# Archival Approach to IaaS Cloud Services

**Hrvoje Stančić, Adam Al-Hariri**
Faculty of Humanities and Social Sciences
University of Zagreb
Ivana Lučića 3, 10000 Zagreb, Croatia
{hstancic, aahariri}@ffzg.hr

**Edvin Buršić**
Financial Agency – FINA
Vrtni put 3, 10000 Zagreb, Croatia
Edvin.Bursic@fina.hr

**Abstract**. *Increasing growth of new technologies and new technological solutions follow increasing demand for new business. The increased usage of new technologies and the higher uses' awareness of the potential problems they bring along have resulted with the need to establish trust in these new services. By services the authors specifically mean Infrastructure-as-a-Service (IaaS) cloud services. The trust can be established in a number of ways such as through the implementation of relevant archival standards, principles of trusted digital repositories, applying to legal procedures, careful service implementation planning etc. In this article the authors focus on the research of the problems and critical points of cloud services, especially paying attention to the areas of governance, compliance, trust, architecture, identity and access management, software isolation, data protection, availability, and incident response.*

**Keywords.** cloud service, IaaS, archive, long-term preservation, trust in digital records

## 1 Introduction

According to the National Institute of Standards and Technology (NIST), cloud computing is "a model for enabling ubiquitous, 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. This cloud model is composed of five essential characteristics, three service models, and four deployment models." [7, 2]

Project Records in the Cloud [10] identifies five essential characteristics of cloud solutions:

1. **On-demand self-service** allows users to access as many computing capabilities as they need
2. **Broad network access** allows users to access the cloud from any machine that has an Internet connection
3. **Resource pooling** allows the multi-tenant model supporting multiple users at the same time
4. **Rapid elasticity** allows users to change the amount of computing resources they need at any time
5. **Measured service** allows precise measuring of utilised resources in terms of storage, processing, bandwidth etc. These resources can be monitored, controlled and reported to the users, who are only charged for what they use by pay-as-you-go model. In most cases this approach reduces costs.

Stancic, Rajh and Milosevic [12, 110-111] differentiate between three service models as follows:

- *Software as a Service (SaaS)* – ability to deliver applications from cloud-based physical infrastructure, accessible via various client software tools or devices. The user has no awareness or control of the underlying physical components or software configuration capabilities outside the delivered application.
- *Platform as a Service (PaaS)* – ability to deliver complete environments (operating systems and required tools) for testing or development of external applications. The user, however, has no control over the configuration settings of the application-hosting environment.
- *Infrastructure as a Service (IaaS)* – ability to deliver complete virtual data centres to the user who is then able to configure and deploy virtual machines and other relevant/corresponding virtual components according to their personalized requirements.

According to the four deployment models Stancic at al. [12] further state that cloud implementations include:

- *Private cloud:* where it is implied that the cloud infrastructure is built and provisioned for private use by a single organization. Private clouds in practice tend to be service-oriented with specific roles and requirements.
- *Community cloud:* where the physical infrastructure is implemented, administered, and operated by several organizations in a certain

community of consumers from organizations that have shared goals and requirements.
- *Public cloud:* the cloud infrastructure is intended for "rent" by the public users, as delegated by the provider usually for profit or other means of compensation for the provider.
- *Hybrid cloud:* the combination of two or more physical cloud infrastructures from different branches of the above listed deployment models that are physically separate but are connected via the means of mutual data and application portability or management hierarchies.

The research results presented in this paper are part of the research activity "Ensuring trust in storage in Infrastructure-as-a-Service" done within multinational research project InterPARES Trust: Trust and Digital Records in an Increasingly Networked Society. The aim of the research in this paper is to investigate the issues of trust in cloud services, particularly in Infrastructure-as-a-Service (IaaS).

# 2 Trust in a Cloud service?

In InterPARES Trust's project terminology database the term *trust* is defined as "confidence of one party in another, based on alignment of value systems with respect to specific actions or benefits, and involving a relationship of voluntary vulnerability, dependence and reliance, based on risk assessment". [4] This means that the users of cloud services should have enough information on a particular service (e.g. in Terms of Service) in order to trust it, or the service level agreement (SLA) between users and cloud service provider (CSP) should equally protect interests of both parties involved. Users' interests increasingly shift from the pure use of a service or infrastructure to additional requirements such as trustworthiness of a service model. The InterPARES Trust project is using Society of American Archivists' (SAA) definition of *trustworthiness* saying that "in general, trustworthiness is synonymous with reliable. In archival literature and records, trustworthiness is often defined in terms of reliability and authenticity. This definition loses its apparent circularity when the reliability of records is understood in the diplomatic sense, 'created with appropriate authority, according to established processes, and being complete in all its formal elements'. In the context of electronic records, trustworthiness often implies that the system is dependable and produces consistent results based on well-established procedures." [4] Therefore, it is needed to investigate what makes a cloud service trusted. In this paper we will approach the issues of trust in IaaS from the archival point of view.

The concept of IaaS cloud service is gaining market momentum. Many similar yet different services exist on the market. From the archival point of view scarcely any of those services has been developed having long term preservation in mind let alone the issues of preservation of authenticity during that process. Since the IT development is very fast, hardware, operating systems, software, file formats, metadata structure etc. change along. The files normally used can become obsolete in five to ten years. Whose responsibility is to preserve them, migrate to newer formats while at the same time preserve authenticity of those records – service providers' or service buyers'? How can service providers demonstrate the trustworthiness of such procedures? Is there a service provider willing to invest in additional archival IaaS and what would that mean? Having these questions in mind we are addressing several important issues a company should consider when purchasing a cloud service. Specifically, the company should seek for information that could guarantee trust in the service.

In the next sections we have, adhering to the structure of the NIST's Guidelines on Security and Privacy in Public Cloud Computing [5], investigated ten areas important for establishing overall trust in IaaS. In each area we address the questions that need to be answered if a company purchasing a cloud service can trust the service provider to have trustworthy system implemented. The areas of analysis are: General information, Governance, Compliance, Trust in the service, Architecture, Identity and Access Management, Software Isolation, Data Protection, Availability, and Incident Response. At the end of the research paper we have listed questions that should have a clear answer before purchasing an IaaS Cloud service.

# 3 Trust and general information on a cloud service

General information about a cloud service can offer initial suggestions whether the service can be trusted or not. Ideally, general information should answer some core questions about Infrastructure-as-a-Service.

Components used in IaaS are: storage, servers and network. [14] While considering storage most common solutions are SAN (Storage Area Network) and DAS (Direct Attached Storage). The problem with enterprise-class SAN solutions is that they require expensive dedicated hardware. DAS is fast and inexpensive to procure, but it comes with far more single points of failure, for example a disk controller failure can easily corrupt an entire storage array. DAS disk space utilisation is poor, 36% on average [8], due to inability to independently share data or unused resources with other servers, which SAN solutions do efficiently.

A cloud server is primarily an IaaS-based cloud service model. There are two types of cloud server: logical and physical. A cloud server is considered

logical when it is executed through server virtualisation. A physical server is logically distributed into two or more logical servers, each of which has separate OS, user interface and apps, although they share physical components from the underlying physical server. Whereas the physical cloud server is also accessed through the Internet remotely, it isn't shared or distributed. This is commonly known as a dedicated cloud server. [13]

Cloud computing took concepts from SOA (Service-oriented Architecture), Client-server model, grid computing, utility computing and peer-to-peer networking. The most important service offered in IaaS is virtualisation technology or, to be more precise, provision of unlimited number of virtual machines. Besides that, cloud service providers offer additional resources such as virtual-machine disk image library, raw storage, file-based storage, firewalls, load balancers, IP addresses, virtual local area networks (VLANs), and software bundles. While talking about virtualisation security one must understand that most security issues arise not from the virtualisation infrastructure itself but from operational issues. [11] Virtualisation improves security by making it more fluid and context-aware. This means security is more accurate, easier to manage and less expensive to deploy than traditional physical security. Security in a virtualised data centre can also be more fully automated. Virtualisation security gives data centre administrator' the power to automatically provision secure machines, automatically have security policies follow desktops when they move, automatically set up firewall sets of rules for classes of servers and automatically quarantine compromised or out of compliance assets, etc. [1]

Therefore, the questions to be asked regarding the issue of trust in IaaS in the area of general information should at least be:

1. Which components are used in IaaS?
2. What types of services are offered in IaaS?
3. What technologies are being used?
4. What implications used technologies have on security and privacy of the system?

If a cloud service offering IaaS have addressed these questions and have made publically available information on these issues it would be the first step towards the idea of a trustworthy archival IaaS cloud service. Little by little, by looking at this and the following areas we have researched, the whole picture of the concept of trustworthy archival IaaS will be built. Next, we will concentrate on the area of governance.

# 4 Trust in governance

Good and effective governance is the key factor in assuring security over data produced by a company. Once the company has externalised any part of IT sector, the security becomes an issue. Therefore, the issues considering governance become more important when data is entrusted to a CSP (Cloud Service Provider). The reason why a company would risk such a security issue is, of course, connected with the cost reduction policy of the company.

When dealing with issues of governance in relation with the trust in IaaS, a company should search for information about the CSP's storage and access policies to data, answering a delicate question whether the CSP can prove that the company's data cannot be mixed with the data of another customer of the provided service, or whether it can prove that the employees of different ranks does not have access above their limits. Readily available answers to the following questions would render an IaaS cloud service more reliable:

1. Is it possible for a client to monitor security of computing environment and data security? How?
2. What kind of security assures a client that his data is not mixed with another clients' data?
3. What kind of security assures a client that there is no data shared with employees of a different rank and/or not created by others?
4. What audit mechanisms and tools are used to determine how data is stored, protected and used to validate services, and to verify policy enforcement?

Conclusively, it is important to be able to monitor how the data is stored, protected and used, in order to validate the purchased service, and to verify that the company's policy is enforced.

# 5 Compliance affecting trust

Different countries have different types of security and privacy laws and regulations at all levels – national, state, and local – increasing the responsibility of a CSP to operate in agreement with the established laws, regulations, standards and specifications. For a company considering IaaS it is important to be aware of the fact by which laws the CSP is governed by, where is geographically the data stored, and is any part of the service outsourced etc. In that sense a company should investigate the differences between the legal regulations of its own country versus the ones that a CSP is governed by. It should also investigate whether its data own are allowed to cross border. Once data crosses border it might become unclear which laws and regulations are to be enforced. Also, if the CSP considered is using subcontractors it should be made clear who is handling the data, and is there any business continuation policy they adhere to, i.e. how can the CSP guarantee the data will not be disclosed, disrupted, modified, destroyed or become unavailable if a subcontractor handling the data goes out of business. In this context, the minimum of four questions a company should ask the CSP it is considering as a service provider:

1. Does your service comply with other countries' laws, regulations, standards and specifications for clients outside your country of service?
2. How is your service secured against unauthorized access, use, disclosure, disruption, modification, or destruction of data?
3. What technical and physical safeguards do you assure?
4. Do you use subcontractors for any part of the used technology or offered service?

# 6 Trust in the service

Trust in the service has many different aspects such as availability of service, ownership rights, certification, risk management and physical and logical security.

Assurance of availability of service can be proven by implementing high quality protection from external attacks. Denial of Service (DoS) type of attack on the CSP holding a company's data might prevent or impair the authorized use of networks, systems or applications by exhausting CSP's resources. The CSP should make clear if it has redundant cloud storage implemented. Also, since the cloud services offered are usually carried out via the use of virtual servers and applications, for a company seeking this kind of services it would be good to have information about physical and logical security of virtual servers and applications. All this could significantly improve trust in the availability of service.

Another concern regarding trust in the service is connected with the issues of ownership of the data given to the custody of CSP. It is reported that a well-known social network service has eternal ownership of users' personal content, even if one decide to delete one's account. That means that they have the right to use, copy, publish, stream, store, retain, publicly perform or display, transmit, scan, reformat, modify, edit, frame, translate, excerpt, and adapt any content one have ever uploaded, including the option to use one's name, likeness and image for any purpose. [9] Although social networks should not be mixed with IaaS CSPs, one could learn from the example and look carefully into the issues of possible transfer of ownership between a company and a CSP.

Trust in a cloud service could be established if the service has relevant certificates, e.g. of quality of services, technology, or architecture, given by a third party. If the certificates are relevant for the archival long term preservation a company could regard that CSP as trustworthy. Trust could also be established if risk management procedures are transparent. Risk management processes that should be clearly explained are those that cover areas of framing risk, assessing risk, responding to risk, and monitoring risk. [3] The users opting for an IaaS should be asking the following questions:

1. Does your service secure from denial of service attack?
2. Does your service secure ownership rights over data?
3. Does your service have any certificate relevant to your service?
4. What kind of risk management does your organization provide?
5. What kind of physical and logical security is assured for the virtual servers and applications?

Publically available information covering answers to those questions would be clear indicator that a CSP has implemented a minimum set of needed procedures.

# 7 Architecture influencing trust in the service

Jansen and Grance [5, 22] claim that the architecture of the software and hardware used to deliver cloud services can vary significantly. The design and implementation of the reliability, resource pooling, scalability, and other logic needed in the support framework is determined by CSP. Virtual machines typically serve as the abstract unit of deployment for IaaS clouds and are loosely coupled with the cloud storage architecture. To complement the server side of the equation, cloud-based applications require a client side to initiate and obtain services.

Between an operating system and hardware platform there is an additional layer of software, the hypervisor or virtual machine monitor, which is used to operate multi-tenant virtual machines. The hypervisor conducts administrative operations, such as launching, migrating, and terminating virtual machine instances. For a company choosing between several CSPs it would be wise to investigate security level of hypervisor or virtual machine monitor.

The questions like following need to be asked:

1. How is a hypervisor or virtual machine monitor secured?
2. How do you secure virtual machine images from attack looking for proprietary code and data?
3. Do you use image management process to govern the creation, storage, and use of virtual machine images?
4. How do you secure from attacks on the client side?
5. How do you secure from attacks on the server side?
6. Do you use encrypted network exchange?

By getting answers to those questions it would be possible to understand the technologies the cloud provider uses to provision services and the implications the technical controls involved have on security and privacy of the system throughout its lifecycle. With such information, the underlying

system architecture of a cloud can be decomposed and mapped to a framework of security and privacy controls that can be used to assess and manage risk.

# 8 Identity and access management

Data sensitivity and privacy of information have always been an area of concern for organizations. The identity proofing and authentication aspects of identity management entail the use, maintenance, and protection of PII (personally identifiable information) collected from users. The data collected by the provider of the purchased service include details about the accounts of consumers, data about customer-related activity, data collected to meter and charge for consumption of resources, logs and audit trails, and other such metadata that are generated and accumulated within the environment, data of an organization's initiative (e.g. the activity level or projected growth of a start-up company), metadata collected by the provider etc. A company migrating its business to a cloud should ask a CSP for clear answers to the questions regarding identity and access management.

# 9 Software isolation

In order to achieve the flexibility of on-demand services CSPs have to use high degrees of multi-tenancy over large number of platforms. This multi-tenancy in IaaS cloud computing environment is typically done by multiplexing the execution of virtual machines from potentially different consumers on the same physical server.

Multi-tenancy in virtual machine based cloud infrastructure, together with the way physical resources are shared between guest virtual machines, give raise to new sources of threat. In the man-in-the-middle attack, the intruder uses a program that appears to be the server to the client and appears to be the client to the server. The attack may be used simply to gain access to the message, or to enable the attacker to modify the message before retransmitting it. So, if administrative control of guest virtual machines is obtained then the man-in-the-middle attack approach can be used to modify the code used for authentication. Once the code is modified the attacker has access to the client's data. Also, if the system is attacked, the client is giving the attacker its password freely and with no knowledge or suspicion. Answers to these questions should be made available in order to facilitate trust in the cloud service:

1. How do you prevent man-in-the-middle attacks?
2. How do you secure from attacks on the server that targets passwords?

Publishing clear statements on how CSPs prevent attacks targeting passwords is a step forward to gaining trust in the cloud service.

# 10 Data protection

Data protection concerns both data-in-transit and data-at-rest. Data-at-rest can be easily protected by the encryption mechanism while protecting data-in-transit is a bit more complicated. E.g. declaration of adherence to the NIST's Federal Information Processing Standards Publications [2] is a good indicator a CSP is thinking about data protection. A company considering a CSP could ask the following questions:

1. What kind of encryption do you use to secure data stored in IaaS?
2. Have you conducted deliberate attacks in order to test your system's protection?

From the archival point of view, it is equally important to preserve data providing enough information on its provenance, authenticity, integrity, and reliability as it is to delete the data when that is needed or required. The latter is important in order no data is retrievable after it is deleted. E.g. CSPs could consider NIST's Guidelines for Media Sanitization [6]. For this aspect the following question is relevant:

1. Upon termination of service what procedures do you use for data sanitization, i.e. how do you secure that data, after deletion is not recoverable?

As already mentioned before, the physical, geographical location of data centres as well as of back-up or archival locations are important to know because the CSP can be located in one country and its data centre in another. Therefore, it is important to understand the implications this situation might have on the data being stored, backed-up and archived. The following questions could provide relevant answers in deciding whether a company is e.g. legally allowed to use an IaaS cloud:

1. Where, geographically, is data stored?
2. Where, geographically, is data backup stored?

The information of geographic locations a CSP is using can result in increase or decrease of trust regarding the security of company's data, and in understanding of the laws and regulations that will be carried out in a case of data loss.

# 11 Availability

A CSP can have facility damage or loss due to either nature or human-influenced disasters. Or it can run out of business, go bankrupt, or have other financial difficulties. A legal seizure could also happen. If an organization relies on IaaS service for data storage and processing, it needs to see clear statements in a CSP's contingency plan on solutions to prolonged or permanent system disruptions, especially with mission critical operations, until the restoration of the

service. Answers to the following questions could help a CSP gain trust:

1. In a situation of a seizure how do you assure availability to users not under seizure?
2. What is your policy regarding user data availability in a case of bankruptcy or other facility loss?

# 12 Incident response

Incident response involves an organized method for dealing with the consequences of an attack against the security of a computer system. The CSP's role is vital in performing incident response activities, including incident verification, attack analysis, containment, data collection and preservation, problem remediation, and service restoration. [5, 33] The incident response plan should be publically available and should clearly state how and in what time a CSP plan to conduct the restoration of the service and determine the scope of the incident and affected assets. Also, the incident response plan should cover the plans to repair the security breach and assure it does not happen again. Information given as answers to the following questions should be made available by a service provider:

1. What is your organization's incident response plan?
2. Do you keep track of the data by which you can determine the scope of the incident, and assets affected?
3. Do you keep a forensic copy of incident data for legal proceedings or as needed by the consumer? Or, do you give incident data to your consumers?

# 13 Conclusion

For a company to take an archival approach to choosing IaaS cloud service provider would mean to invest some time and effort to investigate the ten described areas and decide which CSP should be trusted or can be regarded as trustworthy. For a CSP offering IaaS and wishing to become a trusted CSP taking the archival approach explained here would mean investing some time and effort to make its business as transparent as suggested to gain users' trust. Many CSPs already conduct many of the analysed procedures, but that might not be visible from their publically available documents, statements, policies etc. Therefore, it is in CSPs' interest to make their procedures available for consideration since that is the way for them to become trusted cloud service providers.

Information provided as answers to the set of questions devised during this research should become a minimum set of information that a CSP should provide on an IaaS cloud service offered. That information could help users to choose the right CSP, could function as a check-list for CSPs who want to upgrade their services, and could as well serve as a good starting point for developing trust metrics. If a CSP provides all the information affecting trust in the service, as analysed and discussed in this article, than it could rightfully claim that the service it is offering is a trusted archival service.

# Acknowledgments

This paper shows initial research results of the activity "Ensuring trust in storage in Infrastructure-as-a-Service" within the multinational research project InterPARES Trust: Trust and Digital Records in an Increasingly Networked Society. More about the project activities at: http://interparestrust.org.

# Disclaimer

Opinions or points of view expressed are those of the authors and do not necessarily reflect the official position or policies of the institutions they work with.

# References

[1] Catbird. Why Virtualization Security?, http://www.catbird.com/vsecurity/best-practices, downloaded: April 4th 2014.

[2] *Federal Information Processing Standards Publications (FIPS PUBS)*, NIST, http://csrc.nist.gov/publications/PubsFIPS.html, downloaded: March 20th 2014.

[3] *Guide for Conducting Risk Assessments*, NIST Special Publication 800-30, Revision 1, September 2012, http://csrc.nist.gov/publications/nistpubs/800-30-rev1/sp800_30_r1.pdf, downloaded: May 3rd 2014.

[4] Project *InterPARES Trust: Trust and Digital Records in an Increasingly Networked Society*, http://interparestrust.org.

[5] Jansen, W; Grance, T. *Guidelines on Security and Privacy in Public Cloud Computing*, NIST Special Publication 800-144, December 2011., http://csrc.nist.gov/publications/nistpubs/800-144/SP800-144.pdf, downloaded: March 20th 2014.

[6] Kissel, R; Scholl, M; Skolochenko, S; Li, X. *Guidelines for Media Sanitization*, NIST, September 2006., http://csrc.nist.gov/publications/nistpubs/800-88/NISTSP800-88_with-errata.pdf, downloaded: March 20th 2014.

[7] Mell, P; Grance, T. *The NIST Definition of Cloud Computing*, National Institute of Standards and Technology, Gaithersburg, MD, 2011., http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf, downloaded: April 4th 2014.

[8] *Parallels Cloud Storage: The Ideal Storage Solution for Hosters*, http://www.parallels.com/fileadmin/media/hcap/pcs/documents/PCS_The_Ideal_Storage_Solution_for_Hosters_WP_EN_Ltr_10182012.pdf, downloaded: April 4th 2014.

[9] Raphael, JR. Facebook Privacy Change Sparks Federal Complaint, *PCWorld*, February 17th 2009., http://www.pcworld.com/article/159703/facebook.html?tk=rel_news, downloaded: March 20th 2014.

[10] Duranti, L. *Records in the Cloud: Detailed Description*, http://www.recordsinthecloud.org/secure/documents, downloaded: April 8th 2014.

[11] Rendell, R. Virtualization Security and Best Practices, *Netsecure 2008: IT Security and Forensics*, Illinois Institute of Technology, http://www.cpd.iit.edu/netsecure08/ROBERT_RANDELL.pdf, downloaded: April 8th 2014.

[12] Stancic, H; Rajh, A; Milosevic, I. "Archiving-as-a-Service", Influence of Cloud Computing on the Archival Theory and Practice. In Duranti, L; Shaffer, E. (Eds.), *The Memory of the World in the Digital Age: Digitization and Preservation*, pages 108-125, Vancouver, Canada, 2012.

[13] Techopedia. *Cloud server*, http://www.techopedia.com/definition/29019/cloud-server, downloaded: April 4th 2014.

[14] *Understanding the Cloud Computing Stack: SaaS, PaaS, IaaS*, http://www.rackspace.com/knowledge_center/whitepaper/understanding-the-cloud-computing-stack-saas-paas-iaas, downloaded: April 4th 2014.