Clarifying Privacy in the Clouds

Karthick Ramachandran, Thomas Margoni, Mark Perry
Faculties of Science and Law
University of Western Ontario
London, Canada
{*kramach , tmargoni, mperry*}*@uwo.ca*

Abstract—**Concomitant with the increased market appeal of cloud-based services, there is growing concern over issues of privacy within the architecture. In this paper, we analyze what is meant by the term privacy from a legal perspective, and how the meaning of cloud computing and their operation may be affected in at least one jurisdiction. We also look at some possible solutions to addressing privacy in clouds.**

Keywords-**Privacy, cloud computing, compliance with legislation**

I. Introduction

Cloud computing represents a relatively recent architecture and business model in the information technology environment. It is a term that describes having data processed, stored or retrieved in a cloud, where 'cloud' means somewhere on the Internet. The Internet is a very generic term, especially when we are interested in knowing the physical location of data or a particular server. Saying it is in the Internet or in a cloud means that, most of the times, we don't know, or don't care, from a computing perspective where it is. A main feature of cloud computing is that for operational purposes the cloud users are not interested in their location. This is extremely advantageous from a technical perspective as from that viewpoint one needs the job to get done without having to worry about availability of resources. However, from a legal perspective it raises many problems, not least that it is increasingly an issue in most jurisdictions that companies address privacy requirements and comply with privacy regulation. Location is a key factor that must be considered. Here, we are addressing the issues of location-independent computing, such as is part of the fundamental design of cloud computing, in terms of privacy legislation, in order to determine high non-compliance situations, and identify some possible approaches to provide solutions, and best practices. We take Canada as the use case for this analysis.

We present an introduction to Privacy in Section II. In Section III, we look at data protection laws, specific to Canada. An introduction to Cloud Computing and its architectural details are described in Section IV. Section V enumerates some threats to data stored in remote servers. We discuss some of the technical approaches to protect user's privacy in a Cloud computing environment in Section VI. And finally we conclude in Section VII.

II. What is Privacy

Early interpretations of legislation in England outlawed eavesdropping and spying on others. The English courts in deciding about the granting of a warrant to "break open doors, locks, and boxes, and to seize a man and all his books" have held that "we can safely say there is no law in this country to justify the defendants in what they have done; if there was, it would destroy all the comforts of society, for papers are often the dearest property any man can have" [1]. Many international treaties, covenants and declarations recognize privacy as a fundamental human right, such as in Article 8 of the 1950 Convention for the Protection of Human Rights and Fundamental Freedoms "(1) Everyone has the right to respect for his private and family life, his home and his correspondence. (2) There shall be no interference by a public authority with the exercise of this right except as in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health of morals, or for the protection of the rights and freedoms of others". A simple condensing of the privacy as an autonomous right was summed up as the "right to be let alone" [2]. However, it must be said that the same authors recognized that technological evolution carries with new threats that were previously protected by other methods. So, for example, the old tort of trespass (vis et Armis, in its origins around 13th century) can be seen and adapted to protect privacy to some extent, until when in order to know what was happening in the seclusion of a house it was necessary to enter the property of the house owner, i.e. to trespass his own property. No doubt that such a remedy's main function was to protect the person and the property of the owner. However, it was also capable of protecting something that was not identified yet.

Advances in technology, in their very nature, enable actions previously not possible, so it is possible to see through walls with infrared cameras and store such images on computers for easy replication and distribution. The so-called 'smart' electricity meters enable the power supply company to know what and when equipment is being turned on and off inside your house, and possibly build a 'power' profile of a customer. No physical invasion of land has occurred. It is also possible to determine whether somebody is growing marijuana plants in his basement without entering the property [3]. Or, listening to private conversations without eavesdropping at the door but by capturing electronic communications [4][5][6]. All those activities that in past were not allowed because the old technology required an action that was considered illegitimate (eg: entering the property without a warrant), have become available because technology now permits to carry out the activity without performing the prohibited action.

Although the concept of privacy and data protection has developed over the centuries to the point where most countries now have legislation regulating these issues, new technologies such as Deep Pocket Inspection, or traffic sniffing, and sophisticated listening and imaging tools, pose an enormous threat on the protection of privacy and personal data.

III. Privacy and Data protection in Canada

Canada does not have a generally accepted tort of invasion of privacy at common law, although in the USA several approaches by the courts have lent strength to privacy protection. However, Canadian legislation address privacy and data protection with regard to different conditions such as the nature of the obliged subject (public or private), and the type of activity carried out (commercial or not), and the sector within which the activity is being carried out (for example the extensive regulation of health services). The Canadian Charter of Rights and Freedoms is entrenched in Canada's constitution. Although it does not specifically give a right to privacy, it does protect citizens from unreasonable search and seizure by the state. This general principle has been interpreted by the Supreme Court of Canada (SCC), which stated that rights should be interpreted in a broad and liberal manner so as to secure the citizen's right to a reasonable expectation of privacy against governmental encroachments and intrusions [7]. The SCC went further by stipulating that privacy should be at the core of modern societies, and referring a previous Canadian Government Study on Privacy and Computers (that dates back to 1972), in a way that seems to suggest the applicability of the concept of privacy to informational aspects as well:

> *"This notion of privacy derives from the assumption that all information about a person is in a fundamental way his own, for him to communicate or retains for himself as he sees fit"*[8]

During the early 90s the SCC took the opportunity to develop the concept of privacy with regard to governmental intrusions, including the use of then new technologies [9] [10] [11]. However, it must be observed that these cases are based on governmental intrusions, mostly prosecutions looking at the extent to which citizens should be protected from unreasonable measures.

Regarding a different but connected area of collection of data by Federal agencies the reference legislation is the Privacy Act of 1982 [12]. Such legislation main aim is the regulation of the collection and use of personal information by the federal government and a number of federal public agencies. Such statute is coupled with another piece of federal legislation that is geared towards the accessibility by citizens of information stored by government agencies [13]. Both the Privacy Act and the Access to Information Act refer to personal data or records retained by the Federal Government, thereby identifying records that can be either under an analogical or a digital expression form. It must be noted how such legislation refers only to federal bodies, and that on a provincial level similar legislation has been enacted (For example in the [14], [15] and [16]).

Looking to the private sector, the most relevant piece of legislation in Canada, and probably the more relevant in light of this study, is without doubt the Personal Information Protection and Electronic Documents Act of 2000, also referred to as PIPEDA [17], that applies to all private sector entities that collect, use, or disclose personal information in the course of commercial activities (with the exception of those provinces that have enacted equivalent legislation). PIPEDA as many other piece of legislation around the world is generally based on the Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data of 1980 promulgated by the OECD [18]. Those principles, as enacted by PIPEDA may be summarized as follows:

Accountability: the collecting organization is responsible for the collected data

Identifying purpose: the purpose for collecting personal information shall be identified before the information is collected

Consent: individual's consent is required for the collection or disclosure or personal information

Limiting Collection: the collection of data should be limited to those data necessary for the purpose of collection

Limiting Use, Disclosure and Retention: the collected personal information should be used only for the purposes for which it was collected

Accuracy: collected personal information should be accurate and complete. It is a collecting organization duty to maintain such information updated

Safeguards: the personal information collected shall be protected by measures appropriate to the sensibility of the data collected

Openness: the information regarding the organization privacy policies should be readily available to users

Individual Access: individuals shall access the information retained by an organization regarding such individual, and shall also pretend that the information be amended if not correct

Challenging compliant: the individual shall be able to address a challenge regarding the compliance of those principles to the organization designated individual [19].

The implementation of such basic principles that have their roots in the OECD guidelines is particularly important. In fact, jurisdictions such as the European Union, forbid the transmission of personal data when the destination of such flow is a jurisdiction with not acceptable levels of privacy protection, and this has caused some disruption of data trade between the European Union and the United States of America.

IV. Cloud Computing - Introduction

Cloud computing is the style of computing in which the users can rent infrastructure, platform or software services from other vendors without requiring the physical access to them. It divides the responsibilities of managing technologies between two different stakeholders who can be geographically situated in different corners of the world. Owing to this advantage, the cloud computing has been widely adopted. MarketsandMarkets estimates [20] Cloud Computing market will increase from $37.8 billion in 2010 to $121.1 billion in 2015 at a compound annual growth rate of 26.2 percent.

Figure 1 presents the evolution of cloud computing. Early on in the development of the internet there were computers that connected to the internet using dial-up, ISDN, T1 or T3 lines. They were then replaced by powerful servers at the (TCP) Internet access points. A single server was then replaced by a rack of servers for power hungry applications. Later the same rack of servers were shared between two or three applications and users, to optimize its usage of services. Moreover, a new paradigm of software as a service evolved where standalone desktop applications were slowly moved to powerful servers for ubiquity and more reliability. Cloud computing evolved out of this stage, where multiple vendors can dynamically provision resources based on their requirements and the resources allocated to them can grow or shrink like an elastic.

Cloud computing (access) can be implemented in three different models (Figure 2); Infrastructure as a Service, Platform as a Service and Software as a Service.

In Infrastructure as a Service, the users can rent the physical/virtual machines from the cloud computing vendor and the user installs the basic software in the machines. The cloud service provider (CSP) can also expose some of the machine renting capabilities as an public API, which can be utilized by the users for dynamic provisioning [21]. Amazon, Rackspace and Slicehost are some of the popular providers of Infrastructure as a Service.

In Platform as a Service, the provider encourages the users to develop their application using the platform provided by the CSP (eg: Google App Engine, Microsoft Azure). The users, while developing their applications using the platform, need to only worry about the expression of their application through the platform. The provider optimizes their infrastructure for the platform and the application once installed in the platform will seamlessly scale and the scalability will be the responsibility of the CSP.

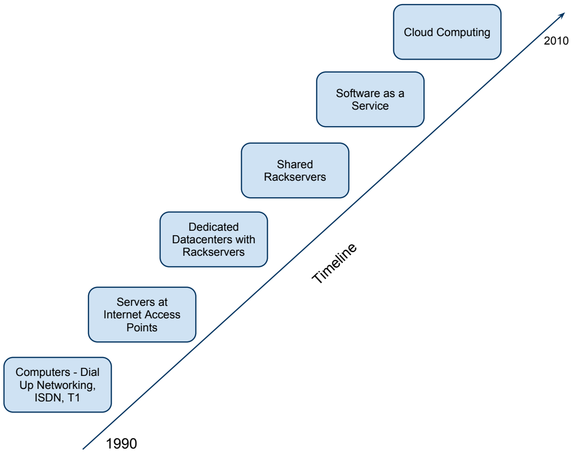
Cloud Computing
2010
Software as a Service
Shared Rackservers
Dedicated Datacenters with Rackservers
Timeline
Servers at Internet Access Points
Computers - Dial Up Networking, ISDN, T1
1990

Figure 1. Evolution of Cloud Computing

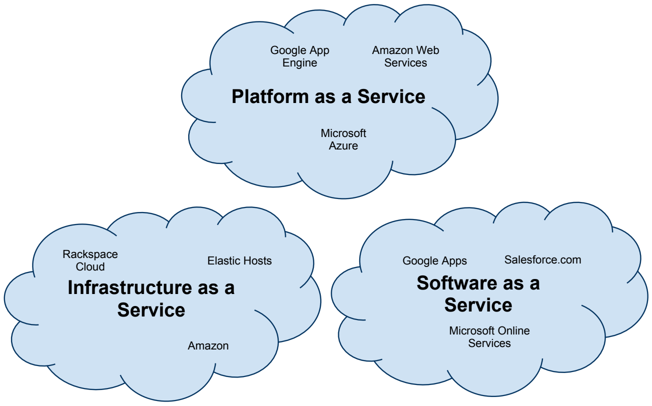

Figure 2. Cloud Architectures

In Software as a Service, the provider implements the software for the client and then provides a virtual container to host the client specific data in the software (eg: Google Docs, Salesforce etc). In this case, the client needs neither to have the technical expertise to host the application or scale nor the expertise to develop the application. The CSP uses its infrastructure to provide the services to the client.

In all the above models, the CSP is responsible for hosting the users data. The user loses the control of the data once it reaches the CSPs data grid. The user is entrusting the provider with data, because either the user has no infrastructure to host the data by themselves or assumes the data will be reliably stored in the cloud providers infrastructure as the cloud provider is trusted to have the necessary expertise to reliably store the data.

This exposes one of the major issues with Cloud computing. Cloud computing paradigm requires disturbing levels of trust by users in the servers that hold their information. Unless there is some means of totally obfuscating the data, the user needs to trust that the data stored by the CSP will be used by them only for the purposes for which it is intended to be used.

V. Threats to Data stored in CSP

There are variety of ways the datas privacy or security can be compromised in a cloud computing environment [22]. Some of them are the following:

A. *Sharing of data with an unauthorized party*

Cloud provider could compromise the confidentiality of the data by sharing the data stored in the system to unauthorized parties. This can go against the terms and conditions of the service and will qualify as the breach of security and contract. The end user could never be aware of such a breach, even if it happened.

B. *Corruption of data stored*

As the cloud computing provider has root access in the physical machine, they will have rights to modify/delete the data. Cloud provider can tamper with the data making the data non-usable or modify the data in a way that system cannot detect the modification. This poses serious threat to the correctness of the application.

C. *Malicious Internal Users*

The employee of a cloud computing provider who has root access to these physical machines, can easily access the data and use it for their advantage.

D. *Data Loss or Leakage*

When a virtual machine is used in an infrastructure, it poses a variety of security issues [23] which could lead to a compromise. Moreover, when the facility which hosts the user's data is subjected to a natural calamity, that would risk the loss of the user's data.

E. *Account or Service Hijacking*

Another risk for the cloud computing provider is, if the service is hijacked, or the computer is hacked by a hacker, the hacker will have full access to the data. As the cloud infrastructure is not under the client's control, it could be more prone to attack as the risk profile of the infrastructure will be unknown to the client.

To summarize storing the data in the cloud, can increase the privacy risks for the following stakeholders:

1) Cloud Computing User
2) Organization using the Cloud Service
3) Implementors of Cloud Platforms
4) Providers of application on top of cloud platforms
5) For the data subject

VI. Approaches to Addressing Privacy Issues in Cloud Computing

There are variety of ways in which the user can ensure that data is protected from the cloud computing provider or the cloud computing provider is made accountable for the data stored. Privacy Enhancing Technologies (PET) can be used by the developers of the application to enhance the privacy of individuals in an application development environment.

Some of PET include:

1) Privacy management tools that enable inspection of server side policies about handling of personal data
2) Secure online access mechanisms to enable individuals to check and update the accuracy of their personal data
3) Anonymizer tools which will help users from revealing their true identity by not revealing the PII (Privately Identifiable Information) to the CSP.

A. *Privacy By Encryption*

Privacy can be enforced by encrypting all the data that is stored in the CSP. The main issue with that architecture is that the cloud provider can be only used for storage of the data. As the data will be unrecognizable for CSP, it will not be possible for CSP to process the data or perform some number crunching tasks on it.

Searchable encryption employs an algorithm that allows users to encrypt the data and then provide the server with trapdoor information [24], so that the server can search for a given string through searchable encryption algorithm. Public Key Encryption with Keyword Search (PEKS) [24] is one of the seminal works in the area of making encrypted data searchable. The authors of PEKS propose to encrypt the message using the Public-Private key infrastructure. Along with this cipher text a Public-Key Encryption with Keyword Search (PEKS) of each keyword is append to the final message. The PEKS has the trapdoor information, which is the extra information sent to the server along with the encrypted keyword for the server to test for the existence of a keyword. Searchable encryption research is at its nascent stage and it is limited only to exact word searches for now.

B. *Privacy By Secure Computation*

Another way to perform computation in the server in a secure way is using secure computation algorithms. The secure computation algorithms enables users to compute use the infrastructure from a insecure environment for computation without revealing the exact input for the computation. Yaos protocol [25] provides some of the basic techniques to perform a computation in a secure way without revealing the inputs. Yaos protocol forces the expression of a computation problem in terms of logical circuit using gates. The input of each gate is randomly encrypted and then then final resulting output is decrypted to get the exact answer of the computation. The encryption and the decryption is done at the clients end. The expression of a simple problem using

Yaos protocol is found to be complex. Hence it still resides in the theoretical realm.

C. Privacy By Using Secure Coprocessors

Secure coprocessors are currently the only realistic way to perform general-computing even when the adversary had direct physical access to the device (in our case adversary can be the CSP itself). It is a very limited computer with its ROM, RAM and battery backup for persistent storage and an ethernet card. When installed in a computer, they can be seen a secure area inside a computer that even the main processor cannot access. Privacy as a Service [26] recognizes these factors and proposes a system architecture in which a coprocessor is installed in every cloud computing system. The data loaded into the cloud is classified based on its significance and security by the cloud user (No Privacy, Privacy with Trusted Provider, Privacy with Non-Trusted Provider). The data tagged with Privacy with Non-Trusted Provider level is processed by the secure coprocessor.

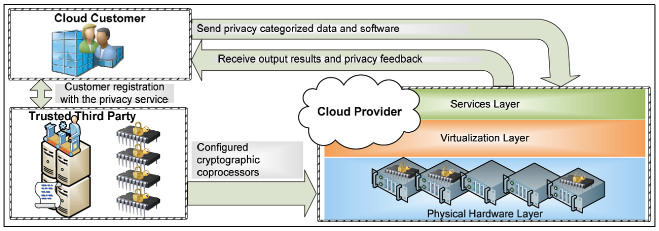
Cloud Customer
Send privacy categorized data and software
Receive output results and privacy feedback
Customer registration with the privacy service
Trusted Third Party
Cloud Provider
Services Layer
Virtualization Layer
Configured cryptographic coprocessors
Physical Hardware Layer

Figure 3. System Model for Privacy by Secure Coprocessors [26]

Figure 3 [26] is an example of a system built using secure coprocessors. Cloud customers, Trusted Third Party and the Cloud Provider are the three main stakeholders of this system. The coprocessor is signed by secret keys by the trusted third party and then is supplied to cloud provider. When a new customer registers with the cloud provider, they share the secret keys with the trusted third party. The co-processors can directly contact the trusted third party for the keys to encrypt the secret data within the coprocessor. The data channel between the co-processor and the trusted third party is secured using a mutually agreed upon public/private key pair during the initial time of supply of co-processors to CSP by trusted third party.

One secure coprocessors cost in the order of hundred thousands, even though PaaS provides some reasoning for the economics behind using them in CSP's machines, for now it looks highly unrealistic to use a coprocessor in the server infrastructure.

We discussed the technical options available to protect user's privacy by having minimum or no trust with the cloud service provider. In all the solutions we noted down the inability of these models to be used in the current cloud environment.

There is a pressing need for the law to provide legal protection to the cloud clients, as they need to trust the cloud provider with their confidential data.

VII. Conclusion

We have discussed some of the issues that confront cloud providers and users, in particular when facing the growing requirement for privacy of data in a growing number of jurisdictions. Although some partial privacy solutions have been suggested, it is unlikely that any of these can be adopted by the providers in the current cloud environment. We are working to develop other approaches to securing privacy for users of clouds, and ensuring that the dangers presented in the clouds are transparent to such users.

VIII. Thanks

The authors thank the Social Science and Humanities Research Council of Canada and the IBM Center for Advanced Studies for their support.

References

[1] *Entick v. Carrington*. 1558-1774 All E.R. Rep. 45.

[2] S.D. Warren and L.D. Brandeis. The right to privacy. *Harvard Law Review*, pages 193–220, 1890.

[3] Kyllo v. United States, 533 U.S. 27 (2001).

[4] T Nabbali and M Perry. Going for the throat: Carnivore in an echelon world. part 1. *Computer Law and Security Report*, 19(6):456–467, 2003.

[5] T Nabbali and M Perry. Going for the throat: Carnivore in an echelon world. part 2. *Computer Law and Security Report*, 20(2):84–97, 2004.

[6] T Nabbali and M Perry. Introducing carnivore: Going for the throat with precision surveillance, TLF v3 0031 (2004).

[7] R. v. Dyment, [1988] 2 S.C.R. 417.

[8] Department of Communications and Department of Justice. Privacy and computers: A report of a task force. Information Canada, Ottawa, 1972.

[9] R. v. Wise, (1992), 70 C.C.C. (3d) 193 (S.C.C.).

[10] R. v. Wong, (1990), 60 C.C.C. (3d) 460 (S.C.C.).

[11] R. v. Duarte, (1990), 53 C.C.C. (3d) 1 (S.C.C.).

[12] Privacy act, (R.S., 1985, c. P-21).

[13] Access to information act, Act R.S.C. 1985, c. A-1.

[14] Ontario province: Freedom of information and protection of privacy act, (1988).

[15] Municipal freedom of information and protection of privacy act, (1991).

[16] Personal health information protection act, (2004).

[17] Personal information protection and electronic documents act, (2000, c. 5).

[18] The Organization for Economic Co-Operation and Development. Guidelines on the protection of privacy and transborder flows of personal data.

[19] Takach G. *Computer Law*, pages 329–330. Irwin Law, 2nd edition (July 2003).

[20] MarketsandMarkets.com. Cloud computing market - global forecast (2010 -2015).

[21] Amazon.com. Amazon ec2 webservices.

[22] Cloud Security Alliance. Top threats to cloud computing v1.0.

[23] T. Garfinkel and M. Rosenblum. When virtual is harder than real: Security challenges in virtual machine based computing environments. In *Proceedings of the 10th conference on Hot Topics in Operating Systems-Volume 10*, page 20. USENIX Association, 2005.

[24] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano. Public key encryption with keyword search. In *Advances in Cryptology-Eurocrypt 2004*, pages 506–522. Springer, 2004.

[25] A.C. Yao. Protocols for secure computations. In *Proceedings of the 23rd Annual IEEE Symposium on Foundations of Computer Science*, pages 160–164. Citeseer, 1982.

[26] W. Itani, A. Kayssi, and A. Chehab. Privacy as a Service: Privacy-Aware Data Storage and Processing in Cloud Computing Architectures. In *2009 Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing*, pages 711–716. IEEE, 2009.