Data Privacy and Security in Cloud Computing

A Survey of Current Research Problems

alexandru totolici
Networks, Systems and Security Lab
Department of Computer Science, UBC
totolici@cs.ubc.ca

Categories and Subject Descriptors
A.1 [**General and Literature**]: Introductory and Survey; E.m [**Data**]: Miscellaneous; H.3 [**Information Systems**]: Information Storage and Retrieval

General Terms
Cloud Computing

Keywords
security, privacy, storage

1. INTRODUCTION

Cloud computing is an established approach to web service development and delivery, having solidified its position as a viable alternative to data centres on the one hand, and client-side software on the other. The term has been overloaded in time to mean a number of (different) things, but in principle it involves the placement of computing resources with another party (the cloud provider) in order to achieve cost, performance and reliability gains [11]. Such a model carries obvious appeal to many tiers of users and service providers, and has given rise to a number of various service offerings that address various specific needs. A brief terminology primer, which describes in further detail the various service models found in cloud computing, is provided to the reader in section 2.

With these new service models, however, new challenges have arisen with regards to data security, and extant concerns inherent in giving third-parties physical access to one's data have been amplified. Users of cloud services must rely on the effective security and moral integrity of the providers, neither of which can be guaranteed. As data are often some of the user's most important assets, the absence of such guarantees makes the overall risks difficult to quantify. However, it is still possible to survey and analyze the broader security aspects, which leads us to identify possible threats in section 3.

As cloud computing is not a new service model, surveys of security, privacy, and associated aspects have been previously compiled; we take a look at some of this related work in section 4.

While the existence of these surveys as well as shorter articles acknowledging ongoing security concerns with the cloud model have been published over the course of the last five years with some frequency, it is important to consider these issues as avenues for further research, rather than show-stoppers preventing the use of cloud computing entirely. We discuss some of this potential in section 5.

2. DEFINITIONS

This section provides a brief set of definitions of cloud computing terminology, but is not meant to be thorough or complete—for that, the reader is directed to the ontology provided by Youseff et al. [37].

To understand the risks and threats to data stored in the cloud, it is important to clarify who the users are, and what services they are provided with. In terms of cloud computing:

- **Users** (also **Clients** or **Customers**) are individuals, companies, or governments seeking the use of infrastructure and services in the cloud.
- **Service Providers** are individuals, companies, or governments with the ability to offer infrastructure and services for general consumption.
- **Services** can be:
  - **Infrastructure-as-a-Service** (`IaaS`) is facilitated by computer virtualization. In `IaaS`, users are presented with one or more virtual machines to which they have full access, and can use as if these machines were stored in a data centre. `IaaS` also enables Virtual Private Clouds [13, 30, 36], where users can create overlay networks based on virtualized hardware.
  - **Platform-as-a-Service** (`PaaS`) is accomplished by the provision of software platforms [7, 35] on which users can write applications. Access to the underlying hardware—visualized or not—is not made available, and specific APIs are usually required to interact with storage and network, etc.

– **Software-as-a-Service** (SaaS) allows companies to offer software via the Internet, as opposed to having clients download and install the applications [3, 9, 26].

– **X-as-a-Service** (XaaS) encompasses everything else; specific offerings, such as **Data-as-a-Service** (DaaS [17, 20]) or **Communications-as-a-Service** (CaaS [28]) are made by various providers in order to answer targeted needs.

As one can observe from the classification above, symmetry is immediately apparent in terms of users and service providers, where the two are often interchangeable. For example, SaaS applications can be built on top of PaaS offerings; these, in turn, are customers of (one or more) IaaS providers [12]. Such nesting creates interesting economic dynamics, but these are not without additional and novel security challenges of their own.

Lastly, it is useful to think about the nature of the data stored in the cloud, in order to better understand the risks involved in case of improper access:

- **Raw** data are kept in DaaS and storage containers [17, 20, 25], and are usually accessed directly by the user.
- **Implicitly-stored** data are part of an offering (e.g., IaaS, SaaS, etc.) and is usually addressable through the respective service to which it belongs. Data in this form are usually not directly useful outside of the service itself, as they may be encrypted, lacking structural meaning etc.
- **Cache/transient** data are stored in *Content Delivery Networks* (CDN [21, 27]) as part of an effort to ensure fast delivery.

We focus our survey on privacy and security risks to **data**, rather than issues related to end-user tracking, data mining, aggregation, or disclosure; these have been discussed elsewhere [8, 14] and are not analyzed herein. While many of these new threats to user privacy can be attributed to the advent of cloud computing—many are dictated by emerging economic models that surround SaaS—they are in an orthogonal class of problems outside of the current scope.

3. THREATS

There are a number of different types of threats to the security and privacy of data stored in the cloud, ranging from other users of a shared service and extending all the way to state governments [29]. Some of these threats are extensions of extant concerns, carried over to the new paradigm, while others are entirely new challenges provided by this model.

3.1 Threat Classification

Threats can be classified based on both what they target, and who the responsible agents are. Various incentives for compromise exist, and it is important to understand the ways in which these differ in the cloud computing context, when compared to previous service models.

3.1.1 Targets

Broadly-speaking, any of the parties involved in cloud computing is at risk in terms of security breaches and privacy violations.

- **Customers**

  Clients of cloud services, whether private (home users) or public (companies and institutions), own the bulk of the data available in these services. For certain users, this information may be very sensitive, such as patient information or financial data stored with a third-party service provider like Mint [16]. Access to extensive personally-identifying information often may lead to identity theft. As the United States' Federal Trade Commission reports [5], credit card fraud is the most common form of ID theft, with serious impact to victims [33]. At minimum, it may be assumed that the user's privacy would be violated if a leak were to occur.

  For public clients, such as businesses, improper access to data may lead to significant business losses; source code and other proprietary information may be disclosed to competitors, to the company's detriment; if user data is leaked, the company may suffer a loss of reputation and even face financial penalties. In the case of a governmental agency, a leak could be tremendous, even lead to the endangerment of citizens or the state.

- **Service Provider(s)**

  The service provider itself is a very attractive target for attackers, as control of the infrastructure would allow for virtually unlimited access to the services running on it.

3.1.2 Agents

Most of the incentives for attackers remain unchanged from pre-cloud computing days, and revolve largely around financial profit.

- **Private Actors**

  Other users of the services may attempt to access data to which they are not privileged, such as documents on a spouse or friend, or a competing business, where the target may be software or customer information *en masse*.

- **Service Provider(s)**

  Intentionally or not, the provider(s) can access users' private data, causing leaks or possible corruption. Considering that some users trust multiple providers[1], the risks are quickly compounded, and often difficult to ascertain.

- **Governments**

[1]Consider a SaaS provider employing IaaS, DaaS, and CDNas part of their overall infrastructure.

Jurisdiction over the service provider(s) allows governments to request the data associated with certain users [34], even in instances where the government does not have jurisdiction over the individual(s) in question. This attack existed in the data centre model, where one company's server racks could be raided in the presence of appropriate warrants. But considering that many more private users are using online services in place of previously client-side software (email [10], office/productivity [9]), this out-of-jurisdiction access could be abused for political and intelligence gains.

3.2 Vectors

The vectors that can be employed by would-be attackers range from technological to legal, reflecting the wide spectrum of cloud computing users.

- **Insider Threats**

  Insider threats are very difficult to guard against. It is up to the service provider to ensure that adequate mechanisms are in place, in order to prevent rogue employees from tampering or exfiltrating customer data. This threat is not particular to cloud environments, however, but clients of such services must be willing to trust the provider's mechanisms to be sufficient.

- **Virtual Machine Layer**

  Particularly in `IaaS` environments, compromising the virtual machine can lead to an attacker breaking out of the isolation provided by the hypervisor, resulting in access to, and possibly control of, the other tenants in the shared environment.

- **Application Layer**

  Bugs in software have been present for as long as software itself, and cloud applications are not much different. Access to user data can be the direct result of such bugs, when authentication and authorization controls do not function correctly [6].

  To some extent, phishing attacks are made more successful because so much of a (private) user's data are stored online, from emails to banking and social networks. While certainly cloud computing is not the reason for the existence of phishing, the reliance on browsers for more and more of the activities users take part in on their computers has increased the prevalence and success of this attack vector.

- **Policy**

  Some companies begin to migrate only part of their data into the cloud, in order to take advantage of cost and reliability, but plan to isolate some of their data form the cloud. However, failures in policy specification, propagated by insufficient technological mechanisms, can lead to accidental transfer of highly-confidential information into the cloud.

- **Caching**

  Content Distribution Networks act as distributed caches, and satisfy some of the reliability and performance requirements for content-intensive applications. However, since most commercial CDN offerings, such as Akamai [19], lack access control, knowledge of a URL is usually sufficient to gain access to data.

- **Legal**

  The presence of data on servers within a particular jurisdiction subjects it to that country's laws. This is made illegal in parts of the world, most notably the European Union [18], and cloud providers must ensure that geographical boundaries for data are respected.

4. RELATED WORK

A number of surveys have been written on cloud computing security and privacy, from both technological and legal points of view. Some of these surveys are broader in scope, attempting to analyze the general security of cloud computing infrastructure at each level of the technological stack and for every type of service offering [31, 24]. Security concerns of this nature are still one of the principal detractors for many companies when it comes to choosing to move part of their business to the cloud [2].

A list of the most relevant legal acts with regard to data in the cloud is compiled in [38] and [29]. The latter also provides some insight regarding the extent to which governments can go in order to obtain access to the data they are interested in. The Canadian Privacy Commissioner has considered some of the user-centric issues around cloud computing [**?**]. The legal community has also been interested in the cloud model and the implications it has in terms of data privacy and protection, as illustrated in [22, 32, 14]

A good summary table of the kinds of risks—policy, technological, and legal—is provided in [15].

5. CONCLUSION

While it may appear that numerous security concerns plague the storage of sensitive data in the cloud, it is important to recognize that such issues are to be expected in a relatively young technology such as this. Operating systems, for example, have been under active research for over four decades, and still there are security issues left uncovered.

Each of the attack vectors presented in subsection 3.2 can be addressed to some extent already. Virtual machine technology aims to further increase the separation of the control and 'worker' VMs [4]. By encrypting and/or storing data via erasure coding and redundancy [1] it is possible to reduce the effects of some breaches. The legal challenges can also be addressed by erasure coding across geographical regions, or by specifying mechanisms that enforce geographic containment; if that is not possible, smaller, geospecific cloud providers are likely to enter the market, in order to satisfy demand for localized services.

Many of the issues that surround data security and privacy can also be addressed by those building services in the cloud, by considering their choices in terms of storage formats, and choosing sensible and secure approaches where possible [23].

6. REFERENCES

[1] H. Abu-Libdeh, L. Princehouse, and H. Weatherspoon. *RACS*. ACM Press, New York, New York, USA, June 2010.

[2] Avanade. Global Survey: Has Cloud Computing Matured? (June), 2011.

[3] Basecamp. Retrieved February 9, 2011, from http://basecamphq.com/.

[4] P. Colp, M. Nanavati, J. Zhu, and W. Aiello. Breaking up is hard to do: security and functionality in a commodity hypervisor. *Proceedings of the*, pages 189–202, 2011.

[5] U. F. T. Commission. Consumer Sentinel Network Data Book for January-December 2010. (December), 2011.

[6] Dropbox Security Blog. Retrieved April 10, 2011, from http://blog.dropbox.com/?p=821.

[7] Google App Engine. Retrieved February 9, 2011, from https://code.google.com/appengine/.

[8] R. Gellman. Privacy in the clouds: Risks to privacy and confidentiality from cloud computing. *World privacy forum*, 2009.

[9] Google Docs. Retrieved February 9, 2011, from https://docs.google.com/.

[10] Google Mail. Retrieved April 10, 2011, from https://gmail.com/.

[11] R. Harms and M. Yamartino. The Economics of the Cloud. *Technology*, (November), 2010.

[12] Adam Wiggins on Building Heroku on Top of Amazon EC2. Retrieved April 10, 2011, from http://www.infoq.com/interviews/wiggins-heroku-ec2-cloud.

[13] F. Krautheim. Private Virtual Infrastructure for Cloud Computing. In *Proceedings of the 2009 conference on Hot topics in cloud computing*, page 5. USENIX Association, 2009.

[14] P. Lanois. Caught in the Clouds: The Web 2.0, Cloud Computing, and Privacy? *Nw. J. Tech. & Intell. Prop.*, 9(2):4, 2010.

[15] C. L. Liu, W. H. Chen, and D. K. Tung. Identification of Critical Security Issues for Cloud Computing. *Applied Mechanics and Materials*, 145:272–276, Dec. 2011.

[16] Mint — Personal Finance. Retrieved April 10, 2011, from http://mint.com.

[17] MongoHQ. Retrieved February 7, 2011, from https://www.mongohq.com/.

[18] J. Noltes. *Data location compliance in cloud computing*. PhD thesis, Aug. 2011.

[19] E. Nygren, R. Sitaraman, and J. Sun. The Akamai Network: A platform for high-performance Internet applications. *ACM SIGOPS Operating Systems Review*, 44(3):2–19, 2010.

[20] Parse. Retrieved February 7, 2011, from https://parse.com/.

[21] G. Peng. CDN: Content distribution network. *ArXiv*, pages 1–26, 2008.

[22] R. Picker. Competition and Privacy in Web 2.0 and the Cloud. *SSRN eLibrary*, 414(June), 2008.

[23] R. Popa, C. Redfield, and N. Zeldovich. CryptDB: protecting confidentiality with encrypted query processing. *Proceedings of the*, pages 85–100, 2011.

[24] K. Popovic. Cloud computing security issues and challenges. *MIPRO, 2010 Proceedings of the*, pages 344–349, 2010.

[25] Amazon Simple Storage Service. Retrieved February 7, 2011, from http://aws.amazon.com/s3/.

[26] Salesforce. Retrieved February 9, 2011, from http://www.salesforce.com/.

[27] S. Saroiu, K. Gummadi, R. Dunn, S. Gribble, and H. Levy. An analysis of internet content delivery systems. *ACM SIGOPS Operating Systems Review*, 36(SI):315–327, 2002.

[28] Amazon Simple Notification Service. Retrieved February 9, 2011, from http://aws.amazon.com/sns/.

[29] C. Soghoian. Caught in the cloud: Privacy, encryption, and government back doors in the web 2.0 era. *Journal on Telecommunications and High Technology Law*, pages 359–424, 2010.

[30] B. Sotomayor, R. Montero, I. Llorente, and I. Foster. Virtual infrastructure management in private and hybrid clouds. *Internet Computing, IEEE*, 13(5):14–22, Sept. 2009.

[31] S. Subashini and V. Kavitha. A survey on security issues in service delivery models of cloud computing. *Journal of Network and Computer Applications*, 34(1):1–11, Jan. 2011.

[32] D. Svantesson and R. Clarke. Privacy and consumer risks in cloud computing. *Computer Law Security Review*, 26(4):391–397, 2010.

[33] R. Turnock. Special Report. *Media History*, 16(1):125–134, Feb. 2010.

[34] Twitter Ordered to Yield Data in WikiLeaks Case. Retrieved April 10, 2011, from http://www.nytimes.com/2011/11/11/technology/twitter-ordered-to-yield-data-in-wikileaks-case.html?_r=1.

[35] Windows Azure. Retrieved February 9, 2011, from http://www.microsoft.com/windowsazure/.

[36] T. Wood, A. Gerber, K. Ramakrishnan, P. Shenoy, and J. Van der Merwe. The case for enterprise-ready virtual private clouds. In *Proceedings of the 2009 conference on Hot topics in cloud computing*, page 4. USENIX Association, 2009.

[37] L. Youseff, M. Butrico, and D. Da Silva. Toward a Unified Ontology of Cloud Computing. In *2008 Grid Computing Environments Workshop*, pages 1–10. IEEE, Nov. 2008.

[38] M. Zhou, R. Zhang, W. Xie, W. Qian, and A. Zhou. Security and Privacy in Cloud Computing: A Survey. *2010 Sixth International Conference on Semantics, Knowledge and Grids*, pages 105–112, Nov. 2010.