SBaaS: Safe Box as a Service

Mounira Msahli, Ahmed Serhrouchni,
Télécom ParisTech, Paris, France
Mounira.Msahli@telecom-paristech.fr,
Ahmed.serhrouchni @telecom-paristech.fr

***Abstract*— While paperless is a source of tremendous opportunities for companies, it is also a bearing of many new risks. Indeed, externalization of electronic filing system can expose the company to several vulnerabilities and threats. We propose, in our gSafe (Government Safe) project a new Cloud service, named Safe Box as a service (SBaaS). SBaaS is used for probative value archiving sensitive documents during a defined period in a secure environment. We propose a layered model that intends to satisfy Cloud user requirements and Cloud security challenges. In this paper we define the technical architecture of the service and its basic components. In addition we discuss its feasibility using Hadoop Distributed File System.**

Keywords- ***Safe Box; Cloud Computing; Security; Hadoop; HDFS***

I. INTRODUCTION

The need for a secure digital conservation for special kind of documents such as enterprise payroll, employment contracts, videos and telephone records are usually associated with several tight regulatory considerations. These documents can be either native digital data or derived from a digitization process and the requirement for a secure archiving may be for few hours or for several decades.

In this context, there is a need for a Digital archiving system to ensure the security of digital objects. The main contribution in this paper is new Cloud Service [1,2,3,4,5,6] called Safe Box as a service.

The occurrence of highly distributed systems using virtualized environments lays several challenges for network security. In the case of safety deposit box service externalization, users can store a highly sensitive data, such as digital passport and ID card for individual users or the Payroll and administrative matters for any enterprise, via online safe box. This service is equivalent to common service of physical safe box that we can found in a bank. Therefore the cloud is a kind of data bank and every file stored on it has a specific attributes (for instance the level of sensitivity, access rights, ownership, etc...).

Transmission of data via insecure network is a very risky operation. Indeed it is very important to remember that although the variety of cloud computing security proposals made until now, it remains many threats and vulnerabilities related mainly to data confidentiality and integrity[7,8]. These risks include for example divulgation of confidential data and unauthorized access. Despite the already cited security issues, paperless brings new controversies such as how to demonstrate an action performed on a cloud system? How to ensure the engagement of a user, how to prove the anteriority of document compared to another? And how to present this proof in a court in case of contentious?

In our work we tried to resolve these problems. The gSafe project offers a compact secure architecture. In addition our contribution is: functional architecture, technical security and specification requirements and feasibilities.

The map of this paper is as follows: in the second section we present the related work. In the third section we detail our proposition and our cloud architecture. In the fourth section we focus on the archiving process and security requirements that we have to define in such cloud service. In the fifth section we discuss the feasibility of our proposition using Hadoop [9, 10] as solution and finally we conclude in the sixth section.

II. RELATED WORK

Several recent works treated durability and sustainability issues in data storage Clouds . In fact, [11] focus on three main aspects to ensure high end to end durability by spreading data over Cloud platforms and performing end to end audit. It focuses specially on audit to resolve physical and software faults in storage clouds.

The DepSky in [12] gives a storage service that improves availability and confidentiality provided by commercial cloud services using encryption, encoding and replication of stored data.

Depot presented in [13] focuses on the trust challenge over Storage Cloud. It uses the Fork-Join Causal consistency (FJC)to make consistent ordered update for nodes in network and to make durability and recovery properties in the storage cloud.

[14] identifies user audit service as a solution for the data security in cloud data storage by detecting the misbehaving servers.

S.Kamara and K.Lauter in [15] presents a summary about recent researches stimulated by storage Cloud like attributes used encryption ABE[16,17] and Symmetric searchable encryption. In this same research context, we propose the SBaaS service.

The specificity of the SBaaS proposed service is: it targets at new users and meets a new Cloud market need. These users require online archiving service in charge of secure filing of data for a period of time already specified and to make the proof of archiving operation, hence it is a probative value

COLLABORATECOM 2013, October 20-23, Austin, United States

DOI 10.4108/icst.collaboratecom.2013.254190

cloud. It is intended for physical person and enterprises, we can take the example of an individual who archives his identity card and wants to recover it whenever needed anywhere all over the world.

III. SBAAS ARCHITECTURE

The key driver for our SBaaS architecture is: to reduce the risk of accidental or malicious access to sensitive stored data and its discloser, to resolve the problem of indexing encrypted data by proposing an XML metadata file for every stored file and to find a solution for the problem related to the trust of storage Cloud.

Generally, SBaaS contains a distributed file systems, network equipments, Cloud Controller and servers. These components are either logical or physical whence the proposed layered model architecture below. Indeed, Fig. 1 presents the layered model for the Safe Box as a Service:

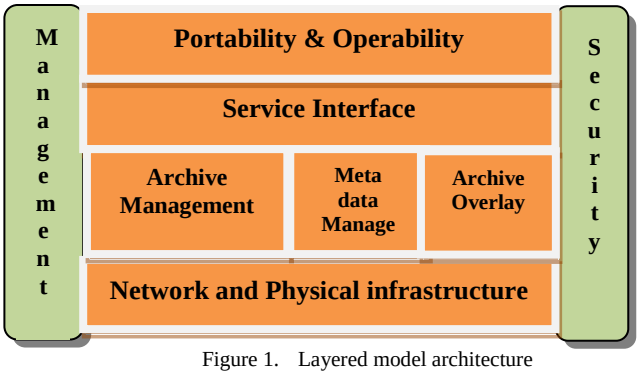

Figure 1. Layered model architecture

In this architecture we observe different layers:

1) The Network and Physical Infrastructure layer: presents the hardware part of the Cloud; this layer includes all network equipment, archiving server and all physical components used in the SBaaS Cloud.

2) The archive Management layer: includes all geographical parameters and attributes defining the archive location.

3) Meta Data management layer: includes all operations and metadata attributes concerning all archived files.

4) Archive Overlay layer: concerns all additional platforms such as virtualization components [18] usually used in Cloud services.

5) Service Interface layer: is in charge of delivering the archiving service to the SBaaS user. The service interface layer manages the required cloud infrastructure to provide a requested service.

6) Portability and operability layer: provides all types of interoperability with other Clouds for instance (Mobile Cloud [19]) and covers all communications inter-cloud like signaling operations.

7) Management layer: is a cross layer that covers all monitoring operations and audit management in order to provide a careful supervision over the entire Cloud.

8) Security layer: is also a cross layer since we need security through all the Cloud layers. The security layer defines several security requirements in the SBaaS Cloud such as:

- Integrity [19, 20]: verify the integrity of the documents using its digest,
- The availability and sustainability of documents through the redundancy archiving and periodic control,
- Confidentiality: the use of a secured transmission means in order to avoid the disclosure of sensitive documents
- Traceability: the ability to follow or reconstruct a history of events that took place in a system and make a precise answer to three questions who? when? and how?

IV. TECHNICAL ARCHITECTURE

A. *Technical architecture*

After presenting the global layered model of the SBaaS cloud service, this section is devoted to describe the associated network architecture:

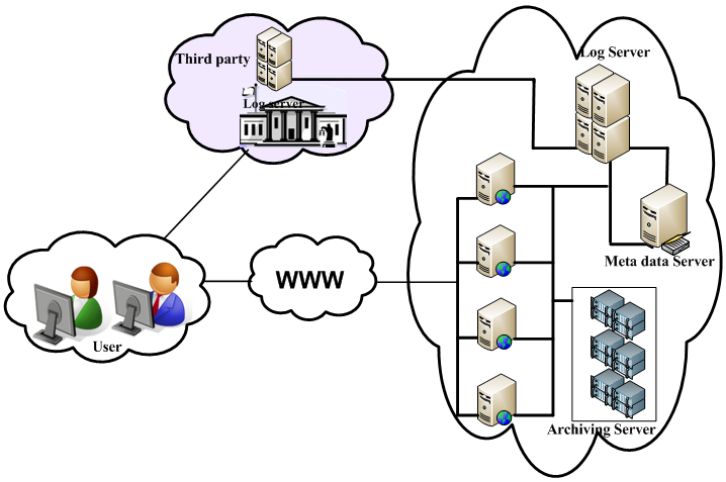

Figure 2. Network architecture

In Fig. 2, there are three actors: the Cloud provider, the cloud user and the third party proof manager . The cloud provider has three kinds of servers:

1) The Metadata server: is in charge of storing metadata related to the archived data, managing archive space (by controlling the place and the duration for every file in the archiving server) and saving the protection profile related to each archived document.

2) The Archiving server: is used to apply some security mechanisms for archived files, e.g. encryption, archive and delete the data when its archiving delay is reached.

3) Logging server: is used for the traceability. Indeed, the event log files can be used later as a proof in case of litigation. The log files have to contain all information regarding

archived documents including changes affecting it, e.g. the block-Id, user Id, files Id, time of the operation. Log files have to be signed in order to ensure its integrity.

The Cloud Proof Manager is composed of two kinds of servers: the logging server and the proof server used to generate the proof of storage.

Moreover, there are four main functions in the SBaaS:

1) Deposit: its goal is to verify the integrity of data and to encrypt the document, to be archived, before the safe deposit box,

2) Read: its objective is to get a copy of the archived file,

*3) Research:*is not performed directly on archived documents but on the classification scheme and metadata. In fact classification gives logical organization of archived files in the archiving server,

4) Delete: the shelf life of archived documents depends on its nature and Regulatory obligations. document.

B. Archiving Process in SBaaS

Fig. 3 displays the first step of the archiving process which covers different procedures ranging from authentication to the transmission of the document to be archived. In fact, the user has to authenticate to the metadata server in order to get a valid token. The token will be later used to make authentication in the archiving server and to submit the archived document. After receiving the valid token, the user sends the security profile and metadata concerning the file, object of the archiving process.

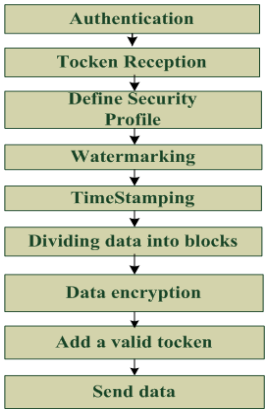

Figure 3. Archiving process in the Client Side

The security profile defines the security level of the file to be archived. It is a combination of four security attributes: the archiving time depending on the type of the document, the duplication number of an archived file, the distance between the duplicate copies of the file and the access control to the file.

With respect to the notion of the distance between the duplicate copies, the following example explains the concept: let (b1, b2) be two blocks, (r1, r2) be two racks and (c1, c2) be two different SBaaS clouds, three scenarios are possible:

- ✓ Two blocks in the same rack and in the same cloud: distance (b1/r1/c1, b2/r1/c1) =0
- ✓ Different blocks in different racks but in the same Cloud: the distance (b1/r1/c1,b2/r2/c1)=1
- ✓ Different blocks in different Clouds: the distance (b1/r1/c1, b2/r2/c2) = 2

For the access control attribute, we use the UCONABC [21] (Usage Control Model "obligation, Authorization and Conditions") based on the access control matrix with dimensions four (S: subject, O: object, C: condition and A: authorization). The matrix is defined by a subject S has a specific authorizations A over the object O under a condition C. As an example, let a cloud user (S) has the Authorizations (read, write) over his file (O) providing that the archiving time not yet expired (C).

The other information that will be transmitted in the security profile is the metadata, which is the title of the document, its creator, for which company, its language, the date of its creation and the scheduled date to erase it.

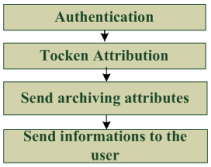

Figure 4. Archiving process in the metadata Server

After sending the security profile, the file is firstly watermarked [22] and time-stamped [23], and secondly divided into blocks and compressed. Subsequently, the file, to be archived, will be encrypted and sent to the archiving server with a valid token already taken from the metadata server.

The archiving process in the metadata server begins by the authentication operation, then the assignment of a valid token to Cloud user. The metadata server receives later on the security profile and sends the archiving metadata (see Fig. 4).

The archiving process, in the archiving server (see Fig. 5), starts by the verification of the token validity by checking its shelf life and its credibility. The server ensures on the integrity of the received data, encrypts and stores it in blocks identified by the bloc-id. Finally, the archive server sends an acknowledgment to the metadata server and to the user providing him a token useful later to retrieve his documents again if needed.

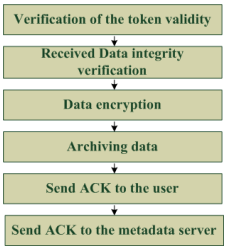

Figure 5. Archiving process in the Archiving server

V. SPECIFICATIONS AND REALIZATION

The feasibility and the validation of the proposed safe box service are based on the Hadoop Distributed file system (HDFS) [24]. The main components of master and slave HDFS architecture are the namenode and the datanode. Subsequently, in the case of gSafe we consider the namenode as the metadata server and the datanode as the archiving server.

Table 1 shows several security specifications and services that will be deployed in future works. Authentication mechanism is based on Kerberos [25], already deployed in Hadoop. The logging events and availability services provided by Hadoop are also retained. The gSafe project has added the Xades [26], the Pades [27] signatures and the time stamping module to Hadoop. The Fig. 6 presents the archiving scheme.

VI. CONCLUSION

This paper presents a new cloud service SBaaS. The specificity of this service compared to the other cloud services is best summed up by two question: how a cloud user might entrust his sensitive documents to a cloud provider? and how can the user prove the archiving operation? This paper tries to answer questions by giving an appropriate architecture with new cloud actor which is the proof manager, pointing out security requirements and highlighting safety aspects needed for online archiving service. In this paper, we defined layered model architecture and its technical specifications: encryption, access control, time stamping, authentication, watermarking …

The control access portion and the security profiles are already developed based on Hadoop API. XadeS and Pades signature has already been developed in the framework of gsafe project. Future works will be devoted to study and enhance the event log service and watermarking mechanisms.

TABLE 1: TECHNICAL SPECIFICATIONS OF THE GSAFE ARCHITECTURE

Features	Technical Specifications
Authentication	Kerberos already exist in Hadoop
Integrity	Time Stamping signature
Availability	Archiving duplication and distance creation already supported by Hadoop
Encryption	RSA [28, 29]
Watermarking	Depending on the file type
Access Control	UCONABC Usage Control (Authorisation, oBligation and Condition)
Traceability	Logging events

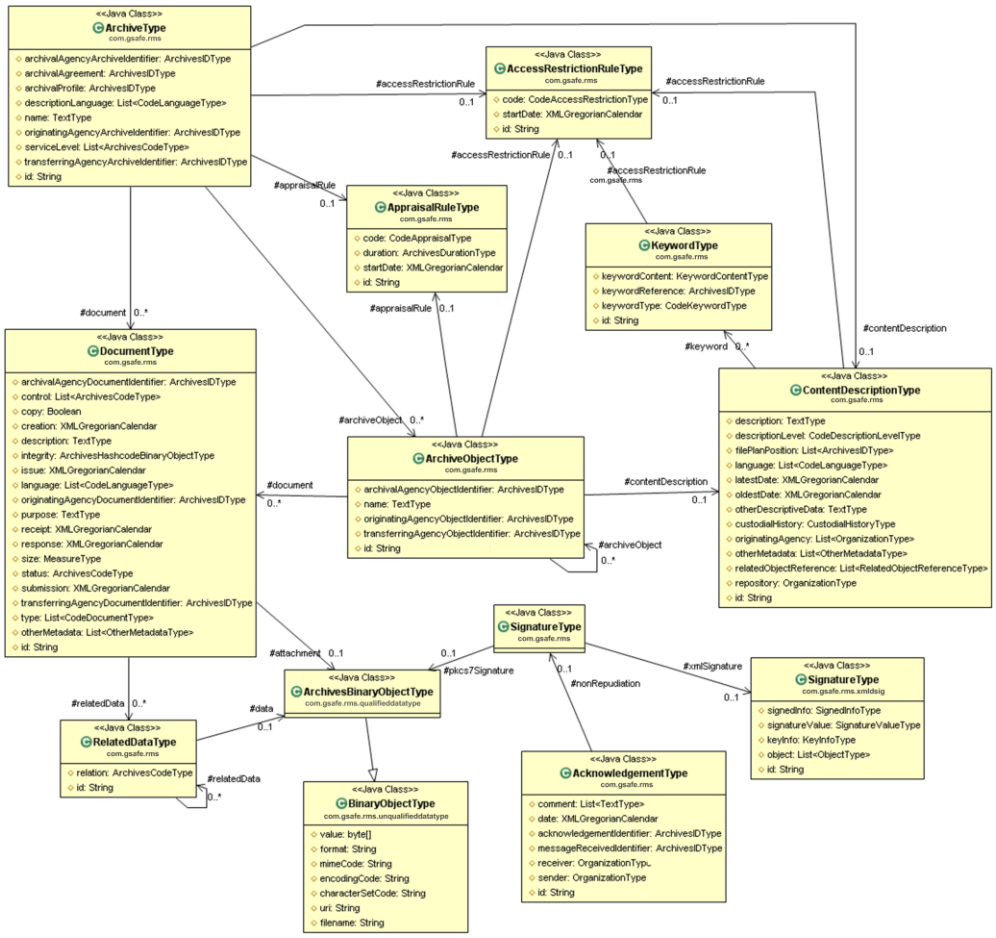

Figure 6. Archiving scheme

ACKNOWLEDGMENT

This work has been supported by the gSafe (government Safe) project sponsored by European Regional Development Fund.

REFERENCES

[1] Bhaskar Prasad Rimal, Eunmi Choi, Ian Lumb, A Taxonomy and Survey of Cloud Computing Systems, INC, IMS and IDC, 2009. NCM '09

[2] B. Khasnabish, J. Chu, S. Ma, Y. Meng, N. So, P. Unbehagen, M. Morrow, M. Hasan, Cloud Reference Framework draft-khasnabish-cloud-reference-framework-02, December 27, 2011

[3] C. N. Höfer, G. Karagiannis, Cloud computing services: taxonomy and comparison, Journal of Internet Services and Applications September 2011

[4] NIST CCRATWG 004 v2, Cloud Architecture Reference Models: A Survey, January 25, 2011

[5] Robert L.Grossman, Yunhong Gu, Michael Sabala, Wanzhi Zhang, Compute and Storage Clouds using wide area high performance networks, Future Generation Computer Sytems .2008 Elseiver

[6] FG group, Part1: Introduction to the cloud ecosystem: definitions, taxonomies, uses cases and high level requirements, ITU 2012

[7] ZHAN Ying, SUN Yong, Cloud Storage Management Technology, ICIC '09

[8] Wayne Jansen, Timothy Grance, Guidelines on Security and Privacy in Public Cloud Computing, Special Publication 800-144

[9] T White, Hadoop: The definitive guide, O'REILLY, 3rd edition

[10] Owen O'Malley, Kan Zhang, Sanjay Radia, Ram Marti, and Christopher Harrell, Hadoop Security Design, October 2009

[11] Ramakrishna Kotla, Lorenzo Alvisi, and Mike Dahlin, SafeStore: A Durable and Practical Storage System, 2007 USENIX

[12] A.Bessani, M.Correia Bruno, Q.Fernando, A.Paulo Sousa, DEPSKY: Dependable and Secure Storage in a Cloud-of-Clouds, EuroSys'11

[13] P. Mahajan S. Setty, S. Lee, A.Clement, M. Dahlin and M. Walfish, Depot: Cloud Storage with Minimal Trust, ACM Transactions on Computer Systems (TOCS), 29(4), December 2011

[14] C Wang, Q Wang, K Ren, N Cao, W Lou, Toward Secure and Dependable Storage Services in Cloud Computing, Services Computing, IEEE Transactions on, April 2012

[15] Seny Kamara, Kristin Lauter, Cryptographic Cloud Storage, Springer-Verlag Berlin Heidelberg 2010

[16] Amit Sahai, Brent Waters, Fuzzy Identity-Based Encryption, EUROCRYPT 2005

[17] Melissa Chase, Multi-authority Attribute Based Encryption, TCC'07: the 4th conference on Theory of cryptography

[18] M.msahli, A.fadhlallah, A.serhroucheni, G.Pujolle, F.mganem, OpenFlow and Ondemand network, NOF'2012

[19] Le Guan, Xu Ke, Meina Song, Junde Song, A Survey of Research on Mobile Cloud Computing, 2011 10th IEEE/ACIS International Conference on Computer and Information Science

[20] Dorothy Elizabeth Rob, ling Denning, Cryptography and data security, Addison-Wesley Publishing Company, 1982

[21] Jaehong Park, Ravi Sandhu, The UCON ABC Usage Control Model, ACM Transactions on Information and System Security, 2004, New York, NY, USA

[22] Van Schyndel, R.G., Tirkel, A.Z., A digital watermark, Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference Novembre 1994

[23] Stuart haber, Swott stornetta, How to Time-Stamp a Digital Document, journal of cryptology 1991

[24] Chuck Lam, HADOOP IN ACTION, 2011 by Manning Publications

[25] Jennifer G. Steiner, Clifford Neuman, Jeffrey I. Schiller, Kerberos: An Authentication Service for Open Network Systems, 1988

[26] XML Advanced Electronic Signatures (XAdES), ETSI TS 101 903 V1.3.2 (2006-03)

[27] Electronic Signatures a nd Infrastructures (ESI); PDF Advanced Electronic Signature Profiles; Part 1: PAdES Overview - a framework document for PAdES

[28] J. Jonsson, B. Kaliski, Public-Key Cryptography Standards (PKCS) #1: RSA Cryptography Specifications Version 2.1, The Internet Society (2003)

[29] M. Myers, R. Ankney, A. Malpani, S. Galperin, C. Adams, X X.509 Internet Public Key Infrastructure Online Certificate Status Protocol - OCSP, RFC 2560, june 1999