

Data Risks in the Cloud

Roger Clarke

Xamax Consultancy Pty Ltd, Canberra, Australia
Visiting Professor in Computer Science, Australian National University, Canberra, Australia
Visiting Professor in Cyberspace Law & Policy, University of N.S.W., Sydney, Australia
Roger.Clarke@xamax.com.au

Received 6 April 2013; received in revised form 30 July 2013; accepted 22 August 2013

Abstract

Cloudsourcing involves considerably greater risks to data than do insourcing or conventional outsourcing. A generic data risk assessment identifies key concerns in relation to harm arising from threats impinging on vulnerabilities in the cloud. Guidance is provided as to appropriate safeguards to address those risks. Most services lack those safeguards, implying that individuals and user organisations need to be far more careful in their use of cloud services.

Keywords: Cloud computing, SaaS, Data security, Data security risk assessment, Vulnerabilities

1 Introduction

In conventional outsourcing, a supplier hosts equipment that supports a relevant stack of software, and stores and maintains data. During the last decade, a form of outsourcing has emerged that is commonly referred to as *cloud computing*. It has been applied to many forms of eCommerce, eBusiness and eGovernment. The literature offers a range of definitions of cloud computing [14], [48], with that proposed by NIST [33] perhaps being the most influential. The interpretation used in this paper is broadly consistent with mainstream definitions. However it follows [13] in locating cloud notions within the frame provided by outsourcing, and reflects the widespread dissatisfaction among IT professionals with the current state of play in the area.

In cloud computing, the supplier changes the focus of the offer from the equipment to the processes. Those processes may be run in any of a wide range of devices, and the location of those devices is primarily determined by the needs of the supplier, not those of the customer. The supplier scales the number of processes and the processing speed and storage capacity to meet the customer's varying needs over hourly, daily, monthly and annual cycles, to reflect growth and decay factors, and in the face of demand uncertainty. The supplier may offer a tariff based on usage, because instead of unused capacity being locked up in hosts that have been pre-allocated to a customer, the supplier can make more efficient use of the available computing resources.

Three categories of cloud computing service model are conventionally distinguished [33]. Infrastructure as a Service (IaaS) refers to the provision of a bare (but virtualised) machine, without software or with little more than a specific operating system and version. Amazon's EC2 and Rackspace were early movers in a marketplace that is becoming densely populated. Platform as a Service (PaaS), on the other hand, offers a configured platform on which organisations can develop and/or install their applications. Examples include Microsoft Windows Azure, Google Apps and a range of services offering specific application development and/or execution environments. The third category, Software as a Service (SaaS), makes specific application software available. SaaS is offered as an alternative to applications running on the organisation's or individual's own host devices or workstations. Examples targeted at organisations include Salesforce, Google Gmail, Zoho, Google Apps, MS Office 365, Dropbox and MYOB LiveAccounts. Examples of SaaS offered directly to consumers, include Zoho, Gmail, Google Docs and Dropbox.

SaaS commonly involves complete dependence on an autonomous external service-provider. At the other extreme, cloudsourcing may be subject to reasonably tight control by the user-organisation, e.g. if it is used solely as a means of replicating data, in such forms as a backup service or a multi-server environment to achieve regional distribution. Between these two extremes are various arrangements in which the dependence on the service-provider may be mitigated by maintaining up-to-date local copies of data, and even by the retention of some degree of local processing capability, e.g. as a fallback arrangement when the service is inaccessible.

A great many risks arise in relation to all forms of insourcing, outsourcing and cloudsourcing. Some risks relate to IT infrastructure, some to the services that the technology enables, and some to the data that it maintains. This article focusses on data risks. From this perspective, cloudsourcing encompasses various configurations. In particular:

- Data may be held entirely in the cloud, with any copies on the user's own device(s) being temporary, partial and non-authoritative;
- The authoritative copy of the data may be held in the cloud, but with one or more secondary copies on the user's own device(s); and
- One or more secondary copies of the data may be held in the cloud – in particular as backup, or to facilitate the synchronisation of multiple copies on the user's own devices – but with the authoritative copy of the data held on one or more of the user's own-devices.

An organisation may use different configurations to support different business functions. It may even switch between different configurations, e.g. to cope with peaks in demand for access or processing. The purpose of this article is to investigate the data risks that arise in all of the variants of cloudsourcing identified above.

2 Research Method

Cloudsourcing has been the subject of active marketing during the period since 2006. Marketers have a natural tendency to overplay the benefits and underestimate the disbenefits and risks. Many exaggerations and misrepresentations have been swallowed by uncritical reporters in the trade press. Moreover, the enthusiasm has at times spilt over into universities and even the academic literature [3].

During the course of a 4-year research program into cloudsourcing, the author has developed templates to support organisations in identifying benefits, disbenefits and risks [10]. Risks faced by consumers were examined in [43], and the templates were adapted and applied to the consumer segment in [11]. The templates were further developed in

[13], which examined cloud computing from the perspective of outsourcing theory and security theory. The purpose of this paper is to examine the specific issue of data risks that arise from cloudsourcing, and the extent to which adopters and service-providers appear to be managing them.

The research reported on in this paper commenced with inspection of the relevant refereed literature on data risks, both generally and within the contexts of outsourcing and cloud computing, followed by summarisation of the known information. The emergence of formal literature lags well behind technological and market phenomena. Reasons for this include the slowness of the cycle of research, writing, review, revision and publication, together with the difficulties of studying small populations of diverse and unstable new phenomena.

To complement the slim body of relevant refereed work, this project accordingly included monitoring of and searches for media reports on the experiences of adopters of cloudsourcing, during the life of the movement to date, since 2005. Articles were sought that reported experiences of user organisations, as distinct from those based on promotional material issued by technology providers. Searches were undertaken using the search-engines of selected, reputable IT media outlets, and Google News. The primary search terms used were *cloud computing*, *cloud* and/or *SaaS*, *PaaS* or *IaaS*, in conjunction with *risk*, and with other terms that appeared in relevant articles. The results were coded, using conventional data security terminology as presented in the later parts of this paper.

It is of course highly desirable that stronger approaches be developed to the gathering of empirical data about the performance and malperformance of cloudsourcing service-providers. The industry is still immature, however, and the collection of reliable data is a challenging undertaking. It is untenable to delay investigations of this nature, because information is needed now, to enable informed decision-making. The author accordingly contends that the best available evidence needs to be used, and the conclusions from the research qualified in order to reflect the inadequacies in the available data.

The paper commences by reviewing the literature on data security, both that dealing with it in a generic sense, and in the particular context of cloudsourcing.

3 Data Security

As a framework for the analysis, a variant of the conventional security model was adopted. The model is closely related to, but differs in a couple of respects from, those in [21] and [15] pp. 38-39. Figure 1 provides a schematic representation. It is based on the propositions that:

- A Stakeholder's perception of the value of an Asset may be harmed by a Security Incident
- A Security Incident results from a Threatening Event impinging on a Vulnerability
- A Threatening Event is a specific instance of some generic category of Threat, including exogenous events (*Acts of God* such as lightning strikes), accidents caused by human agents, and intentional acts of human agents, commonly referred to as Attacks
- Safeguards (referred to in some parts of the literature as *controls*) are used to address the Threatening Events and Vulnerabilities, variously by prevention, deterrence, detection, mitigation and documentation
- Safeguards are subject to Countermeasures by Attackers

The well-established process of Security Risk Assessment was applied. This considers in turn the Assets, Harm, Threats, Vulnerabilities and existing Safeguards, in order to guide the development of a strategy and plan to assure protection that is reasonable in the circumstances [4], [31], [36], [40]. The purpose of a Risk Strategy is to ensure that a network of Safeguards is devised, implemented and maintained in order to appropriately manage at least those risks that are judged to be of the greatest concern. The concept of *risk* that is conventional within the professional security community is very specific, and somewhat counter-intuitive. Risk is a measure of the likelihood of Harm arising from a specific Threat. Risk, defined this way, is used as a guide in prioritising the Safeguards that an organisation's inherently limited resources should be invested in. An appropriate trade-off needs to be made among costs and benefits that can be estimated with some degree of confidence, and abstract Threats whose impact is uncertain.

The application of the Risk Assessment process in this paper differs from its common usage in that it is intentionally generic rather than focussed on the specific context of a particular organisation.

The primary focus of the paper is on additional risks that arise where cloudsourcing is adopted. When applying the generic analysis reported on in this paper to a particular context, it is also necessary to consider the extent to which risks that apply to insourcing, and to forms of outsourcing other than in the cloud, may be avoided or mitigated by cloudsourcing.

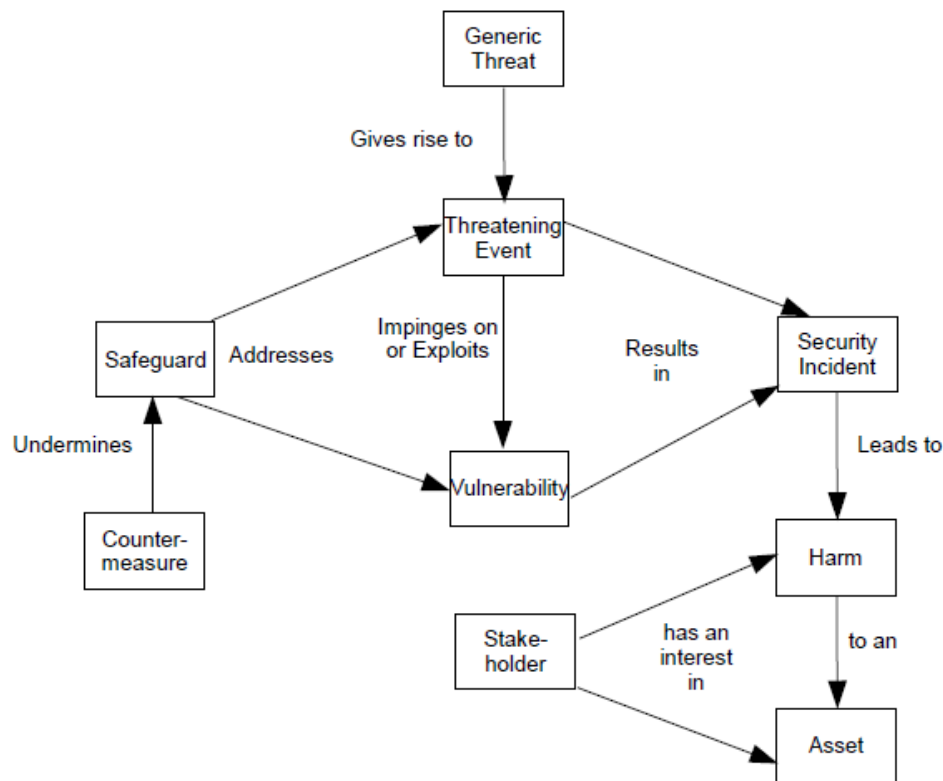


Figure 1: The conventional security model

Organisations' perceptions of cloud computing risks have been investigated in a number of articles [45]. The literature offers various approaches to identifying and structuring data risks confronting organisations. An upbeat article that is very widely cited is [3]. It identified ten security *Obstacles* and *Opportunities*, of which two were data related. They argued that *Data Lock-in* could be addressed by standardising APIs, and by compatible software to enable *Surge or hybrid Cloud Computing*; and that *Data Confidentiality and Auditability* issues could be addressed by deploying encryption, vLANs and firewalls. The paper's analysis of data risk was inadequate, e.g. it used the inappropriate concept of *data theft*, and it failed to encompass the outright loss of data.

Paquette et al. [38] considered the use of cloud computing by US government agencies specifically, and proposed a four-element framework: access, availability, infrastructure and integrity. Hardy & Williams [26], on the other hand, used a six-element risk framework, comprising continuity, compliance, auditability, reputation, intellectual property and content risks. Subashini & Kavitha [42] provided a comprehensive discussion of security issues in SaaS offerings, but the structure imposed on the ideas was very muddy. They identified a range of risks in relation to unauthorised access to data in storage, variously by hackers, by unintended members of the organisation's own staff, by other users of the service-provider's facilities, and by the service-provider's staff – but, remarkably, the authors overlooked access by the service-provider and by governments. Other relevant issues identified were interception during transmission, jurisdictional location (as a matter of legal compliance), data integrity, and data availability – but, like [3], they overlooked outright data loss.

All of [1]-[2], [10] and [13] built on the analysis of Aviziensis et al. [5]. The Ackermann Security Risk Items and Dimensions are reproduced in Table 1. Although the Ackermann model has some advantages for researchers, it conflates threats, vulnerabilities and safeguards, and does not provide useful guidance to organisations that are considering the adoption of cloudsourcing. It also combines IT security, service security, and data security into one melange. Ackermann's Risk Items 10-12 and 15-19 are service security risks, and 20-21, 25-27 and 29-31 are IT security matters. This paper is concerned with the 15 Items in the Ackermann list that are Data Risks.

A range of initiatives have been commenced within the cloudsourcing service-provider industry sector [7] and beyond it [9]. More pragmatically conceived approaches exist, such as a *Cloud Computing Bill of Rights* [47]. An early but fairly comprehensive analysis is in [37]. There are also some signs of new technologies that may deny access to data by cloudsourcing service-providers, such as Eben Moglen's FreedomBox [18].

Current models and current services are widely recognised in the trade press as falling short of the need. The following section of this paper applies Security Risk Assessment in order to develop a framework for understanding cloudsourcing data risk. Unlike Ackermann et al., the focus of this work is on the delivery of value to practitioners rather than to researchers.

Table 1: The Ackermann security risk Items and dimensions. Extract from [2]

Final Set of Security Risk Dimensions and Security Risk Items			
ID	Brief Risk Description: Risk of...	ID	Brief Risk Description: Risk of...
Confidentiality Risks		Performance Risks	
1	... eavesdropping communications	16	... network performance problems
2	... supplier looking at sensitive data	17	... limited scalability
3	... disclosure of data by the provider	18	... deliberate underperformance
4	... disclosure of internal system data	19	... performance issues of internal systems
Integrity Risks		Accountability Risks	
5	... manipulation of transferred data	20	... identity theft
6	... data manipulation at provider side	21	... insufficient user separation
7	... accidental modification of transferred data	22	... insufficient logging of actions
8	... accidental data modification at provider side	23	... access without authorization
9	... data modification in internal systems	24	... missing logging of actions in internal systems
Availability Risks		Maintainability Risks	
10	... discontinuity of the service	25	... limited customization possibilities
11	... unintentional downtime	26	... incompatible business processes
12	... attacks against availability	27	... incompatible with new technologies
13	... loss of data access	28	... limited data import
14	... data loss at provider side	29	... proprietary technologies
15	...insufficient availability of internal systems	30	... insufficient maintenance
		31	... unfavorably timed updates

4 A Generic Data Risk Assessment of Cloudsourcing

To support user evaluation of cloudsourcing proposals, it is necessary to achieve clarity about the nature of relevant Assets and the Harm that they may suffer. Threat and Vulnerability Analysis can then be grounded in that understanding, existing Safeguards can be evaluated, and additional and enhanced Safeguards can be conceived.

4.1 Assets

The term data is used here to refer to any symbol, sign or measure that is in a form capable of being directly captured by a person or a machine. It may represent some phenomenon in the real world, either by resulting from a measurement of it, or from being postulated as indicating something about it. Alternatively, it may be synthetic data that has no such direct relationship, such as the data used in a Monte Carlo simulation. The term *data* is used in this paper in preference to the term *information*, because it is more useful to limit the term information to data that has value, in particular value arising from relevance to a context such as a resource-allocation decision.

Data is subject to a range of quality factors, which bear on its value as an asset. One of these is its capacity to be relevant to some future decision. Other data quality factors include accuracy, precision, completeness and timeliness. Over time, the quality of any particular item of data may diminish. One reason may be because the real-world phenomenon is subject to change over time, but the recorded data does not reflect that change. Data may also lose quality as a result of processing that takes place in the interim, particularly through the alteration of the data, or the alteration or deletion of other data associated with it. The term *data integrity* is commonly used to refer to the condition in which data quality is sustained.

In order to understand data as an asset, it is important to take into account the distinction between data and the medium on which it is recorded. Another factor is the ready replicability of data, particularly in digital form. For these reasons, data is not an asset of the same kind as real estate or chattels (i.e. goods, made of atoms). Intellectual property laws, in particular copyright, create baskets of rights in relation to data, and those rights can be owned and sold; but the data itself is not an asset to which the notion of *ownership* applies. Rather than data ownership, it is more appropriate to apply such concepts as data possession and data control. Data protection laws are commonly cast in terms of a *data controller*. A data controller may be a corporation, a government agency or a not-for-profit organisation, and may be of any size (conventionally, micro, small, medium or large); or it may be an individual.

When determining the value of a data asset, organisations are almost entirely concerned with economic factors. For individuals, on the other hand, there may be an economic dimension, but more commonly their predominant concerns are social factors and psychological values, including hedonism or pleasure-value. Individuals may often be less concerned about data quality factors than organisations, although low quality data in the hands of an organisation may cause individuals difficulties. Organisations generally want data to be persistent, whereas a significant proportion of the data that individuals are interested in is ephemeral, as is evidenced by the immediacy of social media content, and the scant attention paid to retention and archival.

Reflecting the uses to which organisations and individuals put data, the following sources of value can be distinguished:

- Intrinsic Value, arising from its direct support for the recognition of value, as in debtors ledgers and share registers
- Operational Value, arising from the data's usefulness in performing a function, such as inventory management and scheduling meetings
- Competitive Value, arising both from the data's usefulness to the data controller and its potential usefulness to other parties, including economic competitors and strategic competitors
- Reputational Value, arising from the data's capacity to influence the perceptions that other parties have of the data controller or of some other party
- Compliance Value, arising from legal obligations in relation to the protection of the data
- Personal Value, arising from the concerns of a person to whom the data relates (such as health, financial or criminal record data), whether or not the concerns are rational, and whether the value is associated with economic, social or psychological factors

Central though data is to this analysis, it is not the only Asset relevant to an evaluation of cloud sourcing. Data controllers, and parties to which the data relates, have a range of assets that the mis-handling of data can affect. This aspect is further discussed in the following sub-section.

4.2 Harm

Reflecting the literature outlined earlier, the following list identifies a set of five categories of harm to data that need to be taken into account when selecting among in-, out- and cloud sourcing options.

- Data Loss
- Data Inaccessibility
- Unauthorised Data Modification / Loss of Data Integrity
- Unauthorised Data Access
- Unauthorised Data Replication

As will be shown in the following sub-sections, cloud sourcing creates additional risk exposures. The degree of harm varies greatly depending on a variety of factors. For example, a 5-minute period of inaccessibility to accounting data is of a completely different order of magnitude of harm in comparison with the unauthorised replication of a large database of sensitive personal data that is of sufficient richness to support identity fraud.

Because data serves important purposes, and has a number of different values associated with it, it is necessary to also consider categories of Harm arising to other Assets as a result of Harm to a data Asset. The most direct impacts will, however, generally be on the data controller itself. However, harm to the interests of a variety of dependent parties also needs to be taken into account. For example, an airline may be negatively affected by loss of data by a company that maintains its aircraft; and an individual to whom data relates may be harmed by ill-informed decision-making by a corporation or government agency. The following kinds of harm can be caused to organisations of both kinds when data is subject to a Security Incident:

- Degraded Operational Capacity
- Degraded Customer Service Quality
- Reduced Asset Value (e.g. debtors ledger or prospects database; or asset theft)
- Reduced Revenue
- Increased Costs

- Damaged Reputation, incl. Confidence of Customers, Investors and Regulators
- Negative Privacy Impact on Individuals (e.g. customers, employees), incl. personal safety
- Non-Compliance with Obligations or Commitments

Compliance is an important aspect that has been inadequately treated in many discussions of cloudsourcing. The scope and significance of negative impacts on legal compliance by both data controllers and dependent parties includes the following:

- General Statutory & Common Law Obligations
 - Evidence Discovery Law
 - Financial Regulations
 - Company Directors' obligations re asset protection, due diligence, business continuity, risk management
 - Security Treaty Obligations
- Confidentiality
 - Corporate Strategic
 - Corporate Commercial
 - Governmental
 - Personal
- Privacy
 - Unauthorised Use (whether by the data-controller, a service-provider, or a third party)
 - Unauthorised Disclosure by any party that gains access to it (*data breach*)
 - Storage in Data Havens that have limited data protection safeguards (e.g. India, the USA)

4.3 Threats

A wide variety of Threats exist, conventionally divided into three categories:

- *Acts of God or force majeure*
- Unintentional human and device errors
- Intentional Threats, also referred to as Attacks

From the viewpoint of Risk Management, however, Threats are more usefully catalogued according to the organisational location in which they arise, and in which they need to be addressed. In Table 2, the key categories of Threat are listed in the rows, and the categories of Harm identified above are shown in the columns. The Threats are clustered according to the organisational location in which they arise. By *1st Party* is meant a data controller or dependent party. The term *2nd Party* refers to the cloudsourcing servicing provider, and the term *3rd Party* applies to all other organisations and individuals. In the table-cells, a Y (for Yes) indicates that that particular category of Threat is capable of giving rise to that particular category of Harm.

Within the 1st Party cluster, the data-controller or a dependent party may suffer harm from accidents due to errors in the design or performance of business processes, and from attacks by insiders through abuse of the privileges granted to them as users. The incidence of business process error is likely to be higher with SaaS, because the fit of the application to the organisation's needs is likely to be lower, and the application is likely to be less adaptable as those needs change.

The second bracket of categories refers to actions within the realm of the service-provider(s). Storage error might, with low probability, result in data modification, or unauthorised access or replication (e.g. as a result of errors in the permissions lists). A much more likely eventuality is that data may be inaccessible by the organisation for periods of time due to outages attributable to the service-provider's storage facilities [34]. More severe consequences are likely to arise from loss of the data due to unrecoverable hardware failure. This has occurred with nominally reputable providers like Amazon [6], [17]. Far from being unusual, data loss appears to be a frequently-occurring problem [30] and in the rankings published by the Cloud Security Alliance has been raised to the Number 2 security threat [32].

Table 2: Correlation of threat and harm categories

Harm: Threat:	<u>Data Loss</u>	<u>Data Inaccessibility</u>	<u>Data Modification</u>	<u>Data Access</u>	<u>Data Replication</u>
<u>1st Party</u>					
Business Process Error	Y	Y	Y	Y	Y
Abuse of Privilege	Y	Y	Y	Y	Y
<u>2nd Party</u>					
Storage Error	Y	Y	Y	Y	Y
Availability Failure	Y	Y	Y	–	–
Network Malfunction	–	Y	Y	Y	Y
Interception	–	–	–	Y	Y
Abuse of Privilege	Y	Y	Y	Y	Y
Data Incompatibility	Y	Y	Y	–	–
<u>3rd Party</u>					
Hacking	Y	Y	Y	Y	Y
Injunction	Y	Y	Y	Y	Y
Government Powers	Y	Y	Y	Y	Y
DoS Attack	–	Y	–	–	–

This threat is particularly significant in the case of SaaS. After an organisation has adopted SaaS for, say, its office applications, a single server, database, network or power outage renders unavailable the office applications, office documents, mail-archives, appointments and address-books of every staff-member, not merely those staff-members local to the point-of-failure. Cloudsourcing's effect is *one out, all out* [37].

Data inaccessibility may arise from a brief failure of the service as a whole (e.g. due to power outage or loss of connectivity), or through a temporary network malfunction or overload. Disturbances of such kinds might alternatively result in modification (e.g. recovery to an earlier database state). There is also the possibility of longer-term inaccessibility. This might arise from a suspension of service while a liquidator undertakes sale or barter of the service, or merely the data, as a means of recovering monies owed to the service-provider's creditors. (Undertakings previously given in relation to data are commonly ignored during and after a change in ownership of the service-provider, its business, or its database). Outright data loss has arisen in a variety of cases, where the service-provider simply closes its doors, or withdraws the service, as occurred with Google's Postini cloud backup service [46].

Network malfunction might give rise to unauthorised modification of data, or to unauthorised access to or replication of data (e.g. through delivery to an inappropriate location). Interception of traffic between the data-controller and the service-provider could also result in unauthorised access or replication. These exposures are broader than in the case of conventional outsourcing, because of the likelihood of geographical dispersion of the hosts that are providing the virtualised servers.

Abuse of Privilege by the service-provider is an ever-present possibility, which could result in any of the various forms of harm to the data. A specific instance of such abuse is Verizon's scanning of user's data [22].

A further threat arises from the possibility that the data may be formatted in a manner that is compatible with a particular service, but not with any alternative services. This could give rise to delays in accessing the data, unauthorised data modification due to faulty conversion to a new format, or even complete inability to access the data, equivalent to loss of the data.

The final group of threats relates to parties other than the data-controller and the service-provider(s). A break-in may be followed by unauthorised data access (effectively small-scale copying) or unauthorised replication (large-scale copying). Unauthorised modification could occur. A special case of modification that has recently been in evidence,

and that gives rise to data inaccessibility, is *data-napping*, whereby the hacker encrypts the data, and extorts a fee in return for the decryption key [28]. A malicious hacker may, on the other hand, simply delete the data and perhaps seek out backups and delete them as well. Cloud computing has also given rise to a specialised form of hacking, which is referred to as *isolation failure* or a *guest-hopping attack* [35]. A party that has processes running in the same host may be able to gain access to the data associated with another party, enabling any of the actions described earlier in this paragraph to be undertaken.

Where services are insourced, the threats indicated in Table 2 as 2nd party threats are the direct responsibility of the data controller. With any form of outsourcing, the data controller loses direct control over the data, becomes dependent on one or more service-providers, is subject to increased threats to the extent that the data is transmitted further and more often, and is subject to the additional threat of abuse of privilege by service-providers and their employees. With cloud sourcing, the threats expand further, in that the locations of storage and processing are no longer known to the data controller, and hence transparency, oversight and auditability are undermined.

4.4 Vulnerabilities

The technical vulnerabilities inherent in outsourcing are exacerbated by cloud sourcing. The key factors involved are as follows:

- Large Numbers of IT Components
- The Geographical Dispersion of IT Components
- The Complexity of the Network of IT Components
- The Rapidity with which Changes occur in the Network of IT Components
- Substantial, often Hidden Dependencies on both IT and Infrastructure Technologies, including self-induced harm or *auto-immune disease* [27]
- The Fragility of the Services, as a result of cascade effects, with one outage triggering others [8]

These factors inevitably give rise to reliability issues. In [12], over 100 media reports about cloud service outages were assessed in order to gain an understanding of the frequency, length, consequences and redress aspects of cloud sourcing service reliability and data security. Beyond short-term inaccessibility, the proportion of times in which data loss occurred appears to have been as high as 20% of the 49 outages documented in media reports during the period 2005-11.

Beyond the technical vulnerabilities are operational and commercial factors. The most prominent such vulnerabilities that arise from cloud sourcing are:

- Fixed legal Terms of Service that may not fit the data-controller's needs
- Fixed Service Level Agreements (SLAs) that may not fit the data-controller's needs
- Inadequate Internal Expertise
- Inadequate Transparency, Oversight and Auditability, characterised by the encouragement by service-providers to *set and forget* [49]
- Unknown Physical Location of Data, which undermines transparency, oversight and auditability
- Unknown Jurisdictional Location of Data. This has been determined by various governments to preclude the use of public cloud services by government agencies [16], [29], and also directly affects some private sector organisations, particularly in the banking sector [24], [44]
- Physically Remote Service-Provider(s)
- Jurisdictionally Remote Service-Provider(s)
- Large Service-Provider(s) for which the data controller's business is relatively unimportant
- Dependence / Lock-in / Non-compatible formats / Lack of substitutable services

- Accessibility by Governments foreign to the data-controller

The extent of Harm to data that can arise from these Vulnerabilities depends on the approach that the data controller adopts to data management. As indicated earlier in this paper, the use of cloud sourcing for secondary copies of data results in some additional Vulnerabilities, whereas far more substantial issues arise where the authoritative copy of data is in the cloud, and even greater exposures exist if the sole copy of data is entrusted to a cloud sourcing provider.

In the case of insourcing, the data controller has the capacity to directly manage these risks. With any form of outsourcing, the control becomes indirect, and dependent on contractual terms and the service-provider's conformance with those terms. With cloud sourcing, the terms are generally looser, and dictated by the service-provider rather than being customised to meet the data-controller's needs. The physical and jurisdictional location of the service-provider, the contract and the data also tend to become more remote from the data-controller.

A factor that is of major consequence in some circumstances is the scope for interference by governments. A standard is under development to facilitate law enforcement agency access to data in the cloud [20]. However, government actions may or may not be authorised by law, and may or may not be mediated by the judiciary through court orders or warrants. In some cases, governments may also claim extra-territorial reach. This is particularly so with the USA, under its PATRIOT and Foreign Intelligence Surveillance Amendment (FISA) legislation [19]. The US asserts that all data stored by any US corporation, nomatter where in the world it is stored, is subject to US government demand powers. This is far from a mere theoretical possibility, as demonstrated by the (probably unlawful) closure of Megaupload's services by New Zealand law enforcement agencies at the behest of the US [23].

4.5 Key Data Risks in the Cloud

On the basis of the analysis conducted in the preceding sub-sections, it is possible to identify some threat-vulnerability combinations that are of particular concern to data controllers and dependent organisations and individuals considering the adoption of cloud sourcing. Table 3 provides an overview of them using the structure introduced in Table 2 above.

Table 3: Key data risks in the cloud

Harm: Threat:	<u>Data Loss</u>	<u>Data Inaccessibility</u>	<u>Data Modification</u>	<u>Data Access</u>	<u>Data Replication</u>
<u>1st Party</u>					
Business Process Error	Y	Y	Y	Y	Y
Abuse of Privilege	Y	Y	Y	Y	Y
<u>2nd Party</u>					
Storage Error	Y	Y	Y	Y	Y
Availability Failure	Y	Y	Y		
Network Malfunction		Y	Y	Y	Y
Interception				Y	Y
Abuse of Privilege	Y	Y	Y	Y	Y
Data Incompatibility	Y	Y	Y		
<u>3rd Party</u>					
Hacking	Y	Y	Y	Y	Y
Injunction	Y	Y	Y	Y	Y
Government Powers	Y	Y	Y	Y	Y
DoS Attack		Y			

The reasons for highlighting these aspects are as follows:

- Storage Error and Availability Failure:
 - If short-term, this may seriously impact on operational capacity and customer service quality, with contingent effects on reputation
 - If long-term, this may constitute outright data loss, may thereby impact severely on asset values, and may result in non-compliance with significant regulatory obligations

- Network Malfunction may give rise to similar harm to short-term Storage Error and Availability Failure
- Abuse of Privilege by the service-provider or its employees may result in data being exploited by that party or others with whom they deal, in ways that are harmful to the data-controller or other parties. The nature of the Terms applied by most commercial SaaS providers makes this a matter of particular concern
- Data Incompatibility creates a significant barrier to the multi-sourcing of services, adds to switching costs, and implies delays in access to data when a change in service-provider is undertaken
- Hacking, including Isolation Failures and Data-napping, are a potentially serious concern
- Exercise of Government Power may result in reduced operational capacity, and in breach of confidence in relation to sensitive data. It may also highlight non-compliance with data export provisions. The exposure exists in a wide variety of circumstances, but is exacerbated by trans-jurisdictional clouds, by the placement of hubs in data havens, and by claims by some governments of extra-territorial reach

4.6 Safeguards

Some safeguards are natural, such as the technical challenges that confront casual hackers, and the costs involved in mounting some kinds of attacks. Some safeguards are mainstream, and engrained in corporate and individual behaviour, such as locking doors and authenticating people who seek access to data.

There are also incentives that encourage the implementation of safeguards. In particular, cloud service-providers need to provide a sufficient appearance of reliability that they can attract and retain customers. Some of their clientele are technically and commercially capable, and others hire technically and commercially competent consultancies to assess suppliers' capabilities. It would therefore seem reasonable to expect that some basic level of security would be a feature of all cloud services. Unfortunately, that expectation is somewhat undermined by such studies as have been undertaken to date of Terms of Service, SLAs, standard practices and security incidents.

The following sub-sections consider the extent to which technical and operational safeguards appear to be comprehensive and effective, and the degree to which legal safeguards are able to fill the gaps. The primary focus is on the key data risks identified in Table 3.

4.6.1 Technical Safeguards

Safeguards tend to be fairly specific in the threat/vulnerability combinations that they address. For example, the threat of Data Interception in Transit can be addressed by channel encryption – subject to the qualification of its susceptibility to man-in-the-middle attacks. But channel encryption does nothing to mitigate the risks of breaches of security by the service-provider, insiders, hackers and governments. Similarly, encryption of data in storage represents a safeguard against unauthorised access by a hacker to the complete data-set, but not against abuse of privilege by the data-controller or its agents, or by a service-provider, resulting in loss, inaccessibility or unauthorised modification.

Moreover, if the data is to be processed in the cloud, rather than merely stored and recovered, then it needs to be decrypted on the service-provider's device, which exposes the data to unauthorised access during the period when it is being processed, and also exposes the encryption and decryption keys. This creates vulnerability to third parties, but also to actions by the service-provider, including insecure key management processes.

Another important safeguard against unauthorised modification, access and replication is access control. Passwords are an increasingly weak form of authentication, but the alternatives to date remain expensive or inconvenient. A significant challenge exists in making user authentication processes convenient for people who are authorised, yet very difficult for people who are not.

A rich set of tools exists to enable service-providers to resist the efforts of hackers. On the other hand, data centres that support cloud sourcing are honey-pots of data that attract hacker-bees. There is, and will continue to be, an arms race of safeguards, which stimulate countermeasures, which demand further safeguards, etc. Inevitably, some attempted break-ins will succeed.

Various applications of the redundancy principle are relevant. However, replication of the data across multiple locations, and even across multiple service-providers, while mitigating the risk of loss and inaccessibility, increases the risk of unauthorised access. Multi-sourcing remains very challenging at this stage, although some progress has been made in inter-operability protocols and standards [41].

4.6.2 Organisational Safeguards

Many risk exposures arise from human behaviour. Conventional organisational measures used to address them include staff-selection and training, and careful design of the manual aspects of business processes and of human-computer interfaces. Cloudsourcing generally creates more challenges in implementing and sustaining these forms of organisational control, because the services tend to be less well-customised and hence a poorer fit to the organisation's and the individual users' needs.

Application of the redundancy principle in this area results in process controls such as split responsibilities, dual-entry, reviews, approvals, and reconciliations. These also tend to suffer in a SaaS environment because of the relatively unresponsive nature of applications to the organisation's changing needs.

A crucial aspect of organisational controls is oversight and audit of the processes and outcomes, and particularly of the controls. But oversight and audit are undermined in the cloud sourced environment, because of the lack of transparency, resulting in only a limited amount of information being available to enable the checking to be performed.

A vital application of the redundancy principle is local replication of data, and fallback procedures to enable some continuity of business and customer service during outages. This is generally difficult to achieve with cloudsourcing.

4.6.3 Legal Safeguards

Enforceable legal obligations are only a fallback safeguard, albeit an important one. For them to be effective, a number of conditions must be satisfied. These are summarised in the following list:

- Initial Due Diligence:
 - Checks with reference sites and/or independent certification of quality of service
 - Legal Terms of Service that are expressed operationally
 - Legal Terms of Service that are a reasonable fit to the data-controller's needs, rather than fixed and non-negotiable
 - Service-Level Agreements that are expressed operationally
 - Service Level Agreements that are a reasonable fit to the data-controller's needs, rather than fixed and non-negotiable [25]. (Even a 99% uptime commitment permits a 7-hour outage each month without recompense)
 - Specification of Dispute Resolution Processes
 - Access to tribunals and courts local to the Data-Controller
 - Liquidated Damages Provisions. (Recompense for outages beyond, say, 7 hours p.mo. is commonly limited to a small credit based on the fees the user-organisation pays, not based on the damage done to the user-organisation and its customers)
 - Protections for dependent parties, importantly including access to Dispute Resolution Processes and redress. (The doctrine of privity of contract represents a considerable barrier to the exercise of rights by organisations and individuals that are once-removed from cloudsourcing arrangements)
 - Assurance that the service-provider will comply with the exercise of government powers only subject to the due processes of law
- Ongoing Due Diligence:
 - Performance Monitoring and evidence gathering
 - Oversight and Audit, periodically and when needed
 - Periodic Re-Certification

- Redress Provisions:
 - Effective Operation of dispute resolution procedures
 - Enforcement Mechanisms that are inexpensive but effective

However, it is highly unusual for cloud sourcing services to satisfy those requirements: "Today's internet feudalism ... is ad hoc and one-sided. We give companies our data and trust them with our security, but we receive very few assurances of protection in return, and those companies have very few restrictions on what they can do." [39]

5 Implications

This survey of data risk in the cloud has suggested that many threat/vulnerability combinations exist for which the existing safeguards appear to be far from adequate. Some user-organisations conduct security risk assessments and institute risk management plans that protect their interests. On the other hand, many organisations that have adopted cloud sourcing, especially SaaS, have done so without careful consideration of the risks involved, and without a clear understanding of what Harm will arise to what Assets when what contingencies occur. It is no surprise that many governments and many corporations remain sufficiently concerned about the security of cloud sourcing that they have taken conservative approaches e.g. by applying it only to relatively small, non-core applications, or even avoiding adoption at this stage in the maturation of cloud sourcing. Under current conditions, that would appear to be the appropriate approach.

In the academic arena, a considerable amount of research is being conducted, but much of it is conducted within the frame of reference set by the service-provider sector. It is insufficiently sceptical, and insufficiently reflective of the interests of user organisations. An important implication of the research reported in this article is that academics need to become more attuned to the needs of data-controllers, and to focus much more on ways to assess and manage the data risks inherent in cloud sourcing. The analysis conducted in section 4 identifies a wide range of specific matters in need of detailed empirical research. Table 3 indicates particularly high priorities for high-value academic research.

6 Conclusions

The analysis presented in this paper has applied conventional risk assessment methods in the specific context of cloud sourcing, in order to provide a generic analysis of data risks in the cloud. Each organisation needs to apply the analysis, and the risk assessment process and tools presented above to the specifics of its own circumstances.

The analysis has considerable implications for practitioners. Much more needs to be done by industry associations and service-providers, but possibly also by parliaments and regulators. One approach to identifying the minimum requirements of cloud sourcing's management of data risks is to consider the responsibilities of company directors and their equivalents, and of senior executives of corporations, government business enterprises, and government agencies. They have legal obligations relating to the fulfilment of the organisation's mission, the definition and pursuit of strategic advantage, risk assessment and risk management, compliance, and business continuity. There are many circumstances in which directors could be readily found to be in breach of their responsibilities by adopting cloud computing, at least without a substantial risk management plan in place that brings the levels of data risk back within reasonable bounds. Whichever that statement remains true, cloud sourcing remains unready for *prime time*.

The research literature appears not to be serving practitioners as well as it should, Researchers are running on a parallel track to practitioners, and conducting research and publishing papers that are of interest to one another rather than to organisations and individuals outside academe. As an antidote to that malaise, it is proposed that researchers should focus on the important elements identified in the above analysis, and in particular on the high-priority issues identified in Table 3.

References

- [1] T. Ackermann, A. Miede, P. Buxmann, and R. Steinmetz. (2011, June) Taxonomy of Technological IT Outsourcing Risks: Support for Risk Identification And Quantification. Proc. ECIS, Paper 240. [Online]. Available: <http://aisel.aisnet.org/ecis2011/240/>.
- [2] T. Ackermann, T. Widjaja, A. Benlian, and P. Buxmann. (2012, December) Perceived IT security risks of cloud computing: Conceptualization and scale development. AIS Electronic Library. [Online]. Available: <http://aisel.aisnet.org/icis2012/proceedings/ISSecurity/3/>.
- [3] M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R. Katz, A. Konwinski, and M. Zaharia, A view of cloud computing, Communications of the ACM, vol. 53, no. 4, pp. 50-58, 2010.
- [4] Australian Defence Signals Directorate. (2012) Information Security Manual. Australian Defence Signals Directorate. [Online]. Available: <http://www.asd.gov.au/infosec/ism/>.

- [5] Avizienis, J.C. Laprie, B. Randell, and C. Landwehr, Basic concepts and taxonomy of dependable and secure computing, IEEE Trans. Dependable and Secure Computing, vol. 1, no. 1, pp. 11-33, 2004.
- [6] H. Blodget. (2011, April) Amazon's cloud crash disaster permanently destroyed many customers' data. Business Insider. [Online]. Available: <http://www.businessinsider.com/amazon-lost-data-2011-4>.
- [7] G. Brunette and R. Mogull. (2009, December) Security guidance for critical areas of focus in cloud computing. Cloud Security Alliance. [Online]. Available: <https://cloudsecurityalliance.org/csaguide.pdf>.
- [8] S.V. Buldyrev, R. Parshani, G. Paul, H.E. Stanley, and S. Havlin. (2010, April) Catastrophic cascade of failures in interdependent networks. Nature 464. [Online]. Available: <http://www.nature.com/nature/journal/v464/n7291/abs/nature08932.html>.
- [9] D. Catteddu and G. Hogben. (2009, November) Cloud computing: benefits, risks and recommendations for information security. European Network and Information Security Agency. [Online]. Available: http://www.enisa.europa.eu/activities/risk-management/files/deliverables/cloud-computing-risk-assessment/at_download/fullReport.
- [10] R. Clarke (2010, June) Computing clouds on the horizon? Benefits and risks from the user's perspective. 26th Bled Conference. [Online]. Available: <http://www.rogerclarke.com/II/CCBR.html>.
- [11] R. Clarke. (2011, June) The cloudy future of consumer computing. 24th Bled Conference. [Online]. Available: <http://www.rogerclarke.com/EC/CCC.html>.
- [12] R. Clarke. (2012, February). How reliable is cloud sourcing? A review of articles in the technical media 2005-11. Computer Law & Security Review. [Online]. vol. 28, no. 1, pp. 90-95. Available: <http://www.sciencedirect.com/science/article/pii/S0267364911001865>.
- [13] R. Clarke. (2012, June) A Framework for the evaluation of cloud sourcing proposals. 25th Bled Conference. [Online]. Available: <http://www.rogerclarke.com/EC/CCEF.html>.
- [14] Cloud Computing Journal. (2009, January) Twenty-One experts define cloud computing. Cloud Computing Journal. [Online]. Available: <http://cloudcomputing.sys-con.com/node/612375/print>.
- [15] Common Criteria. (2012, September) Common criteria for information technology security evaluation – Part 1: Introduction and general model. Common Criteria. [Online]. Available: <http://www.commoncriteriaportal.org/files/ccfiles/CCPART1V3.1R4.pdf>.
- [16] Defence Signals Directorate (2012, September) Cloud computing security considerations. Defence Signals Directorate. [Online]. Available: http://www.asd.gov.au/publications/csocprotect/cloud_computing_security_considerations.htm.
- [17] L. Dignan. (2011, April) Amazon outage ends cloud innocence. ZDNet. [Online]. Available: <http://www.zdnet.com/amazon-outage-ends-cloud-innocence-1339313776/>.
- [18] J. Dwyer. (2011, February) Decentralizing the internet so big brother can't find you. New York Times, http://www.nytimes.com/2011/02/16/nyregion/16about.html?_r=0.
- [19] European Parliament. (2012, October) Fighting cyber crime and protecting privacy in the cloud. European Parliament. [Online]. Available: <http://www.europarl.europa.eu/committees/en/studiesdownload.html?languageDocument=EN&file=79050>.
- [20] European Telecommunications Standards Institute. (2012, April) Lawful Interception (LI); Cloud/Virtual Services (CLI). European Telecommunications Standards Institute. [Online]. Available: http://www.3gpp.org/ftp/tsg_sa/WG3_Security/TSGS3_LI/2012_45_Bratislava/SA3LI12_044.doc.
- [21] D. Firesmith. (2004, January). Specifying reusable security requirements. Journal of Object Technology. [Online]. vol 3, no. 1, pp. 61-75. Available: http://www.jot.fm/issues/issue_2004_01/column6/.
- [22] S. Gallagher. (2013, March) How Verizon found child pornography in its cloud: Scanned files using hashes of known child pornography images. Ars Technica. [Online]. Available: <http://arstechnica.com/information-technology/2013/03/how-verizon-found-a-child-pornographer-in-its-cloud/>.
- [23] N. Galvin. (2012, January) Megaupload closure hits legitimate users. The Sydney Morning Herald. [Online]. Available: <http://www.smh.com.au/technology/technology-news/megaupload-closure-hits-legitimate-users-20120122-1qc7d.html?skin=text-only..>
- [24] J. Gliddon. (2013, February) CommBank rules out public cloud storage. ITnews. [Online]. Available: <http://www.itnews.com.au/News/334362.commbank-rules-out-public-cloud-storage.aspx>.
- [25] B. Golden. (2011, November) Cloud computing and the truth about SLAs. Networkworld. [Online]. Available: <http://www.networkworld.com/news/2011/110811-cloud-computing-and-the-truth-252905.html>.
- [26] C.A. Hardy and S.P. Williams, (2010, June) Managing Information Risks and Protecting Information Assets in a Web 2.0, in Proceedings of the 23rd Bled eConference eTrust: Implications for the individual, enterprises and society, Bled, Slovenia, 2010, pp. 234-247.
- [27] J. Heiser. (2011, May) Yes, Virginia, there are single points of failure. Gartner Blog. [Online]. Available: <http://blogs.gartner.com/jay-heiser/2011/05/30/virginia/>.
- [28] S. Hicks. (2012, December) Russian hackers hold Gold Coast doctors to ransom. ABC News. [Online]. Available: <http://www.abc.net.au/news/2012-12-10/hackers-target-gold-coast-medical-centre/4418676>.
- [29] J. Hilvert. (2012, February) Conroy warns of cloud uncertainties. ITnews. [Online]. Available: <http://www.itnews.com.au/News/290431.conroy-warns-of-cloud-uncertainties.aspx>.
- [30] Investor's Business Daily. (2013, January) Cloud Computing Users Are Losing Data, Symantec Finds. Investor's Business Daily. [Online]. Available: <http://finance.yahoo.com/news/cloud-computing-users-losing-data-205500612.html>.
- [31] ISO (2008) Information Technology – Security Techniques – Information Security Risk Management ISO/IEC 27005:2008

- [32] S. Kar. (2013, March) CSA Report: Top Nine Cloud Security Threats in 2013. Cloud Time. [Online]. Available: <http://cloudtimes.org/2013/03/07/csa-report-top-nine-cloud-security-threats-in-2013/>.
- [33] P. Mell and T. Grance. (2009, October) The NIST Definition of Cloud Computing. National Institute of Standards and Technology, Information Technology Laboratory. [Online]. Available: http://pre-developer.att.com/home/learn/enablingtechnologies/The_NIST_Definition_of_Cloud_Computing.pdf.
- [34] C. Mellor. (2010, September) NetApp and TMS involved in Virgin Blue outage. The Register. [Online]. Available: http://www.theregister.co.uk/2010/09/28/virgin_blue/.
- [35] D. Molnar and S. Schechter. (2010, June) Self Hosting vs. Cloud Hosting: Accounting for the security impact of hosting in the cloud. Proc. 9th Workshop on the Economics of Information Security. [Online]. Available: http://www.weis2010.econinfocsec.org/papers/session5/weis2010_schechter.pdf.
- [36] National Institute of Standards and Technology. (2011, March) Managing Information Security Risk: Organization, Mission, and Information System View. National Institute of Standards and Technology. [Online]. Available: <http://csrc.nist.gov/publications/nistpubs/800-39/SP800-39-final.pdf>.
- [37] R. Needleman. (2011, September) Was brief Google Docs outage a tremor or a tsunami?. CNETNews. [Online]. Available: http://news.cnet.com/8301-19882_3-20103034-250/was-brief-google-docs-outage-a-tremor-or-a-tsunami/.
- [38] S. Paquette, P.T. Jaeger and S.C. Wilson, Identifying the security risks associated with governmental use of cloud computing, Government Information Quarterly, vol. 27, no. 3, pp. 245-253, 2010.
- [39] B. Schneier. (2012, November) When It Comes to Security, We're Back to Feudalism. Wired. [Online]. Available: <http://www.wired.com/opinion/2012/11/feudal-security/>.
- [40] G. Stoneburner, A. Goguen and A. Feringa. (2002, July) Risk Management Guide for Information Technology Systems. National Institute of Standards and Technology. [Online]. Available: <http://csrc.nist.gov/publications/nistpubs/800-30/sp800-30.pdf>.
- [41] Storage Networking Industry Association. (2012, June) Cloud Data Management Interface (CDMI). Storage Networking Industry Association. [Online]. Available: http://snia.org/tech_activities/standards/curr_standards/cdmi.
- [42] S. Subashini and V. Kavitha, A survey on security issues in service delivery models of cloud computing, Journal of Network and Computer Applications, vol. 34, no. 1, pp. 1-11, 2011.
- [43] D. Svantesson and R. Clarke, Privacy and Consumer Risks in Cloud Computing, Computer Law & Security Review, vol. 26, no. 4, pp. 391-397, 2010.
- [44] L. Tay. (2012, November) ANZ builds up 'cypher cloud' strategy. itNews. [Online]. Available: <http://www.itnews.com.au/Video/324020.anz-builds-up-cypher-cloud-strategy.aspx>.
- [45] Troshani, G. Rampersad and N. Wickramasinghe, Cloud Nine? An Integrative Risk Management Framework for Cloud Computing, in Proceedings of Bled e Conference, Bled, 2011.
- [46] L. Tung. (2012, January) Google clips Exchange backup service. itNews. [Online]. Available: <http://www.itnews.com.au/News/287829.google-clips-exchange-backup-service.aspx>.
- [47] Urquhart. (2010, June) The 'Cloud Computing Bill of Rights': 2010 edition. Cnet News. [Online]. Available: http://news.cnet.com/8301-19413_3-20006756-240.html.
- [48] L.M. Vaquero, L. Rodero-Merino, J. Caceres and M. Lindner. (2009, January). A Break in the Clouds: Towards a Cloud Definition. ACM SIGCOMM Computer Communication Review. [Online]. vol. 39, no. 1, pp. 50-55. Available: <http://ccr.sigcomm.org/online/files/p50-v39n1-vaqueroA.pdf>.
- [49] B. Winterford. (2011, September) Transparency: a core tenet of the cloud. itNews, [Online]. Available: <http://www.itnews.com.au/News/272605.transparency-a-core-tenet-of-the-cloud.aspx>.