International Journal of Distributed and Cloud Computing
7 (1) June 2019, 05-13
http://www.publishingindia.com/ijdcc

Archiving-as-a-Service: A Comprehensive Reference Architecture Using Cloud Computing to Monitor and Supervise the Archive Documents from Long Term Point of View

A. H. Shanthakumara[1*] and N. R. Sunitha[2]
[1]Dept. of Computer Science and Engineering, Siddaganga Institute of Technology, Tumakuru, Karnataka, India. Email: ahskondajji@gmail.com
[2]Dept. of Computer Science and Engineering, Siddaganga Institute of Technology, Tumakuru, Karnataka, India. Email: nrsunithasit@gmail.com
*Corresponding Author

Abstract: This article analyzes the needs and proposes a cloud based comprehensive reference architecture based on the concept of "Archiving-as-a-Service". The reference architecture is designed in the context of new technology and organization for storing the documents and ensuring their security, reliability, availability, and authenticity over long periods of time. The article aims to bring changes in the practices of archiving documents by proposing comprehensive reference architecture with possibilities of wider approach to current outsourcing problems. This perspective, addresses the impact of cloud computing on today's archiving business, archiving service gaps, and the need of changes in the archive practice. The prototype implementation of selected modules, consider monitoring and supervising the archive documents. Finally, the authors justify the comprehensive reference architecture with the performance evaluations for some crucial services such as archiving, re-encryption, confirmation of integrity and proof of existence.

Keywords: Archive document, Archiving as a Service, Archive reference architecture, Long term preservation.

I. Introduction

The impact of digital computing and information technology has been remarkably universal. Vast quantities of digital data are produced, exchanged, and used in a variety of settings, for different purposes. Sometimes, the organizations or data owners store these data in either their own archival systems (on premises datacenters) or the networked archival systems (large scale remote datacenters or rented clouds). The survey Stancic *et al.* [1] exposed that the vast quantity of data is archived either on premises datacenters or rented clouds. It also indicates that, 61% of corporate mails, 58% of invoices, 48% of project data and 61% of user files are archived on their premises in the year 2017. The networked archive services are usually provided by third party called Archive Institutions / vendors. These institutions, more often than before, have implemented large scale datacenters for their archiving operations. The security, reliability and availability are obligatory requirements to be served by these institutions. Although those requirements may seem too easy to fulfil, it is not the case, if long term preservation and obsolescence of technology is taken into consideration. Nowadays, Archive institutions are moving archival content from on premises to cloud and outsourcing the archiving operations to cloud service vendors. The survey Stancic *et al.* [1] predicts at least 46% of corporate mails, 45% of invoices, 49% of project data and 54% of user files will be archived in the cloud by the year 2019. The archival system commonly provides not only long term preservation but also services like user authentication, security services, encryption key management etc.

This article is based on a method of literature study of current research articles on archiving principles, practices and their current issues with respect to monitoring and supervising the archive documents in a progressively advanced new technological environment. Archival system is very complex and problematic process. This article focus on the issues of monitoring and supervising the archive documents when archiving activities are outsourced as well as the new possibilities of wider approach based on the concept of Archiving as a Service.

The contribution of this article includes;

- Describe the service gaps with respect to the issues when archive content is outsource.

- Proposing comprehensive reference architecture based on the concept of Archiving as a Service by considering the possibilities of wider approach to current archival practices.
- Provides practical implementation for selected crucial protocols, which are used for monitoring and supervising the archive documents in a progressively advanced new technological environment.
- The practical implications of wider approach on the concept of Archival as a service open up new possibilities to standardize the archival practices and principles.

II. Related Work

Cloud storage has risen as an encouraging solution for not only supporting various business opportunities but also to reduce operational costs and streamlining their business efficiency. Today, organizations have already begun to use the cloud services to store their sensitive data. It is an opportunity for current archive institutions to enhance their business services using cloud storage as cutting edge technology over the internet.

The paradigm of cloud computing has a spectrum of several reference designs and models. The universally accepted classification of cloud infrastructures and subsequent service models are Platform as a service (PaaS), Software as a service (SaaS), Infrastructure as a service (IaaS), Public cloud, Private cloud and Hybrid cloud.

The Present State of Affairs in Private Cloud Infrastructure

The survey Stancic *et al.* [1] shows that the private cloud is a very familiar category of infrastructure for most of the service providers to deal with sensitive data or copyrighted archival materials. Indirectly, cloud vendors are very compatible with private cloud implicit technologies. Survey Stancic *et al.* [1] and Osterman Research [2] shows that the data owners are still careful in trusting cloud solutions for archiving their sensitive data even though a security and authenticity mechanism of the cloud infrastructure is on par with more traditional IT infrastructure implementations. The result of the survey Stancic *et al.* [1] and Osterman Research [2] also highlights that the storage and archiving in cloud environments are upcoming services. The skills of archiving professionals and their preservation related knowledge and experiences should be included in the implementation of the cloud environment itself. Some critical issues such as design of metadata, static (immutable) or dynamic content, content authentication and sensitivity level of the data are the key for long-term preservation and should be addressed by the archiving experts.

III. Archival Systems Background

There are several archive institutions, (Microsoft Azure [19], AWS [20], Google cloud platform [21], to name a few) in the market, provides low priced and secure cloud storage solution to archive documents. Archive document at rest automatically encrypt and provides seamless integration with varied priced storage tiers such as hot tier for more frequently accessible archive documents and cool tier for less frequently accessible archive documents. Fig. 1 illustrates the traditional concept of custody of archive documents by the archive institutions. Traditionally, archiving process is based on practices and principles of custodial acquisition model. Archive institutions carried out custodial task, monitoring and supervision at very large datacenters on their premises on behalf of data owners. Nowadays archive vendors are adopting cloud based archiving solutions than their on-premises counterparts. This process expect monitoring and supervision by the data owner. The data owners are usually insisted to use just standardised formats for their documents archived. The data owners usually do not possess knowledge of continuously changing formats of their documents over a long period time. Moreover, archiving services are becoming more specialized and complicated for data owners. The complexity involved in protecting and preserving the valuable documents is the major reason for data owners to show their interest in services from archive institutions. Further, since custody is outsourced, the archive institutions have no control over the documents archived in the cloud. It becomes much harder for the archive institutions to direct post- custodial activities and principles such as monitoring and supervision.

Service Gaps in Existing Practice

As a result of outsourcing the archiving operations to cloud service providers, it is harder for the archive institutions to monitor and influence archiving procedures. The skills of archiving professionals and their preservation related knowledge and experiences may not be fully utilized. Lack of sufficient levels of supervision on the quality of services with respect to both custodial activities and storage services in which the archive documents are both created and archived. Finally, the data owners are insisted to possess knowledge of monitoring and supervision of archive documents over a long period of time.

IV. Proposed Architecture for Archiving as a Service (AaaS)

The Open Archival Information System (OAIS) model, Lavoie and Brian [4], provides only services like integrity and authenticity. Many researchers (BSI Technical Guideline 0312 [5], Huhnlein *et al.* [6]), made an attempt to broaden the approach, Lavoie and Brian [4], by providing confidentiality and availability services for sensitive data. Some cloud service vender [22] offers long-term data preservation via a mix of on-premises gateways and IaaS based cloud infrastructure. However, the attempts made to improve the architectures focused more on to improve storage service by ignoring archive professional knowledge and experience. The architecture proposed by Stancic *et al.* [1] highlights the importance of archive community knowledge and experience in the activities of archiving process.

Article aims to refine and extend the architectures proposed in Lavoie and Brian [4], BSI Technical Guideline 03125 [5], Huhnlein *et al.* [6], Stancic *et al.* [1] and Shanthakumara and Sunitha [9] by providing wider approach to current archiving practices using SaaS. The possibility of wider approach could be an active involvement in standardization of archiving practices and principles (in lack of standards for this particular area) by archive institutions. The archive institution is required to be a key player in standardization of best practices and principles for archival document preservation. The archive institution should recognize the new technological benefits and deploy adequate archive services. Fig. 2 illustrate the comprehensive reference architecture for archiving practices based on the concept of Archiving as a Service (Stancic *et al.* [1]). The Archive as a Service (AaaS) plays the role of monitoring, supervising and service providing authority. It provides service to archive community by considering the influences of archive institutions and data owner needs time to time.

The AaaS responsibilities include but not limited to:

- Control of archived document without much interference and additional control taken by data owners and archive institutions.
- Ensure the confidence of data owners in the services provided is safe and stable. It also ensures the usage of archiving service is authentic and trusted source of evidence for data owners and archive institutions.
- Follow all professional regulations and provide standardized services through best practices intended for long term preservations. It ensures document should remain accessible by usage of methods such as migration, format conversions, record keepings etc.
- Provide proactive and secured preservation mechanisms, organizational context, high availability and access mechanisms and methods to prevent from content obsolescence.
- Respond to data owners' current needs, offer more than just rented storage space and become a key player in the archival service market.

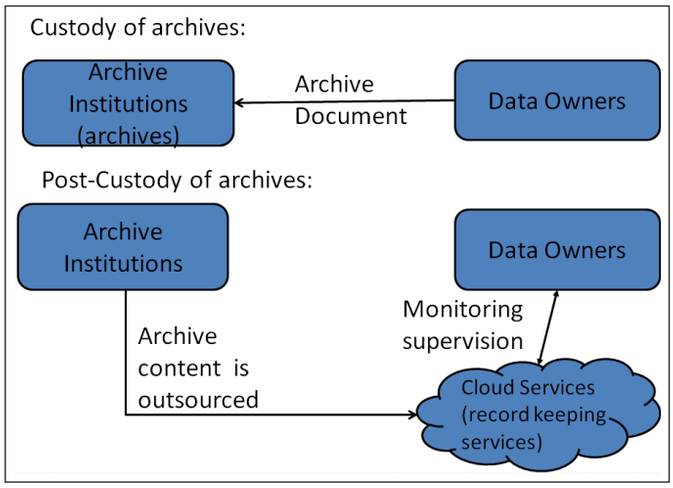

Fig. 1: Traditional Concept of Custody of Archive Documents by an Archive Institutions

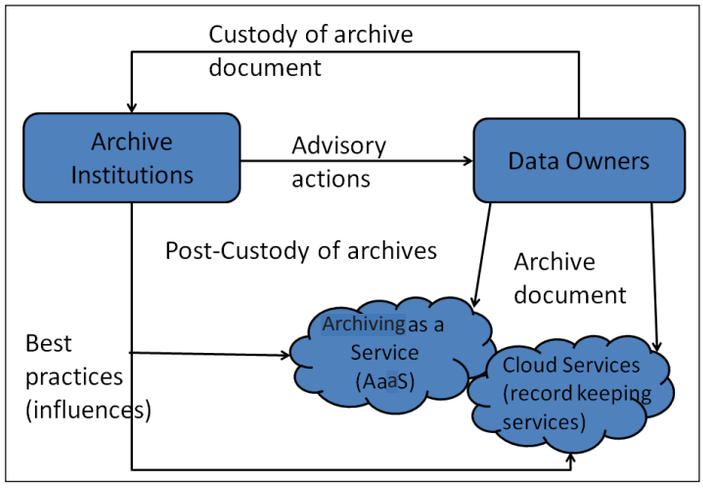

Fig. 2: Architecture for Archiving as a Service

V. Implementation

This section aims to illustrate the concept of AaaS based on the architecture as shown in the Fig. 2. We considered the requirements of archival systems, principles and practices from the previous research work Shanthakumara and Sunitha [7, 8, 9]. As discussed in the Section IV, the responsibilities of AaaS are very complex. Therefore, we implemented a basic prototype to illustrate the concept of AaaS by considering some crucial modules such as archiving, re-encryption and evidence confirmation etc. This basic AaaS consists of five logical modules and twelve interfaces as shown in the Fig. 3. The Application layer (data owner) may interact with the AaaS either through Archive Custody Module provided by archive institution or through AaaS Gateway Module provided by AaaS services.

A. Modules

The framework consists of five logical modules which describe essential activities for monitoring and supervising the archive documents. These logical modules are explained in this section using the web based interfaces as defined in B.

i) Access Control Service Module: This module simplifies the management of access permissions. We have used the familiar concept of role based access control model Sandhu *et al.* [3] to implement access control services. There are four sets of entities called data owner (DW), user (U), permissions (P) and Time period (T). The model has the capability to establish relationship among these entities to predefine user permissions and assign data owner to the predefined users. Data owner creates a time period and activates a subset of the access permissions. Mapping function assigns each time period to a single user or to a set of users dynamically. The interface RegistrationRequest/Response provides on demand security credentials to both registered data owner and archive institution. The interface AccessPermissionRequest/Response monitors access privileges and accordingly updates metadata.

ii) Archive Content Preparation Module: Selection of archive content can be a manual process but preparation service must be automatic and independent of data format. The descriptive metadata is to be useful from long term point view. The process of automatic gathering and representation of metadata is a complex process and it is an area of active research, Paynter [10]. The AaaS must provide interoperability and protect against format obsolescence by using best practice for converting data from its original form into one or more archiving formats.

iii) Reliability Services: The AaaS must reduce the risk of loss by improving overall reliability services using redundancy and active correction. Replication of an archival is the essential component of archival systems from long term point of view. The main inspiration for replications is its expanding availability and reliability in face of defeat. It is very important to decide how many number of replicas of archival to have and where to archive them, which are very stimulating points in the design of archival systems, Shanthakumara and Sunitha [7]. It guards no excess replications in the cloud by automatically incrementing or decrementing the replicas dynamically.

iv) Security Module: There are large numbers of security threats for the archiving system. The counter measures for these threats can be grouped into two categories. First, an attempt to address through the use of encryption in which every data object of archiving system is encrypted. Second, based on secret splitting in which the data object is split into n shares such that m shares (m is less than or equal to n) can be combined to reconstitute the secret and combining fewer than m shares reveals no information. This approach is called (m, n) threshold scheme. Some systems like publius, Waldman [11], use both cryptographic encryption and secret splitting and sharing schemes. The operation overhead incurs in encryption and decryption which is a serious issue for a very large archiving data. Some systems Haeberlen *et al*. [12], Maniatis *et al.* [13], Quinlan and Dorward [14], Lawrence *et al.* [15], are built with no specific secrecy policy and suffer from major security threats. The need of dynamically balanced security is addressed in Shanthakumara and Sunitha [7], since all data objects of archival system may not require a strong security, because sensitive level may not be same for all data objects of archival system. Different security levels for different archives must be used, depending on the sensitivity of archives and also to change the level of security dynamically as sensitivity of the archives changes from time to time. The interface ArchivalClassificationRequest/Response classifies archives dynamically based on their sensitivity. The sensitivity of some archives may gradually decrease and for some archives, it may gradually increase over a span of time. Hence this procedure must make sure that the sensitivity level of archives is regularly outlined. The interface Re-encryptionRequest/Response assigns distinct security levels to distinct archives based on dynamically classified sensitivity levels. Some archives such as public documents need not have same security level as private documents. This procedure ensures substantial security to high level sensitive archives and reduced or mild security level to low level sensitive archives.

v) Archive Management Module: The archived documents may be shared and stored in different cloud service providers. Guaranteeing that the data will persist unaltered and remains available for long time period is a difficult problem. The AaaS requires active management in the form of the following functionalities.

a) Secured Metadata Management: This manages information about the archived documents, keyword search rights and organizing index of each shares of archived documents stored in different clouds. The interface DocumentSubmissionRequest/Response creates a metadata node for each document to be archived. We used the data structure described in the scheme, Shanthakumara and Sunitha [8], to manage metadata node of each archived documents. This data structure is updated by the interfaces Re-encryptionRequest/Response, Access PermissionRequest/Response, ArchiveReadRequest/Response

and ArchiveDeleteRequest/Response. The representation of this metadata must be in one or more archival formats and securely stored.

b) Integrity and Proof of Existence: The AaaS must provide a process that periodically verifies and confirms the integrity and proof of existence of archival to an authorized entity. Simply distributing the shares of archival documents among different clouds or periodically encrypting the archival is not enough to guarantee that the data will get unaltered or remain available from long time point of view. The loss or damage of archival may not be realized until the document is reconstructed to original form. Therefore, AaaS takes a proactive approach to monitor the integrity and provides a proof of existence of archived document to an authorized entity. The interface EvidenceRequest/Response provides only essential data to any challenger to confirm integrity and proof of existence at any point of time.

c) Repair When Data Altered or Loss Occurs: The AaaS takes a proactive step to repair the archived documents when data is altered or loss occurs. An archive institution or data owner may confirm integrity and existence of archived document whenever they wish to confirm by collecting essential data using the interface EvidenceRequest/Response. There are many protocols for dynamic reconstruction of shares of archived documents without requiring to reconstruction of the original document, Wong *et al.* [16]. When data is altered or loss occurs, it can be recovered by invoking RepairRequest/Response.

d) Recover Archived Document: The AaaS performs necessary steps to recover the archive documents in case of a detected failure. It is very important to decide how many number of shares to generate and threshold for recovery that involves issues like privacy protection, Ganger *et al*. [17].

e) Retrieval of Archival Document: The archived documents must remain accessible to authorized entity regardless of state of the submission system, independent of document format and time period of archival. There are several methods to implement access control to archived documents Sandhu *et al.* [3]. The interface SearchPermissionRequest/Response delegate keyword search rights on a set of archived documents, the interface SearchPermissionRequest/Response enables to perform keyword search on a set of archived documents and the interface ArchiveReadRequest/Response provides whole document after checking security credentials.

B. Interfaces

This section describes the operations which can establish the interactions among various modules defined in A. The following interfaces are defined as basic request and response types and can be realized in web services. The interfaces are shown as number inside parenthesis in the Fig. 3.

i) RegistrationRequest/Response: This interface is invoked when Archive Institution (AI) raise the request for registration. It generates and returns a pair of initial Secret Key (SK) and Public Key (PK) for each registered archive institution. Initial keys for Archive Institution are $SK_{AI(t,\ 0)} \rightarrow {}^{R}Z_N^*$, $PK_{AI,\ t} = SK^{2^{T+1}}_{AI,(t,\ 0)} \bmod N$. where N is product of two large prime numbers and Time interval T is divided into equal periods T_i, where $i = \{1, 2, \ldots, t\}$.

ii) DocumentSubmissionRequest/Response: This operation will be invoked when a new document is to be archived. This service may be requested by registered archive institution by uploading documents to be archived. In turn, AaaS will setup the security parameters for each submission request by invoking SetSecurity() and Encryption() method. The response contains the certificate for proof of integrity and existence obtained by invoking Evidence () method.

SetSecurity$(1^{\lambda}, m)$**:** It takes security parameter 1^{λ} and m is maximum possible number of documents as input and produces the public system parameters, *params*. It consists of $(B, PK_{DW,\ t}, H)$ where

- B = (p, G, G_1, e) and p is the order of G and $2^{\lambda} \le p \le 2^{\lambda=1}$. Let G & G_1 be two cyclic groups of prime order p and g be the generator of G and $e : G \times G \rightarrow G_1$ is the bilinear map, Bellare and Miner [18].
- $PK_{IA,\ t} = SK^{2^{T+1}}_{IA(t,\ 0)} \bmod N$ and $SK_{IA,\ (t,\ 0)} \leftarrow {}^{R}Z_N^*$.
- H: $\{0, 1\}^* \rightarrow G$ is a chosen one way hash function.

Encryption(): This method is used by AaaS to encrypt each document $i \in \{1, 2, 3, \ldots m\}$ with his public key $PK_{AI,\ t}$ using pre-defined algorithm and generate its keyword ciphertexts, C_w, for each document with index $i \in \{1, 2, 3, \ldots m\}$.

Evidence(): This function computes the proof of integrity and existence, Shanthakumara and Sunitha [8], the result is stored as metadata along with simple description (date of creation, time stamp of archiving). The certificate for proof of integrity and existence, $E_i \leftarrow (F_i = H(D_i), S_i = H(\text{Encryption } (D_i))$, is computed for each document D_i with index $i \in \{1, 2, 3, \ldots m\}$, where H is a chosen one way hash function.

iii) AccessPermissionRequest/Response: A method maps each Time periods T_i, where $i = \{1, 2, \ldots t\}$ to a set of user $(Ti) \subseteq \{u \mid (\text{Data Owner } (T_i), u) \in UA\}$ and Time period Ti has the permissions $U_r \in$ user (Ti) $\{p|(p, u) \in PA\}$ where $PA \subseteq P \times U$, a many-to-many permissions to user assignment relation and $UA \subseteq DW \times U$, many-to-many data owner to user assignment relations.

iv) SearchPermissionRequest/Response: This interface delegates keyword search rights on set of archived documents to the authorized entity. It takes archive institution secret key $SK_{AI,\ (t,\ i)}$, where $i = \{1, 2, \ldots t\}$, and a set $S \subseteq \{1, 2, 3, \ldots m\}$ which contains the indices of documents as input, then produces the aggregate key k_{agg}. The aggregate key k_{agg} is provided to the authorized data owner.

v) KeywordSearchRequest/Response: This interface takes aggregate key k_{agg} and a keyword, w, to be searched and produces either true or false to indicate the keyword w found in i^{th} document of S or not, where $i = \{1, 2, \ldots t\}$, and a set $S \subseteq \{1, 2, 3 \ldots m\}$.

vi) ArchiveReadRequest/Response: This operation retrieves archive content to an authorized entity by checking security credentials.

vii) ArchivalClassificationRequest/Response: This interface is used to classify an archive document into High Sensitive or Medium Sensitive or Low Sensitive based on dynamically determined sensitivity of the archival. We considered the simple archive classification method, Shanthakumara and Sunitha [7], which considers the archive document history, like number of reads, number of access permissions and number of re-encryptions in a certain period of time. The response contains the level of sensitivity of the archived document as either high or medium or low sensitive. With this classification result, an archive institution may influence right level of security to the archived document. It makes sure a substantial security to high level sensitive archival and condensed or mild security level to low level sensitive archival.

viii) Re-encryptionRequest/Response: This operation runs re-encryption on selected archival under two circumstances. Firstly, when an authorized entity (Data owner) needs a re-encryption with a new key. Secondly, an archive institution may influence re-encryption periodically or as a result of periodical classification of a archived document. We considered the method described in Shanthakumara and Sunitha [7], which take the input from data owner / archive institution as the level security required, apply the re-encryption algorithm and returns the security credentials to authorized entity. The certificate for proof of integrity and existence, Shanthakumara and Sunitha [8], has to be computed, $E_i \leftarrow (F_i = F_i \oplus H(D_i))$, $S_i \leftarrow S_i \oplus H(\text{Re–}Encryption\ (D_i))$, and correspondingly update the metadata for each document D with index $i \in \{1, 2, 3 \ldots m\}$ re-encrypted.

ix) EvidenceRequest/Response: An archive institute or a data owner, who wish to confirm the integrity and proof of existence for the archive document, D_i can collect the record $E_i \leftarrow (F_i, S_i)$, current hash value $H(D_i)$ and Authentication Path from AaaS. The challenger can construct Merkle root, Shanthakumara and Sunitha [8], with the help of authentication path, specific time interval as record index and confirms the existence of their archived document D_i.

x) RepairRequest/Response: This operation does the reconstruction of shares of archived documents. We used the scheme described in Shanthakumara and Sunitha [7], which takes essential security credentials and applies different repair methods for different archive documents based on the security applied to a document.

xi) RecoverRequest/Response: This operation recovers whole archived document using metadata which describes the security credentials and location of redundant copy.

xii) ArchiveDeleteRequest/Response: This operation deletes archive document from cloud by checking security credentials of an authorized entity.

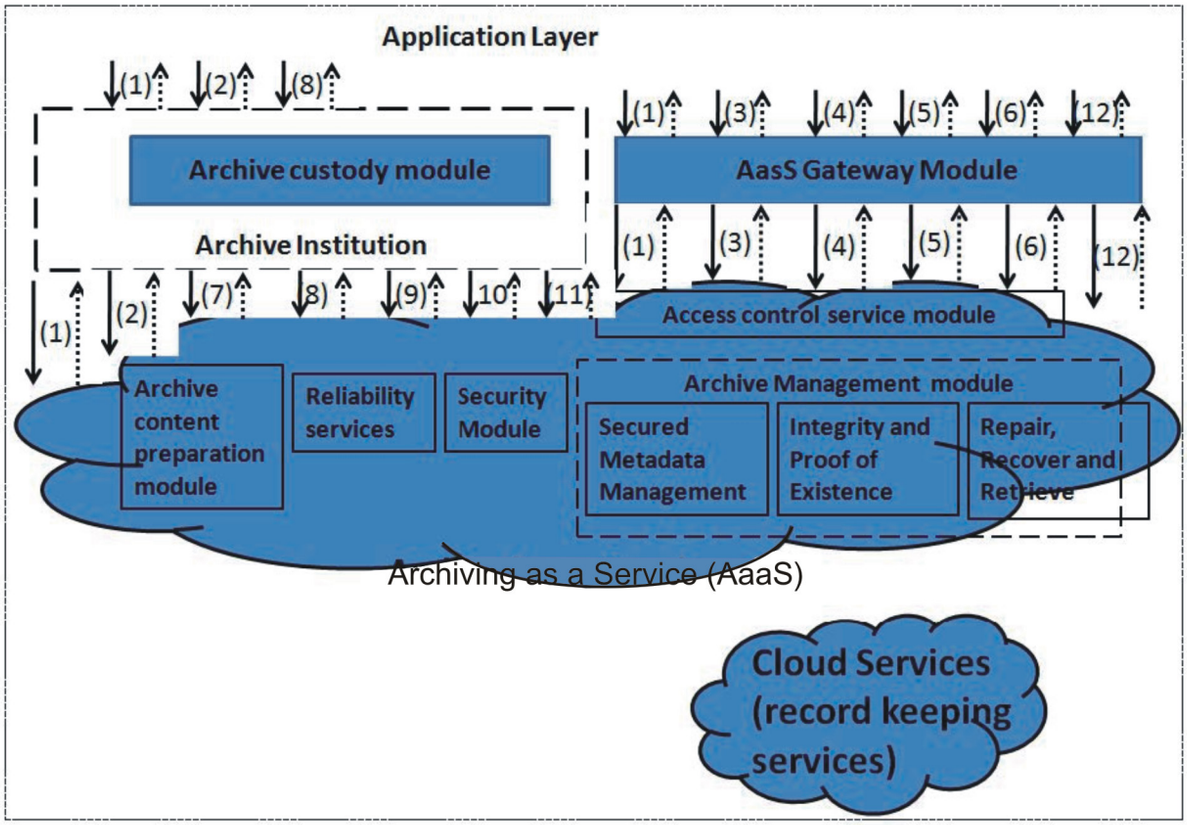

Fig. 3: Overview of Modules and Interfaces for the Services Provided by AaaS

VI. Results and Discussion

The authors implemented some crucial algorithms they selected using JAVA code on their own datacenters and private cloud infrastructure for the simulation purpose. In the simulation, archive institution selected 90,000 emails randomly from Enrons employees' data base [23] with an average size of 1.9 KB. Archive institution treats each email with all its attachment as a single document and outsourced 10 documents initially to AaaS. Once the simulation begins, archive institution dynamically outsources the remaining documents in 30 time intervals as defined in the algorithm in Shanthakumara and Sunitha [7], Over five runs of simulation of Periodical re-encryption algorithm in Shanthakumara and Sunitha [7], it could process up to 81000 emails in 30 time intervals, out of which nearly 23000 documents are classified as highly sensitive data objects (HSD). The Fig. 4 shows the storage requirements for data objects and Meta data with consideration of periodical re-encryption. We compared it to the simulation result of Shanthakumara and Sunitha [8], and conclude that there is not much difference in terms of storage requirements for the archive documents and their metadata whether it is archived in the IT infrastructure or on the cloud.

EvidenceRequest/Response: The Fig. 5. shows the time required to confirm the integrity and proof of existence. The time includes both the collection of the record $E_i \leftarrow (F_i, S_i)$, current hash value $H(D_i)$ and Authentication Path from AaaS and the construction of the Merkle root with the help of authentication path by the archive institution, Shanthakumara and Sunitha [8]. Time difference between on-premises datacenter and AaaS as noted in the Fig. 5 may include the delay due to internet access.

Re-encryptionRequest/Response: The Fig. 6. shows the time required to compute Re-encryption using AES algorithm in DCM syntax, Shanthakumara and Sunitha [7], at both on-premises and AaaS side. The time includes both for the checking security credentials and re-encryption process.

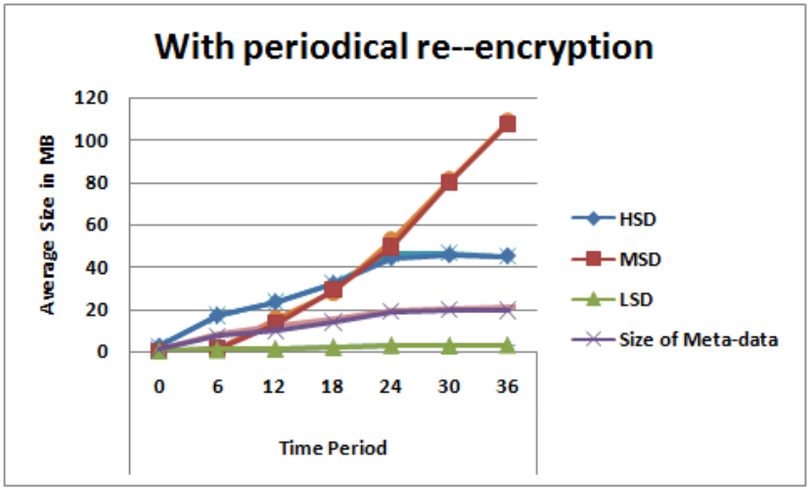

Fig. 4: Storage Requirements for Data Objects and Meta Data When Periodical Re-Encryption is Executed by AaaS

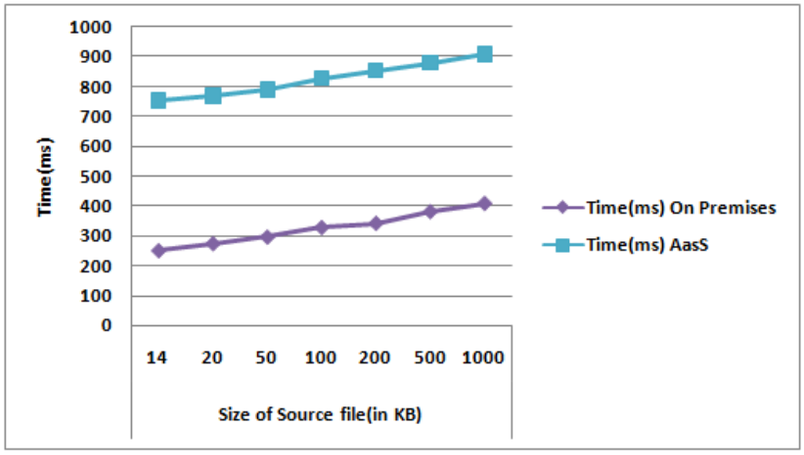

Fig. 5: Time Required to Confirm the Integrity and Proof of Existence

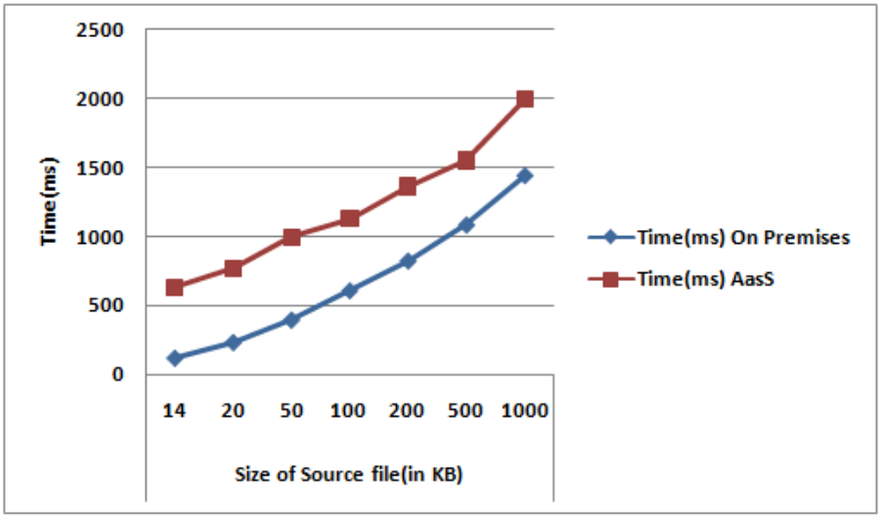

Fig. 6: Time Required to Compute Re-Encryption

VII. Conclusion

The concept of Archiving as a Service provides a new dimension to the process of archiving documents from long term point of view. The practices and principles cultivated for monitoring and supervising the archive documents in a progressively advanced new technological environment plays an important role. When new technological advances happen, the principles and knowledge of the archive community should be considered to perform monitoring and supervisory activities. The private cloud AaaS could take a proactive approach which considers the influences of archive institutions and data owner needs time to time and provide on demand services. The concept of AaaS gives the possibility of wider platform to implement new mechanisms which cloud be developed, tested and become part of best practices. It brings the new technological benefits and deploys adequate archive services to data owners. This paper describes and suggests the need for a comprehensive reference module for monitoring and supervising the archive documents. It describes the simulation results for some selected crucial services such as archiving, re-encryption and evidence confirmation etc. The result shows that there is not much difference, in terms of both storage requirements for the archive documents and their metadata and time required to perform the operations, between whether the document is archived in the IT infrastructure or on the cloud.

References

[1] H. Stancic, A. Rajh, and I. Milosevic, ""Archiving-as-a-Service": Influence of cloud computing on the archival theory and practice," White Paper [Online], 2016. Available: http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/VC_Stancic_et_al_26_B_1430.pdf

[2] Osterman Research, Inc., "Strategies for archiving in hybrid environments," An Osterman Research Whitepaper, published October 2017. Available: https://www.viewpointe.com/uploadedfiles/viewpointe/pdfs/strategies%20for%20archiving%20in%20hybrid%20environments%20-%20viewpointe.pdf

[3] R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman, "Role-based access control models," *Computer*, vol. 29, no. 2, pp. 38-47, February 1996. DOI: 10.1109/2.485845

[4] B. Lavoie, "Meeting the challenges of digital preservation: The OAIS reference model," The OAIS Reference Model, DPC Technology Watch Report, October 2014. Available: www.oclc.org/research/publications/library/2000/lavoie-oais.html

[5] BSI Technical Guideline 03125, "Preservation of evidence of cryptographically signed documents, federal office for information security," October 2015. Available: https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/Publications/TechGuidelines/TG03125/BSI_TR_03125_TR-ESOR_V1_2_EN_Main.pdf?__blob=publicationFile&v=1

[6] D. Huhnlein, U. Korte, L. Langer, and A. Wiesmaier, "A comprehensive reference architecture for trustworthy long-term archiving of sensitive data," *2009 3rd International Conference on New Technologies, Mobility and Security*, Cairo, pp. 1-5, 2009. DOI: 10.1109/NTMS.2009.5384830

[7] A. H. Shanthakumara, and N. R. Sunitha, "A novel archival system with dynamically balanced security, reliability and availability," *2016 International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS),* Bangalore, pp. 27-33, 2016. DOI: 10.1109/CSITSS.2016.7779435

[8] A. H. Shanthakumara, and N. R. Sunitha, 2018, "Techniques to certify integrity and proof of existence for a periodical re-encryption-based long-term archival systems," in N. Rao, R. Brooks, and C. Wu. (eds), *Proceedings of International Symposium on Sensor Networks, Systems and Security (ISSNSS)*, Springer, Cham, 2017. Available: https://doi.org/10.1007/978-3-319-75683-7_16

[9] A. H. Shanthakumara, and N. R. Sunitha, "A comprehensive reference architecture for re-encryption based archival systems with insurability services," *2017 2nd International Conference on Emerging Computation and Information Technologies (ICECIT),* pp. 1-5, 2017.

[10] W. Paynter, "Developing practical automatic metadata assignment and evaluation tools for internet resources," *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'05)*, Denver, CO, pp. 291-300, 2005. DOI: 10.1145/1065385.1065454

[11] M. Waldman, A. D. Rubin, and L. F. Cranor, "Publius: A robust, tamper evident, censorship-resistant, web publishing system," in *Proceedings of the 9th USENIX Security Symposium*, pp. 59-72, August 2000.

[12] A. Haeberlen, A. Mislove, and P. Druschel, "Glacier: Highly durable, decentralized storage despite massive correlated failures," in *Proceedings of the 2nd Symposium on Networked Systems Design and Implementation (NSDI),* USENIX, Boston, MA, May 2005.

[13] P. Maniatis, M. Roussopoulos, T. J. Giuli, D. S. H. Rosenthal, and M. Baker, "The LOCKSS peer-to-peer digital preservation system," *ACM Transactions on Computer Systems*, vol. 23, no. 1, pp. 2-50, 2005.

[14] S. Quinlan, and S. Dorward, "Venti: A new approach to archival storage," in *Proceedings of the 2002 Conference on File and Storage Technologies (FAST),* UNISEX, pp. 89-101, Monterey, California, USA, 2002.

[15] L. L. You, K. T. Pollack, and D. D. E. Long, "Presidio: A framework for efficient archival data storage," *ACM Transactions on Storage*, article 6, July 2011. Available: http://dx.doi.org/10.1145/1970348.1970351

[16] T. M. Wong, C. Wang, and J. M. Wing, "Verifiable secret redistribution for archive systems," in *2002 Proceedings of the First International IEEE Security in Storage Workshop,* pp. 94-105, 2002. DOI: 10.1109/SISW.2002.1183515

[17] G. R. Ganger, P. K. Khosla, M. Bakkaloglu, M. W. Bigrigg, ..., and J. J. Wylie, "Survivable storage systems," *2001 Proceedings of the DARPA Information Survivability Conference & Exposition II (DISCEX'01)*, vol. 2, pp. 184-195, Anaheim, CA, 2001. DOI: 10.1109/DISCEX.2001.932171

[18] M. Bellare, and S. K. Miner, "A forward-secure digital signature scheme," *Advances in Cryptology-Crypto 99*, Springer-Verleg, 1999.

[19] Microsoft Azure (n.d.), Azure Archive Storage. Retrieved from https://azure.microsoft.com/en-in/services/storage/archive/

[20] AWS (n.d.), Data Archive. Retrieved from https://aws.amazon.com/archive/

[21] Google cloud platform (n.d.), Archival Cloud Storage: Nearline & Coldline. Retrieved from https://cloud.google.com/storage/archival/

[22] ZDNet (n.d.), Archiving as a service: A cold-storage solution. Retrieved from https://www.zdnet.com/article/archiving-as-a-service-a-cold-storage-solution/

[23] Enron Data, RData file (enron-mysqldump). Available: http://www.ahschulz.de/enron-email-data/