Data Privacy Issues in Cloud Computing

S. Srinivasan
Texas Southern University, Houston, Texas, USA

Abstract

Cloud Computing is widely used globally today. It has many benefits to offer. However, two of the major concerns for cloud users has been the issue of data security and privacy for data stored in the cloud. Centralized data attracts the attention of government regulators and hackers because of the availability of information in one source. One of the benefits of cloud use is the ability to modify business processes based on data. The main thrust of this paper is to show that when business processes involve the use of cloud then there is the potential for loss of privacy but other resources available in the cloud should help with data protection. We will address this aspect extensively in this paper and suggest some solutions to preserve privacy of individuals and organizations. One aspect of this involves the use of Big Data in the cloud. We will point out that the benefits of using the cloud for large volume storage and handling multiple sources of data. One of the issues with combining data from multiple sources is the ability to re-identify individuals from anonymous data. We point out how privacy is significantly affected when re-identification occurs.

1. Introduction

Cloud computing use is growing significantly today. All the major cloud service providers (CSPs) such as Amazon, Google, Microsoft and Rackspace are from US. One of the laws in US is the USA PATRIOT's Act enacted in the wake of the 2001 attacks. Many foreign governments are concerned that if the CSPs store their data in US then the US government might get access to such data. This would violate the privacy of individuals of foreign nations. This concern resulted in the invalidation of Safe Harbor Agreement between US and Europe. Safe Harbor provided a mechanism for CSPs to transfer data between Europe and US. The new agreement replacing Safe Harbor was agreed upon in 2016. It is called Privacy Shield. As the name suggests, protecting people's privacy is of high importance when it comes to the use of cloud services. As mentioned earlier, cloud offers many benefits to the user in the form of greater availability, access to high-end computing resources, ability to meet elasticity of demand and pay for what is used. People tend to store plenty of personally identifiable information (PII) online for convenience. Often such data are meant for use by closest friends of the users. However, once data is stored electronically, many try to get to it by one means or another. This has resulted in many hacks that have successfully exfiltrated stored data. However, users of online resources have come to expect privacy of their online activities.

Today, technology provides several ways to monitor all online activities, whether it be web surfing, social media communication, email or phone. Many phone services are internet-based now and as such becomes part of the information collected of online activities. Storage has always been cheap and technological advancements have made it possible to store very large volumes of data at an affordable cost.

This has resulted in storage of all types of metadata related to online activities of individuals. Almost all of this storage is in the cloud. The Cloud Service Providers (CSPs) have adapted their business practices to create new Data Centers in various cities within US as well as in several other countries to meet the growing need, based on laws that affect storage of data outside of certain geographical region. Companies like Amazon Web Services (AWS) offer storage at a very cheap cost - 1 cent per gigabyte of storage for infrequent access to data. Table 1 provides a quick comparison of storage costs by major CSPs.

Table 1. Comparison of storage costs among CSPs

Provider	Cost/GB/month	Remarks
Amazon	$0.03	Cost increases beyond 4 TB
Microsoft	$0.01	Up to 100 TB
Google	$0.026	Up to 4 TB
Rackspace	$0.012	

Google, Microsoft, Apple, Dropbox all offer free storage space at varying levels for ordinary users. Table 2 provides a summary of free storage spaces. Given this trend in the industry and the availability of cloud storage at such an affordable rate, businesses tend to accumulate more data about its customers

Copyright © 2016, Infonomics Society

from sources such as social media, email and phone conversations.

Table 2. Summary of free storage spaces by CSPs

Provider	Free storage space
Microsoft	5 GB
Google	15 GB
Apple	5 GB
Box	10 GB
Dropbox	2 GB

In the two tables above we have described the cost for storage and the amount of free storage space major CSPs provide. It is important to note that companies often add costs elsewhere such as accessing the stored data or transferring stored data from one location to another. In this latter aspect there is always an incentive to stay with the same provider for cost savings.

Social media is a source of very large volumes of data that are available rapidly. Thus, social media data fits the Big Data feature of volume, velocity and availability. Big Data involves not only having access to a vast collection of data from multiple sources but also the ability to combine them to draw conclusions. Standalone, these data do not pose much of a threat to individual privacy. However, there is an abundance of technological tools that make it possible to combine such data and derive useful information about individuals. Moreover, researchers at MIT have shown recently that even anonymized data when available from different sources help with re-identification of individuals from such data [7]. This was classically illustrated by the identification of Thelma Arnold from South Carolina based on the AOL anonymized data [3]. This is because certain types of data such as voter registration are publicly available and identify individuals' name, address and phone number. When combined with other information such as Driver's License, more personally identifiable information (PII) is made available. By proper association all such information pertaining to an individual are combined and it exposes more information about an individual's preferences [1]. When taken further, this could lead to exposing an individual's medical history which is then a serious violation of privacy. In protecting people's privacy it is important to note that re-identification possibility is a concern. In fact, Australia considers re-identification capability a threat to individual privacy and as such wants to ban re-identification based on government data.

In Srinivasan [10], there is extensive discussion of Cloud Computing techniques. The most common form of such techniques is Software as a Service (SaaS). A slight variation of SaaS usage comes in the form of installing various Apps on a mobile device. For example, installation of various Apps from Google Play gives the user only one option - to accept the choice of sharing contacts, address book and photos. This practice forces the user to accept these requirements in order to use the App. The consequence of this practice is that businesses collect lot more data about the visitors to their website and customers. This enables the businesses to reach beyond the customer by extending their reach to all contacts of the customer. Furthermore, knowing the information about the contacts and the contents of the address book, the business is able to target the customer with targets ads. Sometimes the customers disagree with this practice as it violates their privacy in the form of keeping their friends not exposed to possible targeting for ads. This is one aspect we would explore further in the rest of the paper.

2. Cloud Related Exposure and Services

The main thesis of this paper is the focus on how cloud services impact the business processes. As pointed out earlier, cloud services are well received and it benefits small and medium sized businesses. Large businesses also benefit from the use of cloud because they can subscribe to Infrastructure as a Service (IaaS). Since all types of businesses benefit from the cloud, business processes would be impacted by the availability of computing services. Major cloud service providers (CSPs) focus their attention on availability of services at 99.9% or higher. Maintaining such high availability greatly benefits all businesses. Security aspects are tied to high availability and CSPs are able to provide advanced computing services and advanced monitoring services for security. Given these benefits of using the cloud, businesses have a great opportunity to modify some of their business processes. First, since data gathering in this set up is simplified and the ability to store large volumes of data in the cloud is enhanced, business process redesign should take into account the availability of personnel to deliver the services [2]. Second, various personnel dealing with customers could provide feedback to the system that could be incorporated in the Big Data analytics in arriving at personalized services to customers.

The ability of cloud to store large volumes of data should be viewed as a benefit to Big Data processing. By keeping the cost low for storage, the large volumes of data present an opportunity to use the

Copyright © 2016, Infonomics Society

data judiciously to derive knowledge that could be used to serve the customers better. Thus, Big Data is not a problem in itself but should be used to benefit the customers by modifying the business processes to handle such data. It is worth pointing out in this regard that storage of data is inexpensive but accessing the stored data costs businesses significant sums. Some of the benefits of cloud storage in cost are thus offset by the additional access cost for the stored data. Protecting stored data is important to preserve its integrity. Several studies have shown that the cloud service providers are able to provide greater levels of security than a typical user would be able to afford because of economies of scale.

Availability of cloud services to a business gives them the ability to redesign their services to take advantage of advanced computing capabilities available because of the use of cloud service. In particular, the business can collect all related data from other sources such as social media and email systems about customer transactions. Using Big Data analysis of large volumes of data, the business can precisely know what a particular customer would want as a service. The ability to personalize service to a large group of customers is thus a direct benefit of cloud service and business processes should be changed to accommodate such personalized service. Typically, such personalized service would be well received by the customers. Because of this benefit the business process should be modified to get customer feedback and incorporate that with related data about their service.

Another benefit of cloud service is the ability to meet the elasticity of demand. Business processes should take into account the availability of open-ended resources at an affordable cost. Use of services such as SaaS enable the business to focus more on their core strengths. Since availability of computing resource is no longer a constraint, the business should take advantage of computing power to experiment with new ideas to serve the customers better. Since the cost is usage based, changing business processes to take advantage of newer technologies is advantageous to a business. Cloud service addresses an important business process for every business, namely backup and recovery. Many businesses do not pay enough attention to data backup and recovery because it is time consuming and does not provide immediate benefit until some disaster strikes, which is rare. However, without such backup businesses would become vulnerable. In the Information Systems management, the businesses should not focus on ROI when it comes to assessing the benefits of backup. With the CSP taking care of all the management aspects of data backup and recovery, businesses tend to focus on their strengths and the CSP provides the essential service of backup and recovery when needed.

We mentioned earlier that CSPs are playing a vital role in information security for the businesses. Common perception is that in order to provide security the user must have control over the devices. This usually applies to physical security. Given the elastic nature of demand for service and the centralization of service, the CSPs are in a better position to provide greater physical security to the hardware. Since very high availability is one of their major strengths, CSPs have the ability to deploy better tools for security. Moreover, businesses have the option of selecting the appropriate encryption technology to protect its data in the cloud, both at rest as well as during an application processing. Use of encryption increases both the cost of service and service latency. Since speed is of essence to businesses, often encryption is bypassed. If a business needs a higher level of assurance on security they have the option of selecting a Virtual Private Cloud (VPC) at a higher cost. The only reason for selecting VPC would be to provide the necessary isolation of hardware from other users. CSPs take advantage of their ability to provide many Virtual Machines (VMs) to their customers using a single hardware. In a public cloud the businesses have the benefit of lower cost but do not have adequate control who else resides in that physical hardware through multi-tenancy.

Thus, businesses have the ability to achieve a higher level of security in the cloud than they could afford otherwise under their control of the computing system. In this regard the customers have the advantage of having independent third party providers evaluate the security arrangements of CSPs and provide a quick comparison capability among various CSPs on their security practices. One well respected independent assessor is Cloud Security Alliance (CSA) which provides such comparison data in the form of STAR (Security, Trust, Assurance Registry). CSA also provides businesses with the Cloud Controls Matrix tool to assess their security practices.

Trust is an important component for many businesses. Trust is built over a period of time and it is not something that is acquired by spending resources. However, there are independent third party auditors who provide such assurances to the users [11]. CSPs provide trust in their services by acquiring compliance certification from sources such as the Big Four accounting firms for their security practices. One such is SSAE 16 (Statement on Standards for Attestation Engagements #16). CSPs carry other compliance certifications such as SOX (Sarbanes-Oxley Act), FISMA (Federal Information Security

Copyright © 2016, Infonomics Society

Management Act), HIPAA (Health Insurance Portability and Accountability Act) and GLBA (Gramm-Leach-Bliley Act) to enhance trust in their operations. These in turn help the various businesses meet their compliance requirements while using cloud services. In this regard we point out that Service Level Agreements (SLAs) are also trust builders.

However, the CSPs do not customize the SLAs for businesses because they deal with thousands of customers and it would be impossible to provide different levels of assurances for various users. This is not a flaw of cloud service since CSPs provide very high availability of service and greater level of security than what a typical business could hope to achieve. To overcome any such concern a business might have, there are Cloud Service Brokers (CSBs) who provide enhanced service to customers with an acceptable SLA. CSBs are a growing presence in the cloud service market and businesses have come to rely on them for many service integration as well in addition to better SLA.

Cloud Service Brokers could also help with enhancing business processes. Because they have the ability to collect and process large volumes of data, the CSB will be in a better position to analyze customer interaction data with the business and notice where business processes could be improved. Businesses should take advantage of this possibility.

3. Business Process Redesign using Cloud

Cloud services have freed up businesses to concentrate on the business aspects and leave the issues associated with managing a computing system to the CSPs. This has turned out to be a major benefit to the businesses because they could now reduce their CapEx (capital expenditures) to the much smaller OpEx (operating expenditures) and at the same time have access to a vast array of computing related services [10]. The main advantage of moving business expenditures from CapEx to OpEx is the speed it affords in business processes because CapEx is always more time consuming to execute than OpEx. Moreover, it enables the businesses to pay for what they use only and have the benefit of enlarging or shrinking their computing needs based on demand. This aspect of cloud service enables the businesses to explore new ways in which they could redesign their business processes. For businesses having the ability to retain certain materials such as emails for extended periods of time is easy with cloud service. The cloud enables the business to be able to have different retention periods for different types of data that they hold based on business and legal constraints. One of the essential features for a business is to have access to their data on demand from multiple sources on multiple devices. Today, many users have a need to access data on smaller devices such as cell phones. Cloud service is able to adapt to this need and enable the business to provide such a service to their users. This requires making changes in their business processes in order to authenticate customers using smaller devices that lack high computing power. The lack of high computing capability on smaller devices is overcome by the use of cloud where the authentication need is transferred. Thus, a business is able to adapt to the customer needs by modifying their business processes using cloud.

Another aspect of business process redesign involves the need for having the ability to provide adequate controls to protect the information [8]. Since the level of control needed for different types of data that a business holds varies, the access control mechanisms must be tuned to realize the security levels needed for authentication. Cloud service provides the necessary computing resources to handle the differing access requirements and provide controls based on a controls matrix. This variability fulfills an important business need for security and at the same time meet the customer need for information stored on their systems. The centralization of data in the cloud enables the users to access the data from many locations.

Businesses feel that if they control the hardware then they can control access to the stored data and thus guarantee confidentiality of information stored. However, this is a myth since CSPs have a higher level of security control and access to information in the cloud is still controlled by the access controls that the business wants to use. Hence, modifying the business process to use the cloud for storing all data and control access to that data using processes similar to what they would use in their internal systems is recommended. CSPs simply facilitate access to stored data and it is up to the business to know the sensitivity of data that they store and so they should adopt differential control for data access. One common concern in this regard is the number of privileged users who will have access to such data. Even though certain people within the CSP will have privileged access to the devices in which data is stored, they would not be able to know what data they are accessing for system management purposes. Moreover, a business would be able to obtain from the CSP all users who had access to the data location and who viewed or modified any stored data. Because of the ability to obtain and analyze this log of data access, businesses should feel confident that they still control access to data stored in the cloud and can verify all those who actually accessed the

Copyright © 2016, Infonomics Society

data. Having access to logs would help a business meet its compliance requirements as well.

4. Privacy Concerns in using Cloud Services

Privacy concerns vary depending on the type of business. Businesses involved in healthcare sector are governed by HIPAA (Health Insurance Portability and Accountability Act) requirements for protecting confidentiality of customer data. Moreover, when a business in the healthcare sector is breached they lose valuable Protected Health Information (PHI). This type of information once disclosed will do irreparable harm to an individual's privacy rights. Cloud service provides various options to the business customers to choose the level of protection needed for their data.

The most common of these approaches is encryption. The customer chooses the type of encryption that they prefer and store the encryption key in a safe place under their control. Because cloud services provide storage at a very low cost, the business processes can be designed in such a way that the business can store the different types of data stored at varying levels of security. Since the purpose of all storage is to use such data at some time, businesses should be aware of the Format Preserving Encryption (FPE) methodology to store certain data so that when they are used, any application using such data will be able to use Format Preserving Encryption [5] [4]. Typically, FPE is more expensive than other forms of encryption.

Thus, businesses that plan to use FPE should expect to use higher prices for cloud service. As seen here, the major resource available to protect privacy is encryption. However, businesses realize that use of encryption increases latency of data access. Consequently, they weaken their use of encryption. Since customers are not aware of the details, they get a false sense of security because of the use of encryption. This became evident when the website of Ashley Madison was breached [13]. Even though some customers used strong passwords, Ashley Madison stored all passwords in lower case, resulting in hackers having an easier time accessing such stored information. This particular data breach also showed that the breach need not have a financial motivation.

It can be ideologically motivated as in the case of Sony 2011 breach. Data breaches which lead to privacy violations are costly for the organizations involved. In the case of Ashley Madison, the breach that occurred in 2015 resulted in a judgment in 2016 against the company in US for a record $17.5 million. The company pleaded its inability to pay the entire sum and instead settled the matter with US Federal Trade Commission (FTC) with a fine of $1.6 million [6].

Privacy of information is a precious commodity. Businesses should protect privacy of individuals and organizations fully. Once compromised, it will be impossible to repair the damage. For this reason, websites like PayPal use a salt before encrypting user data. Since PayPal deals with financial information of users, this added protection provides adequate safeguards to protect people's privacy. In this connection we would test two hypotheses.

H1. Privacy protection comes with a cost to consumers.

H2. Privacy protection is not the top concern of businesses.

Considering H1, when an organization places a higher importance on privacy protection, it takes precautions when such data is in transit or storage. It is much easier to control the storage aspects because of its single location in the cloud. However, cloud services distribute their backup storages in multiple locations for business continuity and disaster recovery purposes. Thus, monitoring all such locations where such protected data is stored is more expensive for the businesses and so it is not pursued. Another aspect to keep in mind is that when a record is deleted by the customer it gets deleted from view but does not go out of storage. This aspect has affected some businesses when people recover deleted information. Microsoft White Paper on this topic shows that the privacy protection in the cloud comes with a cost [9]. Thus, H1 is valid. People note the daily barrage of data breaches involving millions of records being stolen from large businesses, some with very sensitive information such as people's Social Security Numbers (SSN) and dates of birth (DoB). Even though credit card information is also stolen due to data breaches, the damage caused by such loss is repairable. However, when SSN and DoB are stolen there is no possibility of replacing such information.

The common response to data breaches resulting in credit monitoring for a set period does not really protect the privacy of individuals when SSN and DoB are lost. This could result in identity theft, causing greater harm to the privacy of individuals. Since the market place is showing that these types of violations are occurring on a regular basis, it is easy to note that H2 is valid. This claim is validated by the American Bar Association's book on Health Care Data Breaches [12].

Protecting people's privacy when using cloud services comes with applying technical tools and changing business processes. Currently too much emphasis is placed on automation in business

Copyright © 2016, Infonomics Society

processes. With automation, detecting and stopping unauthorized access to stored data is difficult. Hackers take advantage of this weakness in business processes to steal sensitive data. Moreover, without necessary changes to business processes more violations would occur resulting in not only loss of privacy but also in depletion of financial resources. This is more evident when a hacker successfully initiates a financial transfer of large sums of money without anyone monitoring the transaction because of automation. When these frauds are realized, it exposes the vulnerability in internal controls which are then exploited by others for compromising people's privacy. Hence, technical tools alone should not be depended on for protecting privacy but more human intervention and monitoring are needed to prevent loss of privacy.

Cloud service is a major resource that saves businesses cost and provides many advanced services. It enables the business to focus on what they are good at and lets others deal with the chore of managing a computing system. Also, cloud service removes many barriers to entry. However, Cloud Service Providers (CSPs) do not have the necessary information to know the levels of sensitivity for stored information. It is up to each business customer to apply adequate safeguards to stored data in the cloud. Thus, expecting the CSP to provide protection to stored data in the cloud is not tenable. When a user presents valid credentials for access to stored data, there is nothing that a CSP can enforce that would detect unauthorized access. Consequently, the businesses should add additional authentication mechanisms when sensitive data is accessed so that only the valid user will be able to provide such information. An analysis of many breaches shows that the hacker gained initial access through stolen credentials. Since protecting credentials is not easy for SMBs, all large businesses that use business partners lacking such security protection should modify their business processes to look for additional validation when a user presents themselves to a cloud system for access to sensitive data. Thus, the focus should be on business processes when protecting privacy of data in the cloud.

5. Summary

Cloud services have many benefits to offer consumers. It enables businesses to have a cost effective alternative to get high-end computing services. It also puts the burden on businesses using cloud services to know the potential vulnerabilities in storing data in the clear in cloud. Many businesses realize the need for protecting such data but due to the need for rapid access to stored data, compromise on using adequate tools to protect data. This results in privacy violations. Certain privacy violations can be remedied through the use of steps like credit monitoring and replacement of information.

However, when Personally Identifiable Information or Protected Health Information in the cloud is stolen, it does irreparable damage to privacy. Business processes are important in protecting privacy of information stored in the cloud. It is pointed out in the paper that there are technical tools available at a cost but businesses may not be deploying them. Also, businesses depend too much on automated processing of data for rapid response. This aspect should be looked into carefully for protection of people's privacy through the use of modifications to existing business processes.

6. References

[1] Anderson, N. 2009. Anonymized data really isn't - and here's why not, http://arstechnica.com/tech-policy/2009/09/your-secrets-live-online-in-databases-of-ruin/ (Accessed 8/30/16).

[2] Ariwa, E. and Ibe, K. C. 2013. Cloud Computing sustainability and business process reengineering in SMEs: Comparative analysis of UK and Nigeria, 2013 Third International Conference on Innovative Computing Technology.

[3] Barbaro, M. and Zeller, T. 2006. A face is exposed for AOL searcher no. 4417749, http://www.nytimes.com/2006/08/09/technology/09aol.html?_r=0 Accessed 8/30/16.

[4] Bellare, M., Ristenpart, T., Rogaway, P., Stegers, T. 2009. vol. 5867 in LNCS Series, Springer, NY, 295-312.

[5] Black, J. & Rogaway, P. 2002. Ciphers with Arbitrary Finite Domains, vol. 2271 in LNCS Series, Springer, NY, 114-130.

[6] FT. 2016. Ashley Madison Data Breach, https://www.ft.com/content/db7a5c42-c21a-11e6-9bca-2b93a6856354 Accessed 1/25/17.

[7] Hardesty, L. 2015. Privacy Challenges. http://news.mit.edu/2015/identify-from-credit-card-metadata-0129 (Accessed 8/30/16).

[8] Harmon, P. 2007. Business Process Change, 2nd Edition, Burlington, MA: Morgan-Kaufmann.

[9] Microsoft. n.d. Protecting Data and Privacy in the Cloud, http://download.microsoft.com Accessed 9/25/16.

[10] Srinivasan, S. 2014a. Cloud Computing Basics, New York, NY: Springer

Copyright © 2016, Infonomics Society

[11] Srinivasan, S. 2014b. Security, Trust and Regulatory Aspects of Cloud Computing in Business Environments, Hershey, PA: IGI Global.

[12] Thomson, L. L. 2013. Health Care Data Breaches and Information Security, American Bar Association, 253-267.

[13] Weldon, D. 2015. Ashley Madison breach shows that hackers may be getting personal, CIO.com, http://www.cio.com/article/2987830/online-security/ashley-madison-breach-shows-hackers-may-be-getting-personal.html. Accessed 9/2/2016.

Copyright © 2016, Infonomics Society