Protecting Personal Information in Cloud Computing

Miranda Mowbray and Siani Pearson

Cloud and Security Lab, HP Labs, Bristol, UK
{Miranda.Mowbray,Siani.Pearson}@hp.com

Abstract. This paper gives an overview of issues in privacy protection of personal information in the cloud, and describes a variety of approaches that may be used to address these issues. Some of these approaches are available for use now; others are relatively immature, but look promising. The most appropriate approach varies according to the type of data to be processed or application to be run in the cloud.

Keywords: cloud computing, encryption, hybrid cloud, privacy, security, technology, trusted computing, virtual private cloud.

1 Introduction

Cloud computing exacerbates some privacy issues, and there is therefore a need for technological and organizational mechanisms to protect personal information within the cloud. This paper gives an overview of key techniques for addressing this problem. Some of these techniques are also relevant to the protection of corporate proprietary information or state secrets, however in this paper we focus on data privacy for individuals. Section 2 summarises the central issues and problems involved, and Sections 3 and 4 present a number of techniques for protecting personal information in the cloud, divided into those that are available for use now (Section 3), and other techniques that are relatively immature, but are likely to be provided in the future (Section 4). Section 5 discusses three issues related to the successful use of these techniques – key management, design for privacy, and accountability.

First we provide an introductory explanation about privacy and cloud computing.

1.1 Privacy

At the broadest level (and particularly from a European standpoint), privacy is a fundamental human right, enshrined in the United Nations Universal Declaration of Human Rights (1948) and subsequently in the European Convention on Human Rights and national constitutions and charters of rights such as the UK Human Rights Act 1998. There are various forms of privacy, ranging from 'the right to be left alone' [1], 'control of information about ourselves' [2], 'the rights and obligations of individuals and organizations with respect to the collection, use, disclosure, and retention of personally identifiable information.' [3] Personally identifiable information (also referred

R. Meersman et al. (Eds.): OTM 2012, Part II, LNCS 7566, pp. 475–491, 2012.

to as personal data) is information that can be used to positively identify an individual such as name, address, phone number, e-mail address, etc. Others consider privacy violations to be a variety of types of harmful or problematic activities (so that they would not consider unauthorised use of personal data to be a privacy violation if it caused no harm.) [4]. In the context of commercial services privacy for customers entails the protection and appropriate use of their personal data, and the meeting of their expectations about its use; privacy for organizations entails the application of laws, policies, standards and processes by which personal data is managed.

Some personal data elements are considered more sensitive than others, although the definition of what is considered *sensitive personal information* may vary depending upon jurisdiction and even on particular regulations. In Europe, sensitive personal information is called *special categories of data*, which refers to information on religion or race, political opinions, health, sexual orientation, trade-union membership and data relating to offences or criminal convictions, and its handling is specially regulated. In the US, social security and driver license numbers, personal financial information and medical records are commonly treated as sensitive. Health information is considered sensitive in all jurisdictions that have data protection legislation defining sensitive data.

In Europe, the European Data Protection Directive 95/46/EC [5] (and its supporting country legislation) implements the Fair Information Principles [6], along with some additional requirements including restrictions to trans-border data flows to countries with an `inadequate' strength of privacy protection. Legislation similar to the European Data Protection Directive has been, and continues to be, enacted in many other countries. In contrast, the US does not have a comprehensive regime of data protection but instead has a variety of laws —such as the Health Insurance Portability and Accountability Act (HIPAA) — which are targeted at the protection of particularly sensitive types of information. This US approach to privacy legislation is historically sector-based or enacted at the state level and places few if any restrictions on trans-border data flow. The EC considers the US to be 'adequate' for data transfer only under the limitation of the Safe Harbor agreement [7].

At the time of writing, privacy regulations, enforcement activities and sanctions are currently increasing the world over. The US is introducing a Consumer Privacy Bill of Rights [8] and the EU is revising their Data Protection Directive and regulation [9], with the result that FTC enforcement will be strengthened within US and current plans are that European Data Protection Agencies will be able to impose fines of up to 2% of worldwide annual turnover to companies that do not have mechanisms in place to underpin regulatory data protection compliance [9]. Other potential consequences of privacy failure for data controllers include fines or even imprisonment as a consequence of civil and criminal liability, investment risk, loss of business continuity, and reputational damage.

In addition to privacy-specific regulations, other laws and developments in business also impact privacy. The US Patriot Act has the effect that data that is stored on US computers may be viewed by US security services without the data owner being informed – and this may not be compatible with Safe Harbor provisions. The business models of some social media companies are based on earning revenue from their

customers' personal data, so that they have a disincentive to make it easy for their customers to keep this data private. More generally, the possibility of using data to improve enterprises' effectiveness and profitability has led to the Big Data trend: every 18 months, enterprises are doubling the amount of digital data that they store [10], and they are increasingly interested in technologies for harvesting actionable information from this data.

1.2 Cloud Computing

A definition of cloud computing that is commonly accepted is provided by the United States National Institute of Standards and Technologies (NIST):

> "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction." [11]

There are different layers of cloud services that refer to different types of service model, each offering discrete capabilities. The service offered may be the delivery of computing resources such as storage and computing power, where the customer rents virtual machines from the service provider rather than buying hardware and software to provide these resources in the customer's own data centre (this is known as Infrastructure as a Service), the delivery of a solution stack for software developers (Platform as a Service) or the delivery of software applications available on demand and paid for on a per-use basis (Software as a Service).

In addition there are several deployment models for cloud computing, of which the main ones are:

- **Private:** a cloud infrastructure operated solely for a single organisation, being accessible only within a private network and being managed by the organisation or a third party (potentially off-premise)
- **Virtual private:** a private cloud existing within a shared or public cloud
- **Shared:** a cloud that is open to use by selected organisations; for example, it could be a *community* cloud (a cloud infrastructure that several organizations share between themselves by mutual agreement) or a *partner* cloud (cloud services offered by a provider to a limited and well-defined number of parties)
- **Public:** a publicly accessible cloud infrastructure
- **Hybrid:** a composition of two or more clouds that remain separate but between which there can be data and application portability

2 Privacy Issues for Cloud Computing

It follows from the discussion in the previous section that there are privacy issues for cloud computing, because of the potential for unwanted exposure of personal data in the cloud, and because there can be some difficulties in ensuring and/or verifying

compliance of cloud computing services to privacy laws and policies. This can happen for a number of reasons.

First, in cloud computing, data is processed on machines that end users and customers do not own or control. It may be difficult for a cloud service provider (CSP) to ensure that a data subject can get access to all his/her personal data, and there can be lack of transparency about where such data is, who has rights to use it and what is being done with it. Data proliferation is a feature of cloud computing, and this happens in a way that may involve multiple parties and is not controlled by the data owners. CSPs ensure availability by replicating data in multiple data centres. Another reason is that it can be difficult to control (or even know) the exposure of the data that has been transferred to the cloud. For instance, information passing through some countries (including the US) can be accessed by law enforcement agencies. A CSP may not comply with a request for deletion of data, and it is not necessarily clear who controls retention of data. It can also be difficult to get data back from the cloud. There are uncertainties about notification, including of privacy breaches, and ability to obtain redress. It can be difficult to know that privacy breaches have occurred, and to determine who is at fault in such cases. It also can be unclear what rights in the data will be acquired by data processors and their sub-contractors, and whether these are transferable to other third parties upon bankruptcy, takeover, or merger. Some of these uncertainties can be resolved through clearly-worded contracts for cloud services.

Second, there is a threat of theft or misuse of customers' data by co-hosted customers, rogue employees of the service providers, subcontracted services, foreign governments, or attackers breaking into the networks of any of these. Data in the cloud can be encrypted or anonymised to protect against this threat, however there are possible key management issues with encryption, which will be discussed further in Section 5, and de-anonymisation techniques can recover personal data from some types of anonymised data sets [12].

Third, there are legal restrictions on the processing of personal data. For example, in EU law there are constraints and obligations related to notification, data minimization, marketing opt-out, security and geographical restriction of data flow. These restrictions are stronger for sensitive data. In general, the legal situation is subject to change: legislation has not yet been updated to address privacy challenges in the cloud, and for example it is not clear whether the procedure of anonymising or encrypting personal data to enhance users' privacy is exempt itself from privacy protection requirements and courts have not yet ruled many cases specifically related to cloud computing. To further complicate matters, data may travel in the cloud through multiple legal jurisdictions. It is difficult to guarantee that a copy of the data or its backups are not stored or processed in a certain jurisdiction, or that all these copies of data and virtual storage devices are deleted if such a request is made, especially with regards to device reuse. Legal obligations on organizations and individuals who process other peoples' personal data using cloud services vary according to context, and particularly on jurisdiction and the sensitivity of the information involved. However, they commonly require the adoption of technical and procedural safeguards to protect the data, in addition to other mechanisms (due diligence on the CSP's capabilities, legally sufficient contracts with CSPs, restrictions of use, and notification of privacy breaches) [13].

Cloud computing faces many of the same problems as traditional outsourcing, yet the dynamic nature of cloud makes many existing provisions to address this in more static environments obsolete or impractical to set up in such a short time scale. Model contracts are one example of this. It is not clear which party is responsible (statutorily or contractually) for ensuring legal requirements for personal information are observed, or appropriate data handling standards are set and followed, or whether they can effectively audit third-party compliance with such laws and standards. Neither is it yet clear to what extent cloud sub-contractors involved in processing can be properly identified, checked and ascertained as being trustworthy, particularly in a dynamic environment.

To summarise, cloud computing offers significant challenges for organisations that need to meet various global privacy regulations, including the necessity of legal advice on complex global privacy legislation. Cloud computing faces the same privacy issues as other service delivery models, but several of these issues are magnified for cloud computing, especially trans-border data flow restrictions, liability questions, and the difficulties of knowing the geographic location of processing and which specific servers or storage devices will be used.

More broadly, security is an aspect of privacy (in the sense that reasonable security must be used to protect personal data) and so many security issues, including the difficulties in enforcing data protection within cloud ecosystems, lack of training and expertise and unauthorised usage, may be seen to also be privacy issues. At the network, host and application levels, security challenges associated with cloud computing are generally exacerbated by cloud computing although not specifically caused by it. It can be unclear which parties are responsible for which aspects of security. This division of responsibility is hampered by the fact that cloud APIs are not yet standardised. Data security issues for cloud computing depend on the particular service provision and deployment models, but include the risk of data loss, unauthorised collection, unauthorised use, and inadequate protection of the data by the CSP.

Further details about privacy issues for cloud computing may be found in [14]. NIST [15] has recently provided an overview of the security and privacy challenges facing public cloud computing, including an assessment of threats and recommendations to organizations considering outsourcing data, applications and infrastructure to a public cloud environment. Similarly, other guidelines have been produced by ENISA [16] and the Cloud Security Alliance (CSA) [17]. More broadly, trust, security and privacy issues for the cloud are explained and assessed in [18]. Generally speaking, this prior art identifies and categorises privacy and security issues in the cloud, whereas the focus of this paper is on describing solutions to such issues.

3 Protecting Personal Information in the Cloud: Current Approaches

In this section we consider various approaches that can be used to address the privacy issues described above. These approaches involve both technical and non-technical (e.g. procedural) elements. In addition to being used by customers to protect their own data, they may be used to protect the personal data of third parties that customers process using cloud services.

This section describes approaches that are currently feasible. In the next section we will describe some approaches that look promising, but are not yet fully developed.

3.1 Use the Cloud Only for Non-private Data

A straightforward (and widely-used) approach to privacy issues with cloud computing is to avoid the issue altogether by only using the cloud for data that is not personal data. This approach could be extended (although with increased risk) to using the cloud also for personal data that is not sensitive. This is a pretty good solution, although it has the drawbacks that the benefits of cloud computing are not obtained for private data, and also that it is necessary to separate private from non-private data, which may be some contexts be difficult, time-consuming or costly. Much of the data handled by individuals and organizations is in fact not subject to data handling regulations, confidentiality, or policy issues restricting its use. This data can be processed in the cloud without giving rise to privacy issues. Moreover, some applications (for example, applications for testing applications and provisioning networks, which can run on artificial data) do not ever need to use personal or confidential data, and so can be run in the cloud without giving rise to privacy concerns.

3.2 Leave Privacy Protection to the Cloud Service Provider

Another approach is that the CSP is trusted to protect personal information on behalf of the customer. Audit, certification and assurance could be provided as a basis for this trust, in an analogous manner to that which exists for account data: for example, SAS-70 type II certification could be used as a basis for this procedure. However, the type of certification needed is not really in existence to date, although ENISA has provided a framework for what would be needed [19].

Using this approach, security need not necessarily suffer in moving to the cloud model. IT security has economies of scale, and a fundamental part of a CSP's business is keeping their systems secure. So a small or medium-sized business whose main expertise is not in the area of security may well have weaker security than a large CSP. If the business changes from processing data in their own network to processing it in the CSP's network, they may thereby increase their security. However, it is very important to select service providers with suitable controls in place, and if this is not done then there is a strong risk to both privacy and security. A drawback to this approach is that customers' data privacy may not be a high priority for the CSP. Standard business models for the provision of Software as a Service involve gaining revenue from the repurposing of customer data, and cloud computing Terms of Service typically offer no compensation if a customer's data is stolen or misused [20].

3.3 Virtual Private Cloud

A virtual private cloud is a reserved space in a public cloud that is isolated from other customers. It may be a physical machine in the cloud reserved for a particular customer, or may use firewalls and encryption for isolation. It protects against attacks by

co-located customers. Amazon Virtual Private Cloud is one example, and Google App Engine supports similar functionality via Google Secure Data Connector, which allows programs developed by an organization (based upon analysis of the data structures involved) to access information behind its firewall. The main drawbacks are that a virtual private cloud does not in itself protect against attacks other than those carried out by co-located customers, and that it is not always clear what the architectural details of a Virtual Private Cloud are, which can make it more difficult for potential customers of Virtual Private Clouds to assessment of the privacy risks.

3.4 Private Cloud

As considered in section 1, a private cloud is a cloud infrastructure operated solely for an organisation, and can be thought of as automated virtualization within a corporate firewall. More privacy is indeed provided than for public cloud, and for some data and applications it is necessary to use private rather than public clouds. On the other hand, private clouds are less flexible and scalable than public clouds. Using a private cloud means forgoing the economies of scale enabled by sharing the use of very large data centres in the public cloud, and also forgoing the economies of scale for security mentioned above. Furthermore, maintaining, patching and upgrading servers remain the problem of the customer, and are not transferred to the service provider.

3.5 Hybrid Cloud

Hybrid cloud, one of the cloud service delivery models considered in section 1, allows an approach to privacy issues in which a private cloud is used for certain tasks, e.g. storing backup information and processing sensitive information, while a public cloud is used for other less privacy-sensitive tasks.

Hybrid cloud customers can choose which part of the hybrid cloud is most suitable for a particular application or workload. For example, they can use a private cloud for mission critical applications, a virtual private cloud for high availability applications, and a public cloud for test and development or productivity applications. If hybrid interoperability is provided, the customer can move the workload from one service delivery model to the next if desired or if requirements change. The drawbacks of this approach are that customers need to determine the choice of the appropriate service delivery model for each application/workload, and that the weaknesses of the individual models are still present: using a private cloud as part of a hybrid cloud does not eliminate the drawbacks of using a private cloud.

3.6 Encryption for Cloud Storage

While it is in transit from a customer to a cloud service provider, data can be protected by encryption. In fact, it is surprising that not all CSPs encrypt data in transit as the default option. Most applications cannot process encrypted data (although this may change in the future, as will be discussed in section 4.2), and so if the data is processed after arrival in the cloud, it usually has to be decrypted first. However, one

common use for cloud computing is simple storage, in which the customer's data is not processed at all. A feasible, and strongly advisable, approach for protecting personal data that is stored but not processed in the cloud is for the customer to encrypt the data before it is sent to the cloud using a key that is not revealed to the service provider. Hence, the data is not just encrypted in transit, but also in storage. Generally speaking, encryption is practical for information in transit and storage, and techniques include SSL/TSL and disk encryption. Data-at-rest used by a cloud-based application is generally not encrypted, because encryption would prevent indexing or searching that data. In fact, 'searchable encryption' technology [21] does enable search over such encrypted data in the cloud. The drawback of end-to-end encryption of data as an approach to privacy protection is that at present, cloud applications doing more than simple storage and search cannot process encrypted data.

3.7 Just-in-Time Decryption

The Just-in-time Decryption approach is that data is stored in the cloud in an encrypted form and, in order to process or query it, the customer's key is passed to the cloud just in time, and re-encrypted immediately afterwards [22]. The drawback is that there is a time window when the data is vulnerable, and indeed the data might be cached, stored in clear or otherwise accessed as a result.

4 Future Approaches

In this section we briefly describe some approaches to privacy in cloud computing that look promising, but are relatively immature or not widely available at present. In subsection 4.8 we discuss what more is needed before they can be more widely successful.

4.1 Trusted Computing

If trusted computing infrastructure is accessible within the cloud, it can be used to protect private data. Trusted computing uses tamper-resistant hardware (called Trusted Platform Modules – or TPMs) that store identities [23], and can be used to generate privacy-friendly attestation identities, protect stored information and bind information to machine-readable privacy policies and to a trusted platform state [24, 25]. Personal data can be encrypted twice, so that the outer layer can only be decrypted by the TPM, and the inner layer can only be decrypted by software that has been checked by an organization trusted by the customer to meet a privacy policy that the customer has specified. The outer encryption serves to ensure that the data is only processed on trusted hardware, which will not subvert the protection offered by the inner encryption. This is related to the idea of a "seal" function, which can bind access to data to certain properties about the integrity of the platform environment. The inner encryption ensures that the data is only processed by trusted software – or at least by software that has a certificate, signed by an organization that the customer

trusts, that the software conforms to the customer's requirements. In general, trusted computing could be used to ensure that the software processing a customer's data met a wide range of requirements, but for the purpose of privacy protection in the cloud the requirements would be that the software is not a privacy threat, for example that it uses customer data only in ways necessary for the provision of the cloud service.

Trusted computing can also be used to provide a root of trust within a range of systems that protect personal data, and in particular to provide the ability to achieve secure actions on a local system, and to identify and authenticate, generate hashes or watermarks, protect logs through signing and encryption and allow integrity checking of software running in a VM. For example, the cloud infrastructure could be set up as a trusted system where cloud elements are registered by trusted location registries and data is identified and registered as well so that the movement of data can be tracked among elements. This can extend to interconnection of trusted clouds in order to cope with movement of data across clouds, boundaries or jurisdictions.

4.2 Processing Encrypted Data

As mentioned in section 3.6, most existing applications cannot process encrypted data. However, it is possible to design applications to process encrypted data in the cloud in a non-efficient way. Protocols for multi-party computation such as Yao's secure two-party protocol [26] could enable a cloud provider and user to cooperate to calculate a function depending on data known to the user and data known to the cloud provider, without the provider learning the user's data or vice versa. These protocols require several rounds of interaction. In contrast, Gentry's fully homomorphic encryption scheme [27] (which differs in approach from multi-party computation) provides a general way of calculating a function of encrypted data in the cloud, with a single interaction round.

4.3 Obfuscation

The idea of this approach is to automatically obfuscate some of the data before sending it to the cloud for processing, using a key that is not revealed to the service provider, and with the degree of this obfuscation dependent upon the context. The result of the processing is de-obfuscated by the customer to reveal the required answer [28].

Examples of this approach include automatically replacing customer identities with pseudonyms [29], and multiplying SQL database columns by secret factors and then permuting these (which is applicable to 90% of features of SalesForce.com's sales and marketing suite [28, 30]). Although it is not unusual for cloud service customers to obfuscate their data before sending it to the cloud for processing, at present this is typically done by hand or in a semi-automated rather than an automated fashion.

4.4 Sticky Policies

This approach is to attach individual privacy rights, conditions and preferences directly to identity data as it moves around the cloud. So, data handling policies are 'stuck'

to data (either directly or indirectly) as it moves around the cloud and govern its usage. One possibility is that customers allow cloud (service) providers to have access to specific data based on agreed policies and by forcing interactions with interchangeable independent third parties called Trust Authorities [31]. The access to data can be as fine-grained as necessary, based on policy definitions, underlying encryption mechanisms (which are used to stick the policies to the data), and a related key management approach that allows data attributes to be encrypted specifically based on the policy. Access to data is mediated by a Trust Authority that checks for compliance to policies in order to release decryption keys. By these means users can be provided with fine-grained control over access and usage of their data within the cloud, even potentially in public cloud models.

4.5 Policies Travelling with VMs

Service providers now allow customers to create virtual machine (VM) images that can be run in the cloud environments. For example Amazon provides the ability to allow customers to create Amazon Machine Images and execute programs within them in Amazon's environment. These images can be booted up using virtualization technologies provided by Microsoft, VMWare, or Xen. Typically these images contain operating system software (e.g., a Linux or Windows boot image) and programs. Most user data is maintained in separate back-end storage systems.

Due to the mobility of VMs and proliferation of data in the cloud model, encryption needs to be applied in multiple places, and to protect more than just application data. One approach is that privacy and security policies would travel with the VM.

A way of achieving this is that all data being managed would reside within the VM image which is stored in encrypted form. Rather than transporting the data separately, or using separate access management and audit applications, all data and necessary management applications are packaged into a virtual machine image and the entire image is treated as a mobile agent. Because the data moves as part of the image, it can be migrated between service providers as necessary, like a file. The software packaged in the VM image controls usage and movement of the data, and ensures that any associated policies that have been set by the data owner are respected. The software starts as part of the VM when it is booted, performs an integrity check to ensure that none of the software or data on the image has been corrupted or compromised, and offers a single set of communication interfaces to allow access to the data. The software also maintains the keys to decrypting the data as well as managing access policy, authentication, and audit logs. Once the software has been configured by the data owner, the virtual machine is "locked-down" to prevent any access to the data (or the resident software) except through the defined interfaces.

Depending on the level of security desired, either the entire image can be encrypted, or only the data being packaged can be encrypted. In the first case, a specialised boot loader is included in the image to do the integrity check and OS boot. In the second case the OS boots as usual, then the packaged software manages the encrypted data. Encrypting the entire image provides a higher degree of security, since the image would be less susceptible to hacking by examining the image content without running it as a VM. This tradeoff can be made by the data owner.

4.6 Agents That Protect Personal Information

Agents can be used for privacy protection in a number of ways, for example to ask a series of dynamic questions which the user can answer to inform agents about their privacy preferences and to set user policies based on the answers [32]. Ann Cavoukian, the Privacy Commissioner in Ontario, suggests that in order to assure confidence and trust in the privacy of personal information in the cloud, personal devices like cell phones, PDAs, smart cards and other tokens under our physical control should interface with the cloud and act on our behalf, with intelligent software agents within these devices or within the cloud used to automatically and continuously scan, negotiate, do our bidding, and selectively reveal identity information [33].

4.7 Privacy Infomediaries

Another approach is that trusted identity providers would act as privacy infomediaries and carry out audit and policy enforcement. For example, they can strip off identifying information from Internet traffic, protecting the end users' identify by hiding the source computers' identifying information (based on research started by David Chaum [34]). Privacy infomediaries have been used previously to provide anonymity by web proxy (for example, in tools such as Anonymiser [35]); there is also a role for such trusted third parties in auditing and compliance checking when using the sticky policies approach described above [31].

4.8 Evaluation

The approaches described in this section are relatively immature or not widely available at present. Here we describe drawbacks of these future solutions and discuss what more is needed before they can be offered widely.

Drawbacks of using trusted computing in the cloud include that all the organizations which handle the data have to be part of the system, and trusted infrastructure must be present at the points of processing: as a result the choice of service providers will be limited until a wider ecosystem of trusted providers has been built up. At present none of the major CSPs appear to provide trusted computing to customers. Also, the tamper-resistant hardware needs to be optimised to handle a large number of operations, as at present it can act as a bottleneck, slowing down transactions.

Strong enforcement of sticky policies within a cloud infrastructure also requires a trusted infrastructure. In addition, the policy semantics need to be agreed across multiple parties, and there is additional overhead for CSPs (in setting up policies and engaging in additional interactions with user agents, Trust Authorities and other service providers with whom information might be shared).

Both multiparty computation and homomorphic encryption would require cloud providers to re-write their applications, which they may not be willing to do. Although there has been work on improving the efficiency of multiparty computation protocols, when the input data is large they still may require a large amount of computation or storage by the user. This does not sit well with the use of cloud computing

to reduce the amount of computing resources required in-house. Homomorphic encryption is at present is too inefficient to be practical, despite some improvements in overhead (see eg. [36]). If this inefficiency can be overcome, in a few years' time practical and commercial offerings using this approach are likely to become available for selected applications.

The drawbacks of obfuscation are that it gives weaker security than encryption, that different applications will in general require different obfuscation methods, and that the approach is not suitable for all applications. However it would be useful, as a minimum, to automate the obfuscation that is currently done by hand. A range of different techniques are possible that come under the obfuscation category; some cloud obfuscation procedures are starting to be offered within the marketplace (for example, as a token-based obfuscation solution [37] or healthcare services in which patient identifiers are replaced by pseudonyms created using PGP encryption [29]). These are specific solutions and contextualised determination and mapping of obfuscation approaches in the cloud is still work in progress and has not yet been provided.

Policies travelling with VMs look promising, especially as by encapsulating the data in a secure virtual machine image, data could be easily migrated between different administrative entities, stay in compliance with regulatory or enterprise policies, and enable secure audits and access management for the data. However, there is a management overhead. This type of solution has not yet been developed or offered.

The main barrier to privacy-protection agents is probably not technical, but the lack of an incentive for service providers to cooperate. Some previous attempts to standardise and automate online privacy negotiations have foundered for this reason.

Finally, privacy infomediaries are not yet available for cloud environments, but there are not any obvious technical barriers to their existence in these environments, and it seems likely that cloud-specialist privacy infomediaries will arise in the future, probably providing additional functionalities in combination with this role in order to improve business viability.

5 Related Issues

In the previous two sections we have presented a number of different approaches to protecting personal information in the cloud. Each of these has its drawbacks in some situations, and none should be regarded as a panacea. The most suitable solution is context-dependent, although if personal information is never revealed (in clear) to a service provider then most of the privacy issues just do not arise.

In this section we consider three issues closely related to the above options, namely key management, design for privacy, and accountability. Progress in these fields affects the options available for protecting personal information in the cloud.

5.1 Key Management

Contextual considerations will determine whether encryption is an appropriate solution for privacy protection in the cloud and if so, which type of encryption should be

used – for example, in some countries encryption is heavily regulated. Customers (both enterprise and individuals) using encryption within the cloud can face key management issues. There is a distinction between using encryption for which the CSP provides the encryption and key management, and end-to-end encryption for which the service provider does not have the key.

The case where the CSP provides the encryption is common in (public) cloud services, and enables the cloud provider to ensure that customer data is encrypted while it is static in a cloud storage area, while it moves between the cloud and the customer, and while it moves between different parts of the cloud. It makes business sense for public CSPs to have encryption facilities and do the required key management. Virtual machines and complex relationships between multiple service providers in the cloud make key management considerably more complicated, but that is the service providers' problem rather than the customers' problem.

For the second case, where the service provider does not have the key, a full Public Key Infrastructure (PKI) is usually too much for a small business, let alone an individual, to handle and the key management requirements may be prohibitive [38]. A specialist privacy/security service provider may offer encryption and key management for a customer using services provided by other cloud providers. Trend Micro [39] and Porticor [40] offer this kind of service, for example. (Porticor uses split keys, so part of the customer's key is private.) A further complication is that one customer might want to communicate with others via an untrusted cloud, which would mean that the communicating parties would need to agree on the encryption to use and how to do key management [41].

5.2 Design for Privacy

The philosophy and approach of embedding privacy into design specifications, as first espoused by Ann Cavoukian and others [42-43], is sometimes referred to as Privacy by Design. This applies to products, services and business processes. The main elements are:

1. Recognition that privacy interests and concerns must be addressed
2. Application of basic principles expressing universal spheres of privacy protection
3. Early mitigation of privacy concerns when developing information technologies and systems, across the entire information life cycle
4. Need for qualified privacy leadership and/or professional input; and
5. Adoption and integration of privacy-enhancing technologies (PETs).

PETs have been defined (from a UK point of view) as "… any technology that exists to protect or enhance an individual's privacy, including facilitating individuals' access to their rights under the Data Protection Act 1998" [44]. So, PETs include privacy management tools that enable inspection of CSP's policies about handling of personal data, provision of user-centric choice, control, transparency, etc., audit and accountability, as well as pseudonymisation tools that could provide confidentiality at the network layer of cloud infrastructure, and anonymisation or pseudonymisation for

cloud applications such as web browsing, email, payment, voting, etc. and technologies centred around privacy-enhanced identity management. A review of different types of PETs is given at [45].

Current privacy concepts [6] are applicable to cloud computing scenarios and can mitigate cloud privacy risks, but it is necessary to implement mechanisms within the cloud to underpin these concepts. Initial frameworks for designing cloud services for privacy have been provided in [46-47], but these are still being developed.

5.3 Accountability

Accountability in relation to privacy is the acceptance of responsibility for protection of personal information. Hence, in order to be an accountable organisation, a privacy management program (consisting of appropriate policies and procedures that promote good practices) needs to be rolled out within that institution. Guidance on how to do this has recently been provided by the Privacy Commissioners of Canada, Alberta and British Columbia [48]. Privacy by Design may complement, and indeed incorporate, corporate accountability mechanisms [49]. The authors are currently engaged in research aimed at making cloud services accountable [50].

6 Conclusion

We have presented a number of different solutions to protecting personal data in cloud computing. No single solution solves the whole problem, and new approaches are being developed, including by the authors. The most appropriate solution depends upon the types of the application and data involved. Indeed, some data (for example, very sensitive personal data, or some types of financial data for which there are especially stringent processing requirements) probably should not go into the public cloud. Conversely, there is plenty of data for which privacy risks are nonexistent or very low. The numerous privacy issues that we have described in connection with cloud computing are not a problem for this data. Companies and individuals who are (rightly) concerned with cloud privacy issues should not let their concerns prevent them from benefiting from the use of cloud computing to handle data for which these issues do not arise.

References

1. Warren, S., Brandeis, L.: The Right to Privacy. Harvard Law Review 4, 193 (1890)
2. Westin, A.: Privacy and Freedom. Atheneum, New York (1967)
3. American Institute of Certified Public Accountants (AICPA) and CICA, Generally Accepted Privacy Principles (August 2009), http://www.aicpa.org/interestareas/informationtechnology/resources/privacy/generallyacceptedprivacyprinciples/downloadabledocuments/gapp_prac_%200909.pdf

4. Solove, D.J.: A Taxonomy of Privacy. University of Pennyslavania Law Review 154(3), 477 (2006), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=667622
5. European Commission (EC): Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (1995)
6. Organization for Economic Co-operation and Development (OECD): Guidelines for the Protection of Personal Data and Transborder Data Flows (1980), http://www.oecd.org/document/18/0,3746,en_2649_34223_1815186_1_1_1_1,00.html
7. Safe Harbor website, http://export.gov/safeharbor/
8. The White House: Consumer Data Privacy in a Networked World: A Framework for Protecting Privacy and Promoting Innovation in the Global Digital Economy (February 2012), http://www.whitehouse.gov/sites/default/files/privacy-final.pdf
9. European Commission (EC): Proposal for a Directive of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data by competent authorities for the purposes of prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and the free movement of such data (January 2012), http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_10_en.pdf
10. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. Byers, A.H.: Big Data: The next frontier for innovation, competition and productivity, McKinsey Global Insitute Report (May 2011), http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation
11. Mell, P., Grance, T.: A NIST definition of cloud computing. National Institute of Standards and Technology. NIST Special Publication 800-145 (2009), http://www.nist.gov/itl/cloud/upload/cloud-def-v15.pdf
12. Narayanan, A., Shmatikov, V.: Robust Deanonymization of Large Sparse Datasets. In: IEEE Symposium on Security and Privacy (S&P), pp. 111–125. IEEE (2008)
13. Lyon, C., Retzer, K.: Privacy in the Cloud: A Legal Framework for Moving Personal Data to the Cloud. Corporate Counselor (February 14, 2011)
14. Gellman, R.: Privacy in the Clouds: Risks to Privacy and Confidentiality from Cloud Computing. World Privacy Forum (2009), http://www.worldprivacyforum.org/pdf/WPF_Cloud_Privacy_Report.pdf
15. Grance, T., Jansen, W.: Guidelines on Security and Privacy in Public Cloud Computing. NIST Special Publication 800-144 (December 2011)
16. Catteddu, D., Hogben, G. (eds.): Cloud Computing: Benefits, Risks and Recommendations for Information Security. ENISA Report (2009), http://www.enisa.europa.eu/act/rm/files/deliverables/cloud-computing-risk-assessment
17. Cloud Security Alliance (CSA): Security Guidance for Critical Areas of Focus in Cloud Computing. v2.1, English language version (December 2009), http://cloudsecurityalliance.org/guidance/csaguide.v2.1.pdf
18. Pearson, S.: Privacy, Security and Trust in Cloud Computing. In: Pearson, S., Yee, G. (eds.) Privacy and Security for Cloud Computing, Computer Communications and Networks. Springer, London (2012)
19. ENISA, Cloud Computing Information Assurance Framework, http://www.enisa.europa.eu/act/rm/files/deliverables/cloud-computing-information-assurance-framework

20. Mowbray, M.: The Fog over the Grimpen Mire: Cloud Computing and the Law. Scripted Journal of Law, Technology and Society 6(1) (April 2009)
21. Kamara, S., Lauter, K.: Cryptographic Cloud Storage. In: Sion, R., Curtmola, R., Dietrich, S., Kiayias, A., Miret, J.M., Sako, K., Sebé, F. (eds.) FC 2010 Workshops. LNCS, vol. 6054, pp. 136–149. Springer, Heidelberg (2010)
22. Cusack, M.: Information Preservation: Structured Data Archiving: Key Issues. Cloud Camp London (2009), http://www.slideshare.net/cpurrington/mark-cusack-cloud-camp4-london-2
23. Trusted Computing Group, http://www.trustedcomputinggroup.org
24. Pearson, S.: Trusted Computing: Strengths, Weaknesses and Further Opportunities for Enhancing Privacy. In: Herrmann, P., Issarny, V., Shiu, S. (eds.) iTrust 2005. LNCS, vol. 3477, pp. 305–320. Springer, Heidelberg (2005)
25. Pearson, S., Casassa Mont, M., Novoa, M.: Securing Information Transfer within Distributed Computing Environments. IEEE Security & Privacy Magazine 6(1), 34–42 (2008)
26. Yao, A.C.: How to Generate and Exchange Secrets. In: 27th Symposium of Foundations of Computer Science (FoCS), pp. 162–167. IEEE Press, New York (1986)
27. Gentry, C.: Fully Homomorphic Encryption Using Ideal Lattices. In: 41st ACM Symposium on Theory of Computing, Bethesda, Maryland, USA, May 31-June 2, pp. 169–178 (2009)
28. Mowbray, M., Pearson, S., Shen, Y.: Enhancing Privacy in Cloud Computing via Policy-based Obfuscation. J. Supercomputing 61(2), 267–291 (2012)
29. Amazon Web Services LLC, TC3 Health (2009), http://aws.amazon.com/solutions/case-studies/tc3-health/
30. Salesforce.com, Inc.: Sales Force Automation, http://www.salesforce.com/products/sales-force-automation/
31. Pearson, S., Casassa Mont, M., Chen, L., Reed, A.: End-to-End Policy-Based Encryption and Management of Data in the Cloud. In: Proc. CloudCom 2011. IEEE (2011)
32. Irwin, K., Yu, T.: Determining user privacy preferences by asking the right questions: an automated approach. In: WPES 2005: Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society, pp. 47–50. ACM, New York (2005)
33. Cavoukian, A.: Privacy in the Clouds. Identity Journal Ltd. (2008)
34. Chaum, D.: Security without Identification: Card Computers to make Big Brother Obsolete. Communications of the ACM 28(10), 1030–1044 (1985)
35. Anonymizer, http://www.anonymizer.com
36. Gentry, C., Halevi, S., Smart, N.P.: Fully Homomorphic Encryption with Polylog Overhead. In: Pointcheval, D., Johansson, T. (eds.) EUROCRYPT 2012. LNCS, vol. 7237, pp. 465–482. Springer, Heidelberg (2012), http://eprint.iacr.org/2011/566.pdf
37. PerspecSys, http://www.perspecsys.com
38. Pate, S., Tambay, T.: Securing the Cloud – Using Encryption and Key Management to Solve Today's Security Challenges, Storage Networking Industry Association (SNIA) (2011), https://www.eiseverywhere.com/file_uploads/974dc3f1fc021f4f6caa02b20a11b031_Pate_Monday_0940_SNWS11.pdf
39. Trend Micro, http://www.trendmicro.co.uk/
40. Porticor, http://www.porticor.com
41. Barker, E., Smid, M., Branstad, D., Chockhani, S.: A Framework for Designing Cryptographic Key Management Systems, NIST Special Publication 800-130 (April 2012), http://csrc.nist.gov/publications/drafts/800-130/second-draft_sp-800-130_april-2012.pdf

42. Cavoukian, A.: Privacy by Design: The 7 Foundational Principles (January 2011) (revised), http://www.ipc.on.ca/images/resources/7foundationalprinciples.pdf
43. Information Commissioners Office, Privacy by Design, Report (2008), http://www.ico.gov.uk
44. Information Commissioner's Office (ICO): Data protection guidance note: Privacy enhancing technologies (2007), http://www.ico.gov.uk/upload/documents/library/data_protection/detailed_specialist_guides/privacy_enhancing_technologies_v2.pdf
45. Shen, Y, Pearson, S.: Privacy-enhancing Technologies: A Review. HP Labs Technical Report, HPL-2011-113 (2011), http://www.hpl.hp.com/techreports/2011/HPL-2011-113.html
46. Pearson, S.: Taking Account of Privacy when Designing Cloud Computing Services. In: Proc. ICSE-Cloud 2009. IEEE, Vancouver (2009), Also available as HP Labs Technical Report, HPL-2009-54, http://www.hpl.hp.com/techreports/2009/HPL-2009-54.html
47. NEC Company Ltd. and Information and Privacy Commissioner, Ontorio, Canada: Modelling cloud computing architecture without compromising privacy: A privacy by design approach (June 2010)
48. Office of the Information and Privacy Commissioner of Alberta, Office of the Privacy Commissioner of Canada, Office of the Information and Privacy Commissioner for British Colombia: Getting Accountability Right with a Privacy Management Program (April 2012)
49. Cavoukian, A., Taylor, S., Abrams, M.: Privacy by Design: Essential for Organizational Accountability and Strong Business Practices. Identity in the Information Society 3(2), 405–413 (2010)
50. Pearson, S.: Toward Accountability in the Cloud. IEEE Internet Computing 15(4), 64–69 (2011)