computing is also called Internet computing because they both have same symbolic icon. Applications, Operating systems, data, processing capacity, and storage all exist on the Web, ready to be shared among the users (Sadiku et al., 2014). Cloud computing is basically a collection of different e-resources available twenty four hours and accessible from anywhere through browser software. Many companies are getting benefits from cloud computing due to its pay-as-you-go cost model and elasticity of resources, where users pay for only those services that they used (Prakash, 2013), and cloud provides customizable services to users. Compared to the traditional models which provide in-house infrastructure, cloud provides low cost services with high availability (Li et al., 2010). In cloud model the user is exempted from hardware and software maintenance cost.

The National Institute of Science and Technology (NIST) defines the cloud computing in a more appropriate way:

A model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (networks, servers, storage, applications and other services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. (AlZain et al., 2012)

The cloud computing can be defined in a simpler way: “A distributed virtual environment that provides virtualization based IT-as-Services by rent”. That is it is often better to get the required resources on the rent rather than purchasing one’s own resources. The main purpose for users to avail the cloud services is to avoid IT infrastructure purchasing and maintenance cost to get data accommodatable storage space for their large amount of data in the cloud. Beside all cloud services such as Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS) and Infrastructure-as-a-Service (IaaS) (Jansen et al., 2011), cloud also provides storage service as a sub-service of IaaS service model. In storage as a service, the distributed database servers are available on rent to store users’ data. These services are available for all kinds of users without any business discriminations.

Cloud computing is facing a number of challenging threats due to its virtualized multi-tenant nature (Purushothaman & Abburu, 2012). Data security is always the main challenging threat to the quality of services in cloud and may suppress the users’ interest to adopt cloud services for their enterprise benefits (Ransome et al., 2010). All integrated and communicated environment business decisions and operations depend on the quality of the data and information risk management (YanJun & Wen-Chen, 2009), good quality data is required for better decisions (Lam & Chun, 2008).

In cloud environment, all kinds of data are stored on servers, and these services are distributed over the world. User does not have any idea where physically their data are stored. The loss of control on the

18 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/classification-of-file-data-based-on-confidentiality-in-cloud-computing-using-k-nn-classifier/224600

Related Content

Cloud Computing Applications in the Public Sector

Amir Manzoor (2016). *Cloud Computing Technologies for Connected Government* (pp. 215-246).

www.irma-international.org/chapter/cloud-computing-applications-in-the-public-sector/136880

Why We Disclose Personal Information Despite Cybersecurity Risks and Vulnerabilities: Obligatory Passage Point Perspective

Patrick I. Offor (2019). *Cloud Security: Concepts, Methodologies, Tools, and Applications* (pp. 1460-1477).

www.irma-international.org/chapter/why-we-disclose-personal-information-despite-cybersecurity-risks-and-vulnerabilities/224642

Fake Review Detection Using Machine Learning Techniques

Abhinandan V., Aishwarya C. A. and Arshiya Sultana (2020). *International Journal of Fog Computing* (pp. 46-54).

www.irma-international.org/article/fake-review-detection-using-machine-learning-techniques/266476

Evolution of Fog Computing Applications, Opportunities, and Challenges: A Systematic Review

Hewan Shrestha, Puviyarai T., Sana Sodanapalli and Chandramohan Dhasarathan (2021). *International Journal of Fog Computing* (pp. 1-17).

www.irma-international.org/article/evolution-of-fog-computing-applications-opportunities-and-challenges/284861

Necessity of Key Aggregation Cryptosystem for Data Sharing in Cloud Computing

R. Deepthi Crestose Rebekah, Dhanaraj Cheelu and M. Rajasekhara Babu (2019). *Cloud Security: Concepts, Methodologies, Tools, and Applications* (pp. 1393-1407).

www.irma-international.org/chapter/necessity-of-key-aggregation-cryptosystem-for-data-sharing-in-cloud-computing/224638