Privacy in the Cloud: Bridging the Gap between Design and Implementation

Vassilis Manousakis[1], Christos Kalloniatis[1], Evangelia Kavakli[1],
and Stefanos Gritzalis[2]

[1] Cultural Informatics Laboratory, Department of Cultural Technology and Communication,
University of the Aegean, University Hill, GR 81100 Mytilene, Greece
{kavakli,ct08081}@ct.aegean.gr, chkallon@aegean.gr

[2] Information and Communication Systems Security Laboratory, Department of Information
and Communications Systems Engineering, University of the Aegean, GR 83200,
Samos, Greece
sgritz@aegean.gr

Abstract. Bridging the gap between design and implementation stages has been a major concern that deplores designers, analysts and developers for quite a long time during the design and implementation of information systems in traditional environments. This issue grows to bigger dimension with the presence of cloud computing. Designing and modeling an Information System for the Cloud is a major and hard task that most of the traditional software engineering approaches fail to fulfill. In parallel, many respective organisations and respective researchers have highlighted a number of security and privacy challenges that are not present in traditional environments and need special attention when implementing or migrating information systems into a cloud environment. Thus, security and privacy are by themselves two areas that need special attention in the cloud era. This paper moves on to this direction. Specifically, it presents a number of privacy-oriented technical concepts that analysts need to consider when designing and modeling privacy-aware systems in a cloud environment. Also it suggest for every concept a number of implementation techniques that can assist developers in implementing the respective concepts.

Keywords: Cloud Computing, privacy concepts, CSA, Implementation Techniques, Software Engineering, PETs.

1 Introduction

Cloud Computing is without a doubt one of the most significant innovations presented in the global technological map. The number of potential users enrolling and using cloud services increases exponentially on a daily bases. The great demand from online users for cloud services along with the reduced operational costs that the latter offers has motivated many organisations and companies to consider implementing from scratch or migrating organizational services, data and applications on the Cloud. However, despite the various positive characteristics of all cloud service models like reduced costs, better availability insurance, on demand data storage and computation

X. Franch and P. Soffer (Eds.): CAiSE 2013 Workshops, LNBIP 148, pp. 455–465, 2013.

power, cloud users have expressed major concerns regarding the protection of their privacy in such environments basically due to the distributed character of the cloud architecture and the involvement of different stakeholders and providers on specific applications and data processing mechanisms.

According to National Institue of Standards and Technology (NIST), Cloud Computing delivers three different types of services to the end users that derive from three different models. The delivery models are IaaS, PaaS and SaaS, each one of them providing three distinct types of resources, like virtual infrastructure resources, application platforms and software services. Each delivery model is considered as separate layer that is depended from each other and with IaaS being the foundation, PaaS sits on top of IaaS and SaaS sits on top of PaaS. So, as the end users combine different type of services, capabilities form each layer are inherited, so as privacy issues. Also another factor that should be considered is the impact of deployment model on privacy. Privacy risks seem to have bigger impact on public, hybrid and community cloud, compared to the other deployment models. On the other hand, cloud consumers should have in mind that despite the fact that private and cloud deployments are theoretically safer, but still the same threats apply and the only thing that changes is the users group. In this deployment model, the users from the administrator to the simple user are trusted, but that does not mean that proper measures should be considered.

In general, the more low level services the client requests the more responsible for security and privacy is, but still the cloud vendor has an important role on managing and implementing security and privacy measures even in low levels of abstraction.

The scope of the paper is twofold. Firstly, as far as we know, it makes one initial step on identifying and describing the major privacy-related concepts that are newly introduced into the cloud. Secondly, it aims on bridging the gap between design and implementation stages by suggesting for each privacy concept a number of implementation techniques for realizing these concepts on a cloud environment. Specifically the paper is structured as follows. Section 2 describes in text and graphically the privacy related concepts. In section 3 the respective implementation techniques are presented that realize the aforementioned concepts. Finally, section 4 concludes the paper and suggests future extensions.

2 Privacy-Oriented Concepts

In order to preserve privacy inside the cloud, certain requirements need to be realised. This section describes the basic privacy properties that constitute the basic issues that need to be considered when designing or migrating to the cloud. Specifically, the aim of this section is twofold. Firstly, it aims on revealing and describing a number of privacy related concepts derived from related literature as well as respective cloud threats identified so far both in text and diagrammatically.

Secondly, it aims on identifying the applicability of every concept on the respective cloud service model thus assisting the stakeholders on deciding which privacy properties need to be realised in order to satisfy their own goals on every cloud service model respectively. The concepts proposed are mainly derived from the European Commission Draft Report on Security Issues in Cloud Computing [5] as well as from our previous work presented in [6-13]. However, new concepts are also introduced and explained in order to form a complete set for covering all the respective cases. For every concept a brief description is described along with the main privacy issue and the main threats existing from the respective literature regarding this issue.

2.1 Isolation

The specific concept is referred to the complete seal of user's data inside the Cloud computing environment. Isolation is meant to address data disclosure in two ways, firstly, from purpose limitation point of view and secondly from the aspect of the proper technical implementation techniques [5]. Cloud computing resources are shared among a multi-tenant environment. Thus, excessive cloud employee's access rights, posing the risk of any kind of Personal Identifiable Information disclosure and thus violating client's privacy. The specific concept is matched with the following threats derived from [1], Abuse and Nefarious Use of Cloud Computing, Insecure interfaces and APIs, Malicious Insiders, Shared technology issues, Data Loss or Leakage, Privileged user access and Lack of Data Segregation.

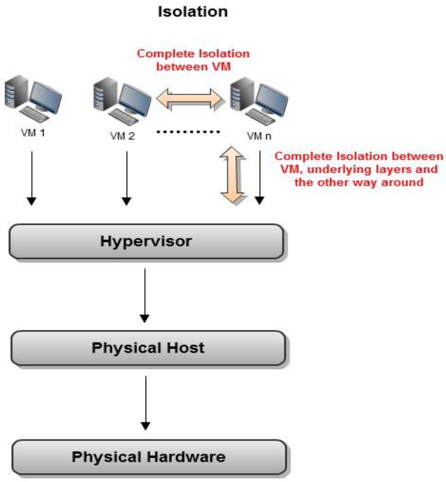
Isolation
Complete Isolation between VM
VM 1
VM 2
VM n
Complete Isolation between VM, underlying layers and the other way around
Hypervisor
Physical Host
Physical Hardware

Fig. 1. Isolation Example

2.2 Provenanceability

The specific concept is referred to the provenance of the data related to the authenticity or identification, the quality of the results of certain procedures, modifications, updates and vulnerabilities, the provenance of certain actions inside the cloud, the detection of origins of security violations of an entity[14], the auditability of client's data and matters that are related to the cloud's sub-system geographical dispersion referred to the legal issues, regulations, policies and each country's rules as far as data processing and protection is concerned. All the above constitute a potential privacy violation if they are not realised properly by implementing the appropriate technical measures. The specific requirement is matched with the CSA threats, Malicious Insiders, Privileged user access, Regulatory Compliance, Data Location, Investigate Support.

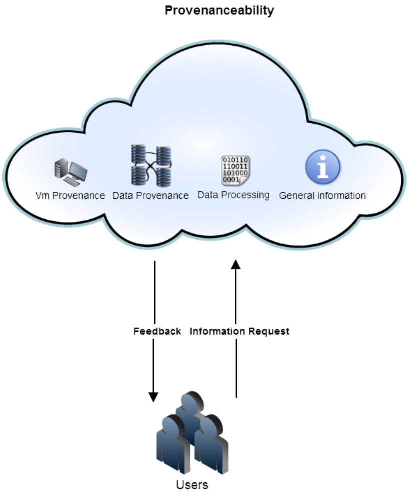

Fig. 2. Provenaceability Example

2.3 Traceability

Traceability concept aims to give the user the ability, to trace her data or not. This property is examined from the proper/improper data erasure aspect, which is a major problem in web-based systems and still continues to exist in clouds. Many cases have been documented for privacy violation due to improper data deletion (documents, photos, etc.). The traceability concept aims to protect privacy, through the ability of tracing them among the data repositories and reassuring that the data have been completely deleted or maintained invisible and anonymized after their deletion[1]. The clients should be able to trace the physical location of their data and to be able to verify that they are processed according to their collection purpose. The specific concept is matched with the CSA threats, Malicious Insiders, Data Loss or Leakage, Regulatory Compliance, Data Location.

[1] In some cases, certain cloud service providers apply retention policies as far as data are concerned. That means that for several reasons, that are stated inside the contract between the cloud provider and the client, the data remain at rest after the clients deletion request for some time and are strictly accessed form specific personnel and only for certain purposes.

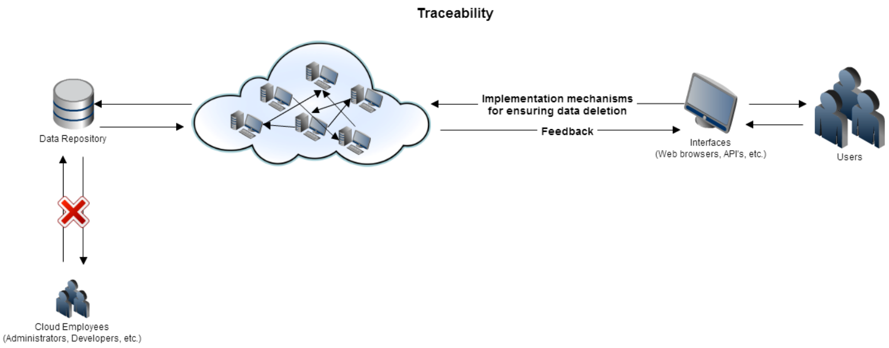
Traceability
Data Repository
Implementation mechanisms for ensuring data deletion
Feedback
Interfaces (Web browsers, API's, etc.)
Users
Cloud Employees (Administrators, Developers, etc.)

Fig. 3. Traceability Example

2.4 Intervenability

Intervenability concept is referred to the fact that, the users should be able to have access and process their data despite the cloud's service architecture. A cloud vendor may rely on other provider's subcontractor services in order to offer her services. That should not be an obstacle for the user to intervene[2] with her data in case she suspects that her privacy is violated by the subcontractors. In fact cloud vendor must be able to provide all the technical, organizational and contractual means for accomplishing this functionality for the user including all respective subcontractors that the vendor cooperates and interrelates [5]. The same applies for the situation that a cloud vendor or the subcontractors are bankrupted and client's data are moved to another provider. The specific concept is matched with the CSA threats, Unknown Risk Profile, Data Location.

2.5 Accountability

Accountability concept is referred to the fact that cloud providers should be able to provide at any given time information about their data protection policies and procedures or specific cloud incidents related to users' data. The cloud architecture[3] makes a complex form of an information system. In terms of management and audit controls, this fact could result in very difficult manageability of the protections mechanisms and incidents. In case of a privacy violation, a cloud provider should be able in any given time to provide information about what, when and how an entity acted and which procedures followed to tackle it [5]. The specific concept is matched with the CSA threats, Abuse and Nefarious Use of Cloud Computing, Insecure interfaces and APIs, Malicious Insiders, Shared technology issues, Data Loss or Leakage, Account or Service Hijacking Unknown Risk Profile, Privileged user access, Regulatory Compliance, Data Location, Lack of Data Segregation, Lack of Recovery, Investigate Support, Long-term Viability.

[2] Access, rectification, erasure, blocking and objection.

[3] International services residual.

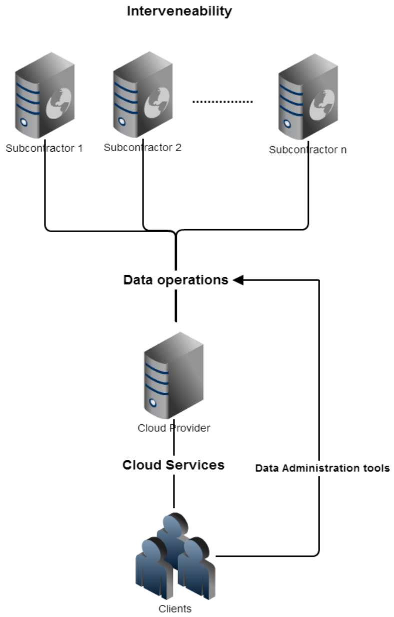

Fig. 4. Intervenability Example

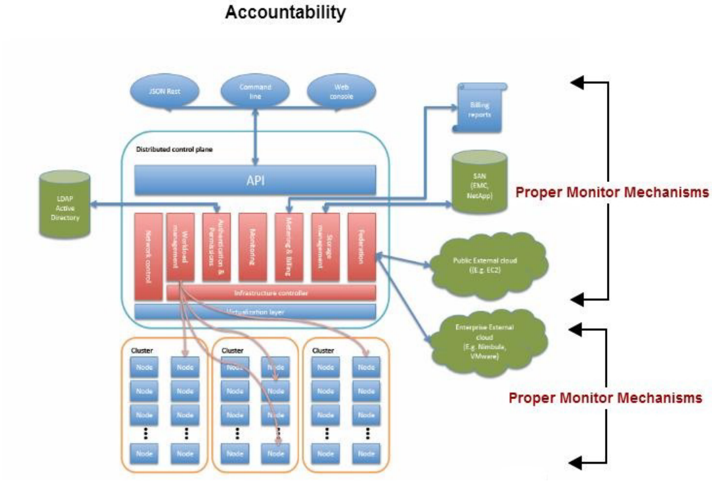

Fig. 5. Accountability Example

Table 1. Matching Security and Privacy Properties with Cloud Services Models

	IaaS	SaaS	PaaS
Property #1: Isolation	x	x	x
Property #2: Provencability	x		
Property #3: Traceability		x	
Property #4: Intervenability	x	x	x
Property #5: Accountability	x	x	x

In table 1 a matching between the aforementioned concepts and the cloud service models is presented. Based on the aforementioned table analysts can identify which the privacy concepts are, that belong to their system, and how these concepts can constitute an initial obstacle during the design of the information system on a cloud environment. Usually when analysts consider the cloud deployment scenario their main goal is to decide on which service model they are interested in deploying to. The identified concepts and the respective matching is a start for creating a holistic process for assisting analysts on receiving the proper decisions.

3 Implementation Techniques

3.1 Data Filtering Techniques (Firewalls)

A firewall is a security guard that is placed between an internal[4] and an external environment. The functions that constitute this mechanism on a simple form are two, data filtering and acceptance or rejection of incoming and outgoing packets. In our case privacy preservation is ensured through the implementation of filtering techniques that aim to achieve isolation between two virtual machines inside a virtual network, through the analysis and detection of malicious traffic that is sent to and from a virtual machine (vm) through the router. Recent editions of firewalls are implementing intrusion detection and prevention inside their core functions, which is a pro in privacy preservation.

3.2 Encryption Mechanism

Encryption mechanisms are used in order to ensure the secrecy of important information [11] inside the cloud environment. Encryption techniques are implemented in various areas of the cloud, in order to encrypt data flow[5] or data at rest[6] and thus protecting privacy through ensuring strong isolation and anonymization of sensitive data [16, 17].

[4] The environment that needs to be protected from the external environment.

[5] Virtual and physical networks.

[6] Databases.

3.3 Hypervisor Hardening, Language, Sandbox, Virtual Machine, OS – Kernel, and Hardware Based Isolation

All the above mentioned implementation techniques provide logical isolation between different entities, procedures and operations inside the cloud. Two types of isolation are implemented, logical and hardware based isolation. Logical isolation is achieved from the first five techniques and attempts to seal all the procedures, operations and the data that flow through the installation of multiple isolation layers between cloud parts, with different programming techniques, inside the cloud environment. On the other hand, hardware based isolation is achieved through hardware controls and it's provided by the processors or by special components combined with the processor [18].

3.4 Privacy Policies and Contracts

Appropriate privacy policies and contracts that benefit client's interest as far as privacy protection is concerned. Cloud users must be very careful about the terms and conditions of the service they are using in order to ensure that their privacy is not violated in case of an incident or a situation that needs to be cleared, e.g. data hosting in foreign countries, what happens in case the cloud provider is bankrupted etc. [5, 7, 17].

3.5 Forensics

Forensics mechanisms are essential in case of an incident, in order to be determined under what circumstances the incident occurred and who is responsible. For example in cloud computing is important to know the origins of the processed data or the detection of fault and security and privacy violations provenance [1, 14, 15, 17].

3.6 Identity Management (IdM)

In this category fall technologies that the use of them combined or individually protect the client's privacy through a solid identity system. Certain techniques that constitute this category are biometrics, smart cards, permission management components, etc. All the above mentioned are techniques that can protect privacy through a solid isolated virtual system and detecting the provenance of certain actions and propably prevent them because of the proper defined identity inside the system [2,6].

3.7 Data Tracking

Data tracking techniques are referred to the technologies that enable data tracing processes in order to inform the client about the route path that their data have followed, where they are hosted at the current time and in what state they have been. Privacy is ensured through the fulfillment of the provencability and traceability requirement that detect the provenance of the data and the provide information about whether client's data are deleted or not and where are located [2, 17].

3.8 Process Operation Identification and Validation

The identification and validation of the processes that modify data is essential to whether the outcomes are reliable or and ensure that privacy is not violated by malicious processes [17].

3.9 Privacy Preserving Data Mining

Most of the times service providers are using collected data from the users, e.g. data traffic, search history, configurations, in order to examine them and make a customer profile for marketing purposes. The fact that personal data are examined is considered as a privacy violation if it's not done properly. This kind of procedures should provide basic anonymization through the data analysis in order for the client's privacy to be ensured [19].

3.10 Monitor and Auditing

Monitor and auditing techniques are widely known and used in order to preserve privacy through monitoring and auditing functions and procedures that occur to an informational system. Monitor and audit procedures are incorporated into security tools and help protecting client's privacy and provide information as to who is accountable about something inside the cloud environment [10].

Table 2. Matching Security and Privacy Properties with Implementation techniques

	Data Filtering (Firewalls)	Encryption	Hypervisor Hardening	Language based Isolation	Sandbox Isolation	Virtual Machine Isolation	Os – Kernel based isolation	Hardware – based Isolation	Privacy Policies and Contracts	Forensics	Identity Management (IdM)	Data Tracking Techniques	Process operations identification and validation	Privacy Preserving Data Mining	Monitor and Auditing
Property #1: Isolation	x	x	x	x	x	x	x	x			x				
Property #2: Provencability									x	x	x	x	x		
Property #3: Traceability		x							x			x			
Property #4: Intervenability									x						
Property #5: Accountability	x	x	x							x	x			x	x

The techniques described above, fall in the category of privacy-enhancing technologies since their main focus is on realizing privacy related concepts as the ones identified in this work. However, these technologies focus on the software implementation alone, irrespective of the privacy issues as well as the cloud services on which the respective software system will be based upon. Thus, this matching aims on providing an initial step of how to bridge the gap between the main privacy concerns and the respective technologies used specifically for cloud environments.

On the other hand, security and privacy requirements methodologies, which address early stages of system design, focus on privacy-related organisational requirements, but do not link these requirements to implementation solutions. Following a number of concepts for understanding the relationship between the user needs in the organisational domain and the capabilities of the supporting software systems is of critical importance and this paper takes an initial step to this direction.

4 Conclusions

The various innovations that cloud computing introduced in its operational environment vary from the traditional "trusted" environment where today's information systems rely on. These innovations hinder new privacy concepts that need to be identified in order to protect the design and implementation of new information systems or even for traditional systems when migrating on cloud environments. Based on this, the specific paper presents an initial effort on identifying the basic privacy-oriented concepts that need to be considered when designing information systems for the cloud. Also, it moves one step further by bridging the gap between design and implementation phases by suggesting a number of privacy-enhancing technologies specifically for the cloud environments. The contribution of this paper can be adopted by a traditional privacy requirements engineering approach in order to be enhanced with the respective concepts aiming on the realization of an approach that deals with the design of cloud oriented systems as it is conducted with the traditional ones respectively. This is also the main future extension for our work. Specifically future steps include the transformation of these concepts on technical requirements and the design of a modeling process for applying these requirements on a real case scenario.

References

1. Cloud Security Alliance, Top Threats to Cloud Computing V1.0, https://cloudsecurityalliance.org/topthreats/csathreats.v1.0.pdf (retrieved September 22, 2012)
2. Heiser, J., Nicolett, M.: Assessing the Security Risks of Cloud Computing, white paper, Gartner group, ID Number: G00157782 (June 3, 2008)
3. Subashini, S., Kavitha, V.: A survey on security issues in service delivery models of cloud computing. Journal of Network and Computer Applications 34(1), 1–11 (2010)
4. Cloud Security Alliance, Security Guidance for Critical Areas of Focus in Cloud Computing V3.0, https://cloudsecurityalliance.org/guidance/csaguide.v3.0.pdf(retrieved September 22, 2012)
5. Draft, EU Directive for Security issues in Cloud Computing (2012)
6. Kalloniatis, C., Kavakli, E., Gritzalis, S.: PriS Methodology: Incorporating Privacy Requirements into the System Design Process. In: Mylopoulos, J., Spafford, G. (eds.) Proceedings of the SREIS 2005 13th IEEE International Requirements Engineering Conference – Symposium on Requirements Engineering for Information Security. IEEE CPS, Paris (2005)

7. Kalloniatis, C., Kavakli, E., Gritzalis, S.: Addressing privacy requirements in system design: The PriS method. Requirements Engineering 13(3), 241–255 (2008)
8. Kalloniatis, C., Kavakli, E., Kontellis, E.: PRIS tool: A case tool for privacy-oriented Requirements Engineering. Journal of Information Systems Security 6(1), 3–19 (2010)
9. Kavakli, E., Kalloniatis, C., Loucopoulos, P., Gritzalis, S.: Incorporating Privacy Requirements into the System Design Process: The PriS Conceptual Framework. Internet Research, Special issue on Privacy and Anonymity in the Digital Era: Theory, Technologies and Practice 16(2), 140–158 (2006)
10. Kalloniats, C., Kavakli, E., Gritzalis, S.: Dealing with Privacy Issues during the System Design Process. In: 5th IEEE International Symposium on Signal Processing and Information Technology, Athens, Greece, December 18-21, pp. 18–21 (2005)
11. Mouratidis, H., Kalloniatis, C., Islam, S., Huget, M.P., Gritzalis, S.: Aligning Security and Privacy to support the development of Secure Information Systems. Journal of Universal Computer Science (2012)
12. Mouratidis, H., Giorgini, P.: Secure Tropos: A Security-Oriented Extension Of The Tropos Methodology. International Journal of Software Engineering and Knowledge Engineering (2007)
13. Mouratidis, Giorgini, P.: Security Attack Testing (SAT) - testing the security of information systems at design time. Inf. Syst. 32(8), 1166–1183 (2007)
14. Wei, L., et al.: Managing Security of Virtual Machine Images in a Cloud Environment (2009)
15. Microsoft Technical report: Privacy in the cloud computing era, a Microsoft perspective, Microsoft Corp, Redmond, USA (November 2009)
16. Sonehara, N., et al.: Isolation in Cloud Computing and Privacy – Enhancing Technologies – Suitability of Privacy – Enchancing Technologies for Separating Data Usage in Business Processes, National Institute of Informatics, Chiyoda-Ku, Tokyo (2005)
17. Zhang, O.Q., et al.: How To Track Your Data: The Case for Cloud Computing Provenance, HP Laboratories, HPL-2012-11 (2012)
18. Viswanathan, A., Neuman, B.C.: A survey of isolation techniques. Draft Copy, University of Southern California
19. Singh, M.D., et al.: A cryptography based privacy preserving solution to mine cloud a data, Infosys Technologies Limited, Bangalore, India (2010)