Computers in Industry 83 (2016) 108–120

Contents lists available at ScienceDirect

Computers in Industry

journal homepage: www.elsevier.com/locate/compind

Engineering of secure multi-cloud storage

Philipp Junghanns [a], Benjamin Fabian [b,*], Tatiana Ermakova [c]

[a] Humboldt-Universität zu Berlin, Institute of Information Systems, Spandauer Str. 1, 10178 Berlin, Germany
[b] Hochschule für Telekommunikation Leipzig (HfTL), Chair of Business Intelligence and Data Science, Gustav-Freytag-Str. 43-45, 04277 Leipzig, Germany
[c] University of Potsdam, Chair of Business Informatics, esp. Social Media and Data Science, August-Bebel-Str. 89, 14482 Potsdam, Germany

ARTICLE INFO

Article history:
Received 16 January 2016
Received in revised form 11 August 2016
Accepted 14 September 2016
Available online

Keywords:
Cloud computing
Data exchange
Security
Privacy

ABSTRACT

This article addresses security and privacy issues associated with storing data in public cloud services. It presents an architecture based on a novel secure cloud gateway that allows client systems to store sensitive data in a semi-trusted multi-cloud environment while providing confidentiality, integrity, and availability of data. This proxy system implements a space-efficient, computationally-secure threshold secret sharing scheme to store shares of a secret in several distinct cloud datastores. Moreover, the system integrates a comprehensive set of security measures and cryptographic protocols to mitigate threats induced by cloud computing. Performance in practice and code quality of the implementation are analyzed in extensive experiments and measurements.

1. Introduction

Cloud computing has revolutionized the way computational resources are made available [1]. It has brought up new delivery models that render new possibilities of renting infrastructure, computing platforms, and software as services, and reduces the need for large-upfront investments. Particularly, this entails benefits such as payment on a per-usage basis, low or no fixed costs, and short time-to-market [2]. Especially public cloud solutions often induce significant economies of scale resulting in cost advantages that cannot be achieved by in-house or mid-sized data centers [3]. On these grounds, cloud computing constitutes an attractive model for small and medium sized companies, and its resource elasticity enables to respond quickly to changing business conditions. But also larger organizations are increasingly integrating cloud services into their daily IT operations while many are already using them for almost all storage needs of company assets such as customer data, operational data, and digital archiving [4]. Cloud computing has become a key enabler for big data since it provides nearly unlimited resources, such as computational power and storage capacity, on demand [5]. Due to cost advantages, massive storage capacities and service levels, Cloud Data Stores (CDS) can also serve as inter-organizational storage centers and exchange hubs [6].

However, with more and increasingly confidential data being stored in the cloud and the shift of responsibility toward cloud services, new security and privacy threats arise. How far can a third party cloud service provider (CSP) be fully trusted in securing data from unauthorized access, secondary use, disclosure, manipulation, or even loss? Can data availability be guaranteed when relying on a single CSP? Recent reports on intelligence operations and Internet surveillance as conducted by, e.g., U.S. secret agencies [7] as well as subjection to foreign law and judicial authorities [8] have reflected concerns about the trustworthiness and privacy of public cloud solutions. Moreover, legislative frameworks, such as the EU Directive 95/46/EC [9] or HIPAA [10], require strong protection of sensitive data and impose requirements on their transmission and processing. This emphasizes the need for practical solutions that increase the security and privacy of data in the cloud.

This article presents the design and implementation of a cloud gateway system for secure storage in a multi-cloud architecture, based on secret sharing techniques and a comprehensive set of cryptographic protocols. It extends earlier work [11,12] on data sharing for healthcare processes by a generalized architecture, an extended and refined proxy design, a much better and mature implementation, and new evaluations on performance and code quality. As for the scope of our Secret Sharing Proxy (SSP), this article will focus on CDSs for binary large objects.

Inter-organizational sharing of data requires the system to incorporate authentication and authorization such as role-based (RBAC) or other fine-grained access control mechanisms. CSPs are considered as semi-trusted parties that may be tempted to break a

* Corresponding author.
E-mail addresses: philipp.junghanns@gmail.com (P. Junghanns), fabian@hft-leipzig.de (B. Fabian), tatiana.ermakova@uni-potsdam.de (T. Ermakova).

http://dx.doi.org/10.1016/j.compind.2016.09.001
0166-3615/© 2016 Elsevier B.V. All rights reserved.

client's security objectives but are not likely to collude in larger groups. Threshold secret sharing describes the concept of securely sharing a secret among a specified amount of players by splitting the secret into n distinct shares that are confidentially distributed to the participating entities. The secret is only recoverable if a threshold of $t \leq n$ shares are available while any $t-1$ shares give no information on the secret. Among other security measures, this concept can be applied to design a SSP that instead of only one CDS uses a set of multiple CDSs for storage of sensitive data, such as an Electronic Medical Record (EMR). This article also addresses novel cryptographic protocols such as Attribute-Based Encryption to realize access controlled collaboration throughout the proxy architecture as well to serve as an additional protection layer to safeguard confidentiality and resist collusion.

The remainder of the article is structured as follows. Section 2 presents the design of the gateway architecture including objectives and major system components. The secure storage workflow and software architecture are discussed in Section 3. The system is evaluated in Section 4 and compared to related work in Section 5. Section 6 concludes the article.

2. Design of the cloud gateway

2.1. Design objectives

The multi-cloud gateway must protect the confidentiality and integrity both of data at rest and data in transit as well as safeguard the availability of the system and the data stored. The gateway should suite application scenarios where records of size ranging from kilobytes to several hundred megabytes are stored. From a usability perspective, the data retrieval process must be executed fast in order to provide a satisfying user experience. Client systems may be heterogeneous in type and processing power as well as numerous in quantity. The gateway should provide accessibility for a large range of systems including mobile clients.

Relevant cloud storage services are usually based on classic web service protocols such as REST [13] and SOAP [14]. Many providers offer both protocols simultaneously [2]. The survey in [15] lists 76 storage APIs for public cloud storage services from which 64% are consumable as RESTful web services, 11% are provided as SOAP services, and the remaining are based other protocols such as XML-RPC or JSON-RPC. Currently, the most prominent enterprise-level CDS providers include Amazon's AWS Simple Storage Service (S3), Google's Cloud Storage, Microsoft's Windows Azure Blob Store, Rackspace's Cloud Files, among others. Therefore, the SSP must be designed in a way that allows for future extensibility and easy integration of changing storage service APIs and new providers.

2.2. Components of the cloud gateway architecture

The cloud gateway consists of several interacting, self-contained components that are separated for architectural, security, and organizational reasons as will be explained in the following.

SSP Server. The SSP Server (simply referred to as SSP) is the central part of the system. It receives confidential data from authorized clients on which it performs secret sharing algorithms in order to disperse the document into a configured amount of distinct shares that are to be transferred and stored at several CDSs. The integrity of created shares is safeguarded using a two-staged tampering detection. The SSP server exposes a web service interface toward clients that may connect to the SSP's secure storage, retrieval, update, and delete services. The SSP server acts as an intermediary proxy between clients and datastores so that it consumes web services provided by the CDSs, which are realized through storage connectors, and provides a unified storage web service for authorized and authenticated clients.

SSP Clients. The SSP Clients can be installed and run on a variety of different operating systems or platforms, such as workstations, laptops or handheld devices, and enable access to the SSP system. Clients can be used by human users who would like to securely store sensitive data in the CDSs. Clients can also be realized as software systems, such as a document management system, which utilize the SSP for data storage by integrating the SSP Client *Application Programming Interface* (API).

CP-ABE Key Authority. One of the SSP core security layers involves Ciphertext-Policy Attribute-Based Encryption (CP-ABE) that provides both fine-granular access control and data confidentiality. CP-ABE requires the issuance of secret keys used for decryption that are generated over a set of user specified attributes. In order to guarantee correctness of key issuance and attribute specification (i.e., to avoid deliberate key request for fraudulent or inapplicable user attributes), a trusted key-issuing entity named *Key Authority* is required.

SSP Datastore Connectors. Public cloud storage solutions offer a wide range of service interfaces. CDSs also differ in the granularity of access control for data stored and protection mechanisms in place. The SSP offers a flexible and extensible datastore plug-in architecture so that new or changing CDS APIs can be connected. This allows for future extensibility and, more notably, enables connectors of different nature such as locally or network attached datastores (e.g., FTP servers or NAS), to become part of the SSP's storage architecture as well. The latter is of particular importance for limiting public share exposure in a highly critical threat scenario.

PKI. Integrity of data stored within the SSP's architecture is safeguarded at multiple levels both at the client and server side. SSP clients use digital signatures based on certificates so that upon cross-SSP requests, other participants than the storage-initiating (or data owning) clients are capable of verifying the signature of requested data. Furthermore, clients authenticate against the SSP using X.509 certificates [16]. Both certificates should be issued by a Certificate Authority (CA) as part of a dedicated Public Key Infrastructure (PKI).

External CDSs. In the SSP's storage architecture, multiple instead of only one public CDSs are used to mitigate the risk of confidentiality compromises and increase availability through threshold secret sharing. The CDSs are considered as semi-trusted datastores so that it is assumed that security and privacy threats generally affect a CDS, however, it is assumed that distinct CSPs are unlikely to collude in compromising confidentiality of customer data. However, even in this case protection layers remain.

2.3. Deployment architecture and variants

The architecture depicted in Fig. 1 shows one possible and favorable deployment scenario of the SSP including its subsystems. It shows the SSP being operated within a trusted environment, e.g., as part of an organization's network infrastructure. In this setting, the SSP instance is dedicated to one organization. All participating entities are obliged to request for client authentication and signature certificates of the PKI, which can also issue certificates for external participants so that inter-organizational access and data sharing can be accomplished.

The main idea behind operating the SSP within a trusted network is that the SSP server manages access credentials, such as security tokens or CDS request signature private keys. Therefore, access to this material must be properly secured within a controlled environment. Secondly, if the SSP is used by a limited set of participants, operation within a demilitarized zone (DMZ) through a double-staged firewall would protect clients by controlled isolation from malicious Internet attacks.

Another particular benefit when operating the SSP in a trusted environment is that its extensible datastore connector mechanism