Cloud computing: Challenges and future directions

Kim-Kwang Raymond Choo

A computer's operating system, applications and data are typically installed and stored in the 'traditional' computing environment. In a cloud computing environment, individuals and businesses work with applications and data stored and/or maintained on shared machines in a web-based environment rather than physically located in the home of a user or a corporate environment. Lew Tucker, Vice President and Chief Technology Officer of Cloud Computing at Sun Microsystems, explained that cloud computing is 'the movement of application services onto the Internet and the increased use of the Internet to access a wide variety of services traditionally originating from within a company's data center' (Creeger 2009: 52). For example, web-based applications such as Google’s Gmail™ can be accessed in real time from an Internet-connected machine anywhere in the world.

Use of cloud services creates a growing interdependence among both public and private sector entities and the individuals served by these entities. This paper provides a snapshot of risk areas specific to cloud services and those that apply more generally in an online environment which clients of cloud service providers should be aware of.

Cloud computing

It is not clear when the term cloud computing was first coined. For example, Bartholomew (2009), Bogatin (2006) and several others suggested that ‘cloud computing’ terminology was, perhaps, first coined by Google™ Chief Executive Eric Schmidt in 2006. Kaufman (2009: 61) suggests that cloud computing terminology ‘originates from the telecommunications world of the 1990s, when providers began using virtual private network (VPN) services for data communication’. Desisto, Plummer and Smith (2008: 1) state that ‘[t]he first SaaS [Software as a Service] offerings were delivered in the late 1990s...[a]lthough these offerings weren’t called cloud computing’. There is, however, agreement on the definition of cloud computing.

The National Institute of Standards and Technology defines cloud computing as a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (eg networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction (Mell 2009: 9).
Architectures and deployment models
Cloud architectures can be broadly categorised into:

Infrastructure as a Service (IaaS) is the foundation of cloud services. It provides clients with access to server hardware, storage, bandwidth and other fundamental computing resources. For example, Amazon EC2 allows individuals and businesses to rent machines preconfigured with selected operating systems on which to run their own applications.

Platform as a Service (PaaS) builds upon IaaS and provides clients with access to the basic operating software and optional services to develop and use software applications (eg database access and payment service) without the need to buy and manage the underlying computing infrastructure. For example, Google App Engine allows clients to run their web applications (ie software that can be accessed using a web browser such as Internet Explorer over the internet) on Google’s infrastructure.

Software as a Service (SaaS), builds upon the underlying IaaS and PaaS provides clients with integrated access to software applications. For example, Oracle SaaS Platform allows independent software vendors to build, deploy and manage SaaS and cloud-based applications using a licensing economic model. Here, users purchase a license and support for components of the Oracle SaaS Platform on a monthly basis.

Cloud services can be used in a private, public, community/managed or hybrid setting (Cloud Security Alliance 2009). Privately-hosted cloud services are generally considered a safer but more costly option than services using a shared-tenancy setting (ie data from different clients stored on a single physical machine). In line with this, the US Government recently announced an initiative ‘to offer cloud-based services that are hosted in private data centers and which could be used to handle more sensitive data’ (McMillan 2009: np).

In a community/managed setting, tenancy can either be single (dedicated) or shared and the IT infrastructure is either managed by the organisation or a third-party cloud service provider. The main difference between hybrid cloud services and other cloud services is that the former ‘is a composition of two or more clouds (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability’ (Mell & Grance 2009: 13).

Benefits
Cloud computing provides a scalable online environment which facilitates the ability to handle an increased volume of work without impacting on the performance of the system. Cloud computing also offers significant computing capability and economy of scale that might not otherwise be affordable to businesses, especially small and medium enterprises (SMEs) that may not have the financial and human resources to invest in IT infrastructure. Advantages include:

- Capital costs—SMEs can provide unique services using large-scale resources from cloud service providers and ‘add or remove capacity from their IT infrastructure to meet peak or fluctuating service demands while paying only for the actual capacity used’ (Sotomayor et al. 2009: 14) on a ‘pay-as-you-go’ economic model.
- Running costs—it can also be significantly cheaper to rent added server space for a few hours at a time rather than maintain proprietary servers. Rental prices for Amazon Elastic Compute Cloud (EC2), for example, are between US$0.10–1.00 an hour. Businesses do not have to worry about upgrading their resources whenever a new version of the application is available. Businesses can also base their services in the data centres of bigger enterprises or host their IT infrastructure in locations offering the lowest cost.

Advantages of using cloud services can also go beyond cost savings as cloud computing allows clients to:
- avoid the expense and time-consuming task of installing and maintaining hardware infrastructure and software applications; and
- allow for the rapid provisioning and use of services to clients by optimising their IT infrastructure (Lewin 2009).

External hosting of applications and storage also ensures redundancy and business continuity in the event of a site failure.

Service level agreements
To ensure guarantees from cloud service providers for service delivery, businesses using cloud computing services typically enter into service level agreements (SLAs) with the cloud service providers. Although SLAs vary between businesses and cloud service providers, they typically include the required/agreed service level through quality of service parameters, the level of service availability, the indication of the security measures adopted by the cloud service provider and the rates of the services.

Cloud computing risks

Attacks targeting shared-tenancy environment
A virtual machine (VM) is the software implementation of a computer that runs its own operating system and application as if it was a physical machine (VMware 2009). Multiple VMs can concurrently run different software applications on different operating system environments on a single physical machine. This reduces hardware costs and space requirements.

In a shared-tenancy cloud computing environment, data from different clients can be hosted on separate VMs but reside on a single physical machine. This provides maximum flexibility. Software applications running in one VM should not be able to impact or influence software running in another VM. An individual VM should be unaware of the other VMs running in the environment as all actions are confined to its own address space.

In a recent study, a team of computer scientists from the University of California, San Diego and Massachusetts Institute of Technology examined the widely-used Amazon EC2 services. They found that ‘it is possible to map the internal cloud infrastructure, identify where a particular target VM is likely to reside, and then instantiate new VMs until one is placed co-resident with the target’ (Ristenpart et al. 2009: 199). This demonstrated that the research team were able to load their eavesdropping software onto the same servers hosting targeted websites (Hardisty 2009). By identifying the target VMs, attackers can potentially monitor the cache (a small allotment of high-speed memory used to store frequently-used information)
in order to steal data hosted on the same physical machine (Hardesty 2009). Such an attack is also known as a *side-channel attack*.

The findings from this research may only be a proof-of-concept at this stage, but it raises concerns about the possibility of cloud computing servers being a central point of vulnerability that can be criminally exploited. The Cloud Security Alliance, for example, listed this as one of the top threats to cloud computing.

Attacks have surfaced in recent years that target the shared technology inside Cloud Computing environments. Disk partitions, CPU caches, GPUs, and other shared elements were never designed for strong compartmentalization. As a result, attackers focus on how to impact the operations of other cloud customers, and how to gain unauthorized access to data. (Cloud Security Alliance 2010: 11)

### VM-based malware

Vulnerabilities in VMs can be exploited by malicious code (malware) such as VM-based rootkits designed to infect both client and server machines in cloud services. Rootkits are cloaking technologies usually employed by other malware programs to abuse compromised systems by hiding files, registry keys and other operating system objects from diagnostic, antivirus and security programs. For example, in April 2009, a security researcher pointed out how a critical vulnerability in VMware’s VM display function could be exploited to run malware, which allows an attacker ‘to read and write memory on the “host” operating system [OS]’ (Keizer 2009: np).

VM-based rootkits, as pointed out by Price (2008: 27), could be used by attackers to ‘gain complete control of the underlying OS without the compromised OS being aware of their existence…[and] are especially dangerous because they also control all hardware interfaces. Once the VM-based rootkits are installed on the machine, they can “view keystrokes, network packets, disk state, and memory state, while the compromised OS remains oblivious”.

### Botnet hosting

Bot malware typically takes advantage of system vulnerabilities and software bugs or hacker-installed backdoors that allow malicious code to be installed on machines without the owners’ consent or knowledge. They then load themselves into computers often for nefarious purposes. Machines infected with bot malware are then turned into “zombies” and can be used as remote attack tools or to form part of a botnet under the control of the botnet controller. Zombies are compromised machines waiting to be activated by their command and control (C&C) servers. The C&C servers are often machines that have been compromised and arranged in a distributed structure to limit traceability.

Cybercriminals could potentially abuse cloud services to operate C&C servers to carry out distributed denial-of-service (DDoS) attacks, which are attacks from multiple sources targeting specific websites by flooding a web server with repeated messages, tying up the system and denying access to legitimate users, as well as other cyber criminal activities. In December 2009, for example, a ‘new wave of a Zeus bot (Zbot) variant was spotted taking advantage of Amazon EC2’s cloud-based services for its C&C…functionality’ (Ferrer 2009: np).

#### Launch pad for brute force and other attacks

There have also been suggestions that the virtualised infrastructure can be used as a launching pad for new attacks. A security consultant recently suggested that it may be possible to abuse cloud computing services to launch a brute force attack (a strategy used to break encrypted data by trying all possible decryption key or password combinations) on various types of passwords. Using Amazon EC2 as an example, the consultant estimated that based on the ‘hourly fees Amazon charges for its EC2 web service, it would cost more than [US$]1.5m to brute force a 12-character password containing nothing more than lower-case letters a through z…[but] an 11-character code costs less than [US$]60,000 to crack, and a 10-letter phrase costs less than [US$]2,300’ (Goodin 2009: np).

Although it is still relatively expensive to perform brute force online password-guessing attacks (also known as *online dictionary attacks*), this could have broad implications for systems using password-based authentication. It may not take long for attackers to design a more practical and cheaper mechanism that exploits cloud services as a launch pad for other attacks, a threat also identified by the Cloud Security Alliance (2010: 8):

> Future areas of concern include password and key cracking, DDoS, launching dynamic attack points, hosting malicious data, botnet command and control, building rainbow tables, and CAPTCHA solving farms.

### Data availability (business continuity)

A major risk to business continuity in the cloud computing environment is loss of internet connectivity (that could occur in a range of circumstances such as natural disasters) as businesses are dependent on the internet access to their corporate information. In addition, if vulnerability is identified in a particular service provide by the cloud service provider, the business may have to terminate all access to the cloud service provider until they could be assured that the vulnerability has been rectified.

There are also concerns that the seizure of a data-hosting server by law enforcement agencies may result in the unnecessary interruption or cessation of unrelated services whose data is stored on the same physical machine.

In a recent example, ‘FBI agents [reportedly] seized computers from a data center at 2323 Bryan Street in Dallas, Texas, attempting to gather evidence in an ongoing investigation of two men and their various companies accused of defrauding AT&T and Verizon for more than US$6 million’ (Lemos 2009: np). This resulted in the unintended consequence of disrupting the continuity of businesses whose data and information are hosted on the seized hardware.

[For] LiquidMotors, a company that provides inventory management to car dealers, the servers held its client data and hosted its managed inventory services. The FBI seizure of the servers in the data center rack effectively shut down the company, which filed a lawsuit against the FBI the same day to get the data back (Lemos 2009: np)

While the above example may be an isolated case, it raised concerns about unauthorised access to seized data not related to the warrant, which can result in the unintended disclosure of data to unwanted parties, particularly in authoritarian countries.
There had been a number of reported incidents of cloud services being taken offline due to DDoS attacks (see Metz 2009). Although DDoS attacks already existed, the cloud computing environment is a new attack sector that may have a more widespread impact on internet users.

The security measures adopted by different cloud service providers varies. If a cybercriminal can identify the (cloud service) provider whose vulnerabilities are the easiest to exploit, then this entity becomes a highly visible target. The lack of security associated with this single entity threatens the entire cloud in which it resides (Kaufman 2009: 63).

Rogue clouds
Just like entrepreneurs, cybercriminals and organised crime groups are always on the lookout for new markets and with the rise of cloud computing, a new sector for exploitation now exists. Rogue cloud service providers based in jurisdictions with lax cybercrime legislation can provide confidential hosting and data storage services for a usually steep fee. Such services could potentially be abused by organised crime groups to store and distribute criminal data (eg child abuse materials for commercial purposes) to avoid the scrutiny of law enforcement agencies.

Hosting confidential business data with cloud service providers involves the transfer of a considerable amount of management control to cloud service providers that usually results in diminished control over security arrangements. There is the risk of rogue providers mining the data for secondary uses such as marketing and reselling the mined data to other businesses. A June 2009 email survey of 220 decision-makers in US organisations with more than 1,000 employees highlighted similar concerns. In the survey, 40.5 percent of the respondents agreed/strongly agreed that '[t]he trend toward using SaaS and cloud computing solutions in the enterprise seriously increases the risk of data leakage' (Proofpoint 2009: 24).

Unfortunately, clients (especially SMEs) are often less aware of the risks and may not have an easy way of determining whether a particular cloud service provider is trustworthy. Tim Watson, head of the computer forensics and security group at De Montfort University remarked that ‘one provider may offer a wonderfully secure service and another may not, if the latter charges half the price, the majority of organisations will opt for it as they have no real way of telling the difference’ (Everett 2009: 7).

Other potential risks
Espionage risks
There is increasing pressure for nation-states to develop cyber-offensive capabilities. The next wave of cyber-security threats could potentially be targeted attacks aimed at specific government agencies and organisations, or individuals within enterprises including cloud service providers. For example, Google and several Gmail accounts belonging to Chinese and Tibetan activists have reportedly been targeted (Google 2010; Helft & Markoff 2010).

Foreign intelligence services and industrial spies may not disrupt the normal functioning of an information system as they are mainly interested in obtaining information relevant to vital national or corporate interests. They do so through clandestine entry into computer systems and networks as part of their information-gathering activities.

Cloud service providers may be compelled to scan or search data of interest to ‘national security’ and to report on, or monitor, particular types of transactional data as these data may be subject to the laws of the jurisdiction in which the physical machine is located (Gellman 2009). In addition, overseas cloud service providers may not be legally obliged to notify the clients (owners of the data) about such requests.

Regulation and governance
The privacy and confidentiality risks faced by businesses that use cloud services also depend to a large extent on the terms of service and privacy policy established by the cloud service providers. Failure to comply with data protection legislation may lead to administrative, civil and criminal sanctions. Data confidentiality and privacy ‘risks may be magnified when the cloud provider has reserved the right to change its terms and policies at will’ (Gellman 2009: 6).

Some cloud service providers argue that such jurisdictional issues may be capable of resolution contractually via SLAs and the like. Clients using cloud services could include clauses in their SLAs that indicate the law governing the SLA, the choice of the competent court in case of disputes arising from the interpretation and the execution of the contract. The Cloud Security Alliance (2009: 28) also suggested that clients of cloud services should require their providers ‘to deliver a comprehensive list of the regulations and statutes that govern the site and associated services and how compliance with these items is executed’.

Businesses should ensure that SLAs and other legally-binding contractual arrangements with cloud service providers comply with applicable regulatory obligations (eg privacy laws) and industry standards, as they may be liable for breaching these regulations even when the data being breached is held or processed by the cloud service provider.

Determining the law of the jurisdiction in which the SLA is held is an important issue. It may not, however, be as simple as examining the contractual laws that govern operations of cloud service providers to determine which jurisdiction’s laws apply in any particular case. Gellman (2009: 19) pointed out that ‘[t]he user may be unaware of the existence of a second-degree provider or the actual location of the user’s data…[and] it may be impossible for a casual user to know in advance or with certainty which jurisdiction’s law actually applies to information entrusted to a cloud provider’.

Businesses should continue to conduct due diligence on cloud service providers, have a comprehensive compliance framework and ensure that protocols are in place to continuously monitor and manage cloud service providers, offshore vendors and their associated outsourcing relationships. This would ensure businesses have a detailed understanding of the data storage information to maintain some degree of oversight and ensure that an acceptable authentication and access mechanism in place to meet their privacy and confidentiality needs.

The way forward
Culture of security
Vulnerabilities in a particular cloud service or cloud computing environment can potentially be exploited by criminals and actors with malicious intent. However, no single public or private sector entity ‘owns’ the issue of cyber security. There is,
arguably, a need to take a broader view and promote transparency and confidence building between cloud service providers, businesses and government agencies using cloud services as well as between government and law enforcement agencies. In addition, an effective cyber-security policy should be comprehensive and encompass all (public and private sector) entities. The public and private sectors should continue to work together to:

- identify and prioritise current and emerging risk areas;
- develop and validate effective measures and mitigation controls. This would involve establishing a standard that mandates certain minimum requirements to ensure an adequate level of electronic information exchange security; and
- ensure that these strategies are implemented and updated at the respective level.

It is reasonable to assume that higher levels of security can only be achieved at higher marginal costs. To encourage a culture of security, governments could incubate and create market incentives for cloud service providers to integrate security into the software and hardware and system development life cycle. An improved level and type of security is likely to increase the marginal cost of security violations, which in turn will reduce the marginal benefits of cybercrime.

An example is to create an environment conducive for cloud service providers to achieve marketing and competitive advantages if they offer products and services with higher levels and more innovative types of security to assist in combating cyber exploitation. This could be accomplished through government tenders. Dealing with insider threats should also be incorporated into the software/hardware and system development life cycle.

Need for reforms to Australia’s privacy regime

It is a near impossible task to fully harmonise privacy and data protection regimes due to the different judicial and legal systems internationally. There are countries that do not have any mandatory data retention or data protection requirements.

To lower operating costs and ensure redundancy, cloud service providers may have data centres in a number of jurisdictions, with the data changing and moving continuously between the provider’s servers. In such cases, cloud service providers have to ensure that they comply with a myriad of regulatory obligations both locally and overseas.

The Eighth Principle of the United Kingdom’s Data Protection Act 1998 (c. 29), for example, provides that personal data shall not be transferred to a country or territory outside the European Economic Area unless that country or territory ensures an adequate level of protection for the rights and freedoms of data subjects in relation to the processing of personal data.

The European Economic Area (EEA) consists of the European Union (EU) member states together with Iceland, Liechtenstein and Norway. An agency or business in Australia must comply with the Australian Privacy Principles (APPs) when using a cloud service and in particular would need to comply with APP 8—cross-border disclosure of personal information. However, Australia is not an EEA country and has not been declared by the EU as having adequate privacy laws.

The Eighth Principle of the United Kingdom’s Data Protection Act 1998 (c. 29), for example, provides that personal data shall not be transferred to a country or territory outside the European Economic Area unless that country or territory ensures an adequate level of protection for the rights and freedoms of data subjects in relation to the processing of personal data. The European Economic Area (EEA) consists of the European Union (EU) member states together with Iceland, Liechtenstein and Norway. An agency or business in Australia must comply with the Australian Privacy Principles (APPs) when using a cloud service and in particular would need to comply with APP 8—cross-border disclosure of personal information. However, Australia is not an EEA country and has not been declared by the EU as having adequate privacy laws. Therefore, a cloud service provider with data centres in Australia and the United Kingdom that wishes to transfer personal data outside of the United Kingdom to Australia, but within their group of companies, would have to do so in a manner which ensures adequacy of security. The provider would have to submit to binding corporate rules for authorisation by the Information Commissioner. This unnecessarily increases the cost of compliance for cloud service providers and discourages them from establishing data centres in countries such as Australia.

The Australian Law Reform Commission, for example, explained that ‘the European Union specifically has cited this unusual [small business] exemption as a major obstacle to Australia being granted “adequacy” status under the European Union Directive on the Protection of Individuals with Regard to the Processing of Personal Data and on the Freedom of Movement of Such Data (the EU Directive)” (ALRC 2008: 113–114). Australia could consider removing the exemption for small business (including not-for-profit organisations) with an annual turnover of $3m or less from the Privacy Act 1988 (Cth). As recommended by the ALRC (2008: 114), this would bring Australian privacy laws into line with laws in similar jurisdictions, such as the United Kingdom (UK), Canada and New Zealand, and could facilitate trade by helping to ensure that Australia’s privacy laws are recognised as ‘adequate’ by the European Union.

In October 2009, the Australian Government released the first stage of reforms to enhance the protection of personal privacy, responding to the ALRC’s (2008) inquiry into the effectiveness of the Privacy Act 1988, which included proposed reforms to privacy that impinge on cross-border data flows (Australian Government 2009).

Research directions

Dr Renato Iannella, principal scientist of National ICT Australia, suggested that we need to say we need a cloud.au. It’s a cloud, but it’s only in Australia, therefore we can protect it and it’s not going to be in the hands of [foreigners] (Tung 2008: np).

Although having the data centres physically in Australia may reduce some of the location risks that cloud service clients may face, such an approach may not be feasible until Australia’s privacy laws are recognised as adequate by the EU.

More research should be funded to find ways to mitigate existing and new security risks in the cloud computing environment. Messmer (2009: np) suggested that in the near future, we might see ‘a shift to using or developing “security as a service” to adapt to new threat scenarios in both public cloud computing and virtualization of their IT infrastructure’.

A team of researchers from Xerox’s PARC and Fujitsu Laboratories of America also warned that these ‘new threats require new constructions to maintain and improve security’ and highlighted the need to design ‘tools to control and understand privacy leaks, perform authentication, and guarantee availability in the face of cloud denial-of-service attacks’ (Chow et al. 2009: 89). For example, encrypting stored data ensures data confidentiality. This, however, prevents cloud service providers from
executing services on this data—‘searching and indexing data is impossible to do with traditional, randomized encryption schemes’ (Chow et al. 2009: 89).

The Australian Government has invested significantly in law enforcement responses, education, science and research, and development. It is hoped that there will be further investment to enable Australian security researchers to play a more significant role in designing state of the art security tools that can be deployed in a cloud computing environment and help position Australia as an international leader in cyber security.

References

All URLs correct at July 2010


Desisto RP, Plummer DC & Smith DM 2008. Tutorial for understanding the relationship between cloud computing and SaaS. Stamford, CT: Gartner


www.aic.gov.au

Dr Kim-Kwang Raymond Choo is a Senior Research Analyst at the Australian Institute of Criminology. He wrote this paper while on a 2009 Fulbright Professional Australia-US Alliance Studies Scholarship in the United States.