A Privacy Impact Assessment Tool for Cloud Computing

David Tancock[1], Siani Pearson[1] and Andrew Charlesworth[2]

[1]HP Labs, Long Down Avenue, Bristol, UK. BS34 8QZ
[2] Centre for IT and Law, University of Bristol, Queens Road, Bristol, UK. BS8 1RJ
{David.Tancock, Siani.Pearson} @ hp.com; a.j.charlesworth@bris.ac.uk

Abstract

In this paper, we present a Privacy Impact Assessment (PIA) decision support tool that can be integrated within a cloud computing environment. Privacy is an important consideration in cloud computing, as actual or perceived privacy weaknesses will impact legal compliance, data security, and user trust. A PIA is a systematic process for evaluating the possible future effects that a particular activity or proposal may have on an individual's privacy. It focuses on understanding the system, initiative or scheme, identifying and mitigating adverse privacy impacts and informing decision makers who must decide whether the project should proceed and in what form. A PIA, as a proactive business process, is thus properly distinguished from reactive processes, such as privacy issue analysis, privacy audits and privacy law compliance checking [1], applied to existing systems to ensure their continuing conformity with internal rules and external requirements.

Key Words: cloud computing; cloud storage; decision support; privacy impact assessment.

1. Introduction

A Privacy Impact Assessment (PIA) is a systematic process for identifying and addressing privacy issues in an information system that considers the future consequences for privacy of a current or proposed action [2]. It is thus, in part, a predictive exercise designed to prevent or minimise adverse privacy outcomes. Typically, PIAs are a series of steps, posing and answering questions and considering options. In some jurisdictions an expected deliverable of the PIA process is a document, such as a PIA report [3]. A PIA is different from business processes, such as privacy issue analysis, privacy audits, or privacy law compliance checking [4]; these processes are intrinsically reactive in that they scrutinize existing projects to ensure their continuing conformity with internal rules and external requirements. A PIA is proactive, permitting organisations to design privacy into new systems during the design and development stages, reducing the risk that costly retrofitting of privacy safeguards will be required after implementation. While a PIA may be perceived primarily as a management tool (i.e. as a threat/risk assessment process), it can be leveraged into a tool for enhancing individual privacy. By surfacing privacy issues at an early stage, and providing system designers with relevant knowledge, and the impetus to tackle those issues at the architectural level, PIAs can facilitate the raising of a system's privacy baseline without undue impact on its functionality [5]. Privacy rights are protected and advanced by convincing agencies and businesses to carry out a PIA for the following reasons: to demonstrate legal compliance, to allow organisations to develop better policies, to save money, to develop a culture of privacy protection, to prevent adverse publicity, and to mitigate risks in advance of resource allocation. In the case of cloud computing, the goal of enhancing end user trust by decreasing the risk of exposure of end user's information is particularly important.

However, at present (August 2010) there is no agreed international standard for a PIA process, and there are critical variations in the implementation of PIAs in the different jurisdictions where they are currently in use. Not the least of these variations is the status of PIAs within regulatory frameworks. Some jurisdictions place a formal legislative obligation on organisations to undertake PIAs (e.g. the Ontario Personal Health Information Protection

978-0-7695-4302-4/10 $26.00 © 2010 IEEE
DOI 10.1109/CloudCom.2010.27

Act 2004) [6], some PIAs are prescribed by binding 'policy' (the Canadian Treasury Board Secretariat Directive on Privacy Impact Assessment 2010) [7], or 'recommended' by those with no legal authority (e.g. the United Kingdom (UK) Information Commissioner's PIA Handbooks 2007/09) [8]. As such, the international PIA landscape can be very complex for organisations to navigate.

This paper considers the possibility of developing a PIA decision support tool in a cloud environment. The structure of the paper is organised as follows. In section 2 we consider privacy and security issues in cloud computing. In section 3 we present our solution, which is in the form of a PIA for cloud computing. Section 4 covers related work previously carried out within the context of privacy and security in cloud computing, and evaluates whether elements of these approaches are suitable for our tool. The paper ends with discussion of next steps and conclusions.

2. Privacy and Security Issues in Cloud Computing

In this section, we provide a brief overview of privacy and security issues in cloud computing. These issues will need to be taken into account by our tool, both in its design and in its analysis process.

To find a definitive definition for cloud computing is hard, as there are many variations described for it. Broadly speaking, cloud computing is Internet-based computing, whereby a large pool of easily usable and accessible virtualised resources (such as hardware, development platforms and/or services) can be dynamically reconfigured to adjust to a variable load (scale), allowing also for an optimum resource utilisation. This pool of resources is typically exploited by a pay-per-use model whereby guarantees are offered by the Infrastructure provider by means of customised Service Level Agreements (SLAs). SLAs are usually a part of a service contract between the provider and customer, where the level of service is formally defined.

Typically, in cloud computing the details (resources) are abstracted from the users, who no longer have need for expertise in, or control over, the technology infrastructure 'in the cloud' that supports them. This usually involves over-the-Internet provision of dynamically scalable, and often virtualised, resources. Such provision often takes the form of web-based tools or applications that users can access and use through a web browser as if they were programs installed locally on a user's own computer [9].

In order to reach the cloud, data moves over public networks where it needs to be protected. Cloud providers, such as Amazon, Google and others, achieve economies of scale by building large data centres and then sharing resources among all of their customers. Because the cloud is a shared environment, providers need to ensure that customers are identified properly and that no customer has the ability to see or change another customer's data. However, by incorporating these shared services at the infrastructure layer all clouds should automatically become multi-tenant (depending on the degree of multi-tenancy offered by the cloud), starting at the hardware layer and going all the way up to the user-interface layer.

There are a number of concerns involving privacy and security within cloud computing. For example, data in the form of Personally Identifiable Information (PII) is typically transferred to, stored and processed in the cloud on machines that consumers do not own or control. This type of scenario emphasises the inherent risks and threats associated with cloud computing. Users lose a degree of control over their 'sensitive information' (this term having a specific meaning in Data Protection circles) as the responsibility for protecting that information is shared between the cloud client and the cloud host. The ultimate responsibility for privacy protection, however, always remains with the controller of the private information -- in this case the cloud client. Privacy, security and trust concerns in cloud computing include: the theft, misuse or unauthorised resale of personal data, loss of organisational trust by consumers and decrease of privacy rights, obligations and status [9].

Furthermore, the cloud may have hidden obstacles such as conflicting privacy laws in the jurisdiction where the processing takes place and the jurisdiction where the data originates [10].The variation of privacy laws between jurisdictions, and the dynamic nature of cloud computing will be an ongoing source of controversy [11]. For further analysis of privacy and security issues in cloud computing, see [12, 13, 14, 15].

3. Our Approach

In this section, we present details of a PIA tool for cloud environments, outlining: what the tool does; how the tool works; its architecture. However, at present (September 2010) the tool is only at the conceptual/design stage.

The PIA tool addresses the complexity of privacy compliance requirements for organisations (both public and private sector), by highlighting privacy risks and compliance issues for individuals within the organisation who are not experts in privacy and security, so they can identify solutions in a given situation. This will allow organisations to identify potential issues at an early stage, and hence avoid costs associated with pursuing development paths that are unlawful or pose a higher risk than an organisation can accept or insure against. Where PIAs are mandatory for public sector organisations, the tool can provide evidence that due process has been followed for the purpose of reporting and audit. More specifically, it can help decision makers within organisations to decide whether a new project (in a broad sense, encompassing scheme, notion or product, etc.), that they wish to develop should go ahead, and if so in what form (i.e. what restrictions there are, what additional checks should be made, etc.) The tool could be run at several stages during the lifetime of a project development process, each time producing different output and advice appropriate to that stage.

The PIA tool can be deployed in the cloud as a service accessible from a web browser, using a Software as a Service (SaaS) model, in which external organisations can ask to use that service (probably on a pay-per-use basis), in order to generate PIA reports based on their input, as required. In this model, security mechanisms are used in order to protect any confidential information that is transferred or stored by the service.

The PIA tool is able to reason about privacy and security risks in the cloud that may be raised, as part of its analysis about the project. This analysis includes those aspects mentioned in the previous section, in relation to the particular context involved: who the cloud service provider is, what their trust rating is, what security and privacy mechanisms they use, as well as other factors that are not specifically cloud-related: for example, to what extent the current project involves sensitive information, for what purposes personal data will be used, etc.

Generally, there are two different types (i.e. levels, forms, gradations) of PIAs conducted in all jurisdictions, although the names and the processes vary. For example, names attributed to a short form of PIA are "small-scale" (e.g. UK), "PPIA" (e.g. Canada) and "Privacy Scan'" or "Privacy Impact Statement" in other jurisdictions. The short form of PIA is similar to a full-scale PIA, but is less formalised and requires less exhaustive information gathering and analysis, usually focusing on specific aspects of a project. A full-scale PIA conducts a more in-depth internal assessment of privacy risks and liabilities. It analyses privacy risks, consults widely with stakeholders on privacy concerns and brings forward solutions to accept, mitigate or avoid such concerns. The process guidelines for a full-scale PIA tend to be more comprehensive [3]. Our PIA tool can assist with carrying out either of these types of PIA. It is able to generate assessment, although the aspects relating to stakeholder consultation must be dealt with in a parallel process that can be integrated within a workflow.

User input for the PIA tool contains project information, such as project name, organisation name, region, brief project description, project lead and contact details. This is followed by a descriptive analysis of the project such as outlining project documents, identifying stakeholders, identifying early privacy risks, in order to determine if a PIA is required. For example, the user may wish to describe how the organisation collects or obtains personal information, or explain if personal information will be transferred outside their jurisdiction including details of the receiving countries. Output for the PIA tool is a report displaying information in several sections: introduction; project summary and contact details; the summary of findings (which indicates if the PIA tool has found the project to be either compliant or not); risk summary (which indicates the levels of risk associated with each privacy domain); details of other compliance/non-compliance issues, such as security, transparency, and transborder data flows. Furthermore, the PIA tool provides detailed information about policies in relation to which the project is not compliant or is only partial compliant. In these situations the tool provides detailed reasons for the partial or non-compliance, by highlighting the specific legislation concerned, risks, standards, policies, etc. Finally, a checklist is displayed indicating what the user (organisation) must do to resolve these issues. Throughout the report large visual indicators are displayed; these indicate the issues that have passed compliance, require further attention, or have failed.

The following section provides more details of how the tool works.

3.1. Architecture and Knowledge Representation

The PIA tool is a decision support system (DSS) based on a type of expert system [16]. The architecture of the PIA

tool is illustrated in Figure 1, which represents the main components of the tool and storage services provided by the cloud provider.

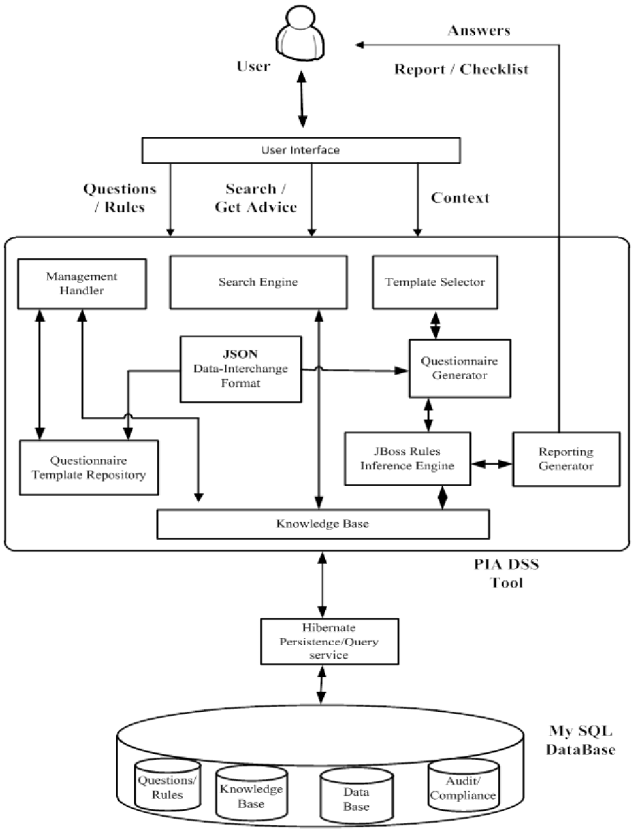

Figure 1. PIA Tool Architecture

The PIA tool has a knowledge base (KB) that is created and updated by privacy experts on an ongoing basis. Thus generic rules for privacy and data protection legislation from a number of jurisdictions (e.g. the United Kingdom (UK) Data Protection Act 1998, the United States (US) Privacy Act 1974 etc.) are created and entered into the KB by the experts using a specific User Interface (UI). This is important as the tool is to be deployed within a cloud environment, whereby organisations from different jurisdictions may ask to use the application. Initially, the tool will cover jurisdictions that currently conduct PIAs including; the UK, the US, Australia, New Zealand (NZ), and Canada.

There are two types of users: end users (who fill in a questionnaire from which a PIA report is generated), and domain experts (who create and maintain the KB). Typically, the end-user interacts with the PIA tool via a Web UI, and answers a series of questions that are contextually generated. For example, the answers provided by the end-user help to build the project profile and the assessment questionnaire, as illustrated in Figures 2 and 3.

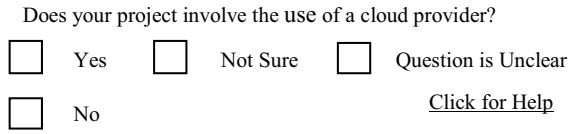

Figure 2. A Simple Question in Project Profile related to a Cloud Provider

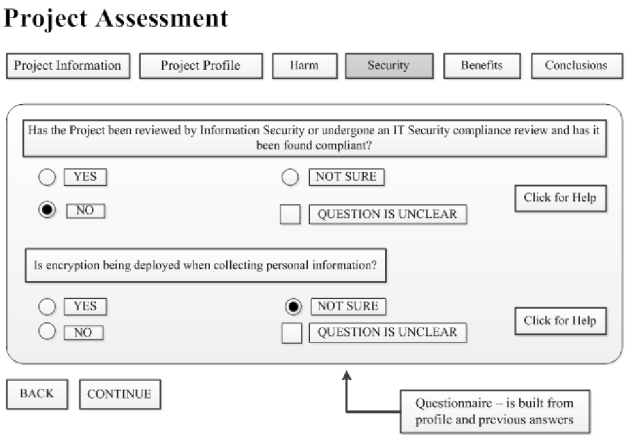

Figure 3. Typical Project Assessment Questions

Questions and answers (QA) in the tool are generated by templates. Different templates correspond to different contexts, and relevant parts of the template can be revealed according to previous user answers. Template-based QA extends the pattern matching approach of natural language interfaces to databases, where the intelligence of the system is embodied in a collection of manually created question templates [17]. This allows the question templates to share the same feature – they are labels, attached to pieces of answer information that mimic the structure of natural language queries and map this structure into the underlying data model. The corresponding output to the user can be static text, data returned by an SQL query or a multimedia presentation.

Our approach uses a JBoss rules (i.e. Drools) engine [18], that makes inferences by deciding which rules (i.e. those created by the domain experts) are satisfied by facts or objects, prioritises the satisfied rules, and executes the rule with the highest priority. The engine uses forward chaining to search the inference rules until it finds one in which the ***'When'*** condition is known to be true. It

concludes the ***'Action'*** condition and adds this information to its data, and continues this until a goal is reached. A meta-level description of the privacy rules for this phase is: 'When <trigger conditions> Then <action>

Although we use this particular inference engine to run rules, the approach is not reliant on any particular inference engine or specific format beyond the processing of ***'if..then'*** rules, so a variety of mechanisms could be used from production rules systems, to 'Clips' or 'Prolog'.

Basically, the tool uses rules to generate an output report and potentially also an audit trail. The output report is ultimately based on the answers provided by end users. The output of the questionnaire (being the answers provided by the user) is matched against the 'when' condition of the rules: the corresponding action within the rules contains code that assesses associated risk and groups output into categories (transborder data flows, compliance with legislation, etc.). This analysis is displayed within the output report as a histogram (bar chart) as illustrated in Figure 4.

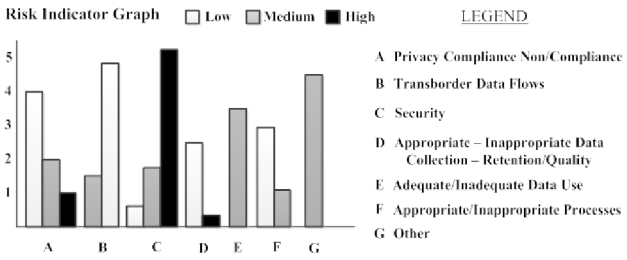

Figure 4. Output Histogram

This indicates the levels of risk associated with each category and colours are used to convey potential privacy risks and threats: black equates to being high risk, grey equates to a warning, and white to low risk associated with that category. In addition to this high-level indication of risk, the output report of the PIA tool displays: more detailed analysis of these risks and how they may be mitigated, report status, a project summary, contact information and online privacy law and PIA guidelines.

Part of the analysis carried out by the tool is to consider legal aspects, such as the UK-US Safe Harbor process for US companies to comply with the European Directive 95/46/EC on the protection of personal data [19]. The tool has to take into account the rules associated with transborder data flows and cross border PIAs [20, 21]; moreover, the tool has to consider global organisations and their binding corporate rules. To achieve this, the tool will have a representation of policies related to different legal jurisdictions, and will take these policies into account as they apply to a given context.

In summary, the PIA tool helps organisations to ensure privacy concerns are met and supports enterprise accountability, supplying employees with sufficient information and guidance to ensure that they design and conduct their projects in compliance with privacy requirements, such as those outlined in the UK PIA Guidelines of 2009 [9]. However, due to the complexities of privacy law and the challenges that organisations face when attempting to achieve multi-jurisdictional compliance in cross-border data processing environments [22], an easier solution to achieve is a PIA tool that merely flags up potential compliance issues rather than a more sophisticated solution that also identifies ,what the user (organisation) must do to resolve these issues.

The following section will provide more details of our specialised tool, including how it may be used in a cloud environment.

3.2. Cloud Deployment

In this section we provide more details of the architecture of the tool, including deployment within the cloud environment.

The tool is deployed in a cloud as a service (e.g. SaaS), whereby the end-users (i.e. customers) of the tool do not actually have to own the platforms. However, there are a number of ways in which our PIA DSS tool may be deployed in the cloud: for example, as a private cloud, public cloud, or hybrid cloud.

It appears that the most appropriate solution for the PIA tool is that it makes use of both private and public cloud resources with external cloud services. This is either on a continuous basis or in form of a 'cloudburst', and is known as the hybrid cloud. This allows the tool to store sensitive data on its own data centre, while making use of public and private cloud services for infrastructure and general computing [23].

To merit discussion and future debate we believe the hybrid model can be improved. The solution is to deploy the PIA tool in a private cloud to protect PII relating to multiple parties, and create a virtualization of our tool, or part of the tool in a public cloud, - that will experience the heavy traffic on a cloud-service infrastructure service as illustrated in Figure 5. Thus the copy of the PIA tool would be a static instance running across a number of

Web servers that could be configured as either a load-balanced set of servers or as a cluster. This would create some latency, but would greatly reduce the effort needed to maintain the servers and the PIA tool, and would eliminate the problem of maintaining state between the internally and externally hosted Web sites.

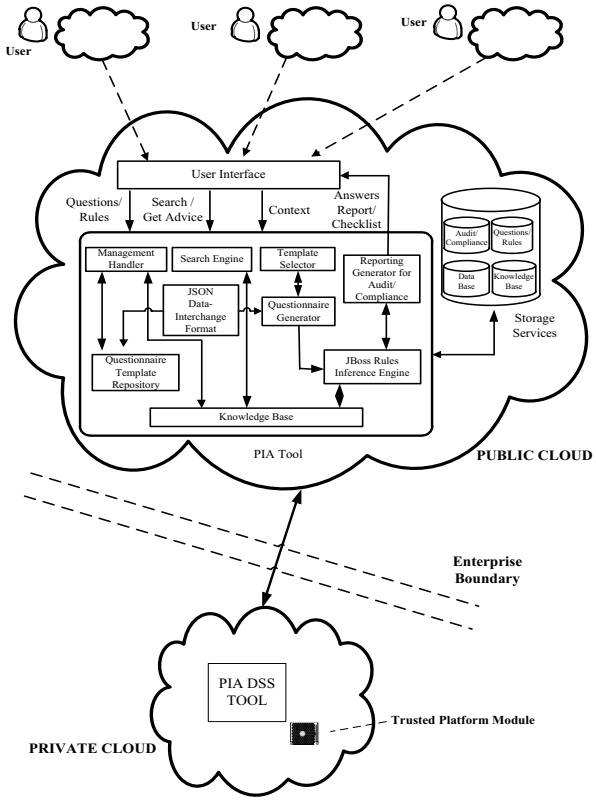

Figure 5. PIA Tool Architecture with Cloud Solution

In order to achieve this solution a number of alternative potential aspects need to be considered. Basically, these aspects are solutions to security and the protection of data (i.e. encryption or obfuscation), and storage.

Therefore, we consider some aspects related to these solutions including: the Trusted Platform Module (TPM), the use of a form of location register that can be deployed in conjunction with elements that govern data storage, the use of a cloud storage gateway.

The distinguishing feature of trusted platforms is the incorporation of cheap 'roots of trust' into computer platforms (in the form of TPMs). They offer facilities for the secure generation of cryptographic keys and limitation of their use, in addition to a hardware pseudo-random number generator [24].

For our PIA tool, the TPM may be virtualised and used to authenticate hardware devices and check that they are operating correctly. Moreover, a TPM can serve as a root of trust for measurement, storage and reporting, and this approach is potentially extensible to support a measure of audit within the cloud.

Another possible approach that may be useful for our PIA tool is a location register [25]. At present this technology is used mainly in mobile (cellular) networks. It is a database of user information that temporarily holds profiles of roaming users (users outside their home area) such as account information, account status, user preferences, features subscribed to by the user, and the user's current location. This approach could be adapted to fit our PIA tool, particularly to detect the user's location. Therefore, rules for transborder data flows could be applied to specific instances. For example, if a US organisation wishes to use the PIA service, then Safe Harbor rules would be included. However, this technique would mainly be useful in an extension of our approach that can provide enhanced proof that the user information used by the PIA tool is correct - in other words, to check the location of the information that is stored and transferred associated with cloud services that are the target of assessment by the PIA tool.

A potential solution to full virtualisation of our tool into the cloud is the cloud storage gateway. A cloud storage gateway can provide encryption, authentication, and authorisation, but it is a server that resides at the customer premises, and exposes cloud storage services as if they were local storage devices [26]. The gateway is typically packaged as a virtual machine (VM) and translates cloud storage Application Programming Interfaces (APIs), including Representational State Transfer (REST) or Simple Object Access Protocol (SOAP), to block-based storage protocols such as Internet Small Computer System Interface (iSCSI) or Fibre Channel. Additionally, the cloud storage gateway uses local caching to alleviate latency issues, and can translate file-based interfaces such as the Network File System (NFS) or Common Internet File System (CIFS) with seamless integration. This is largely due to the fact that cloud storage gateways use standard network protocols, and can translate traditional file based protocols to cached object-oriented storage. An advantage of using this approach is that the administrator (i.e. expert) can modify or update the rules and templates of the PIA tool very easily and quickly without corrupting the application files that are copied by the cloud provider.

This is achieved by the gateway as it allows the main PIA tool files to be kept safely within the private cloud, while secondary copies of the tool's files (i.e. the working set that is used by the tool's customer) can be cached back for fast-access if needed [26]. The working set of files in this instance are pages of the PIA tool that the customers of the PIA tool are expected to access (e.g. a Welcome Page, questionnaires etc).

Another advantage of using a cloud storage gateway for our PIA tool is the ability to update, at regular intervals, application files that are stored in the cloud. For example Nasuni [26] terms this a 'Synchronous Snapshot'. Thus after the initial push (where all files are copied to the public cloud and moved into the cache), the snapshot checks each file chunk for changes within the file tree. It then tags new files and altered, corrupted, old, chunks of data as dirty. New files are chunked, and all of the dirty data is then compressed and encrypted. The snapshot then sends each encrypted chunk to the specified cloud and receives the associated keys that allow it to retrieve files in the event of a restore or a cache miss. Once both files and directories have been pushed to the cloud, the snapshot generates a new root directory and tears down the snapshot, ready to start all over again. Therefore, the snapshot uses a number of protection techniques including: the duplication, compression, and encryption of each file, before sending them to the cloud, as illustrated in Figure 6. However, the snapshot only forwards changes between the original files and the most recent version and pushes out only what is necessary, thus reducing potential storage costs.

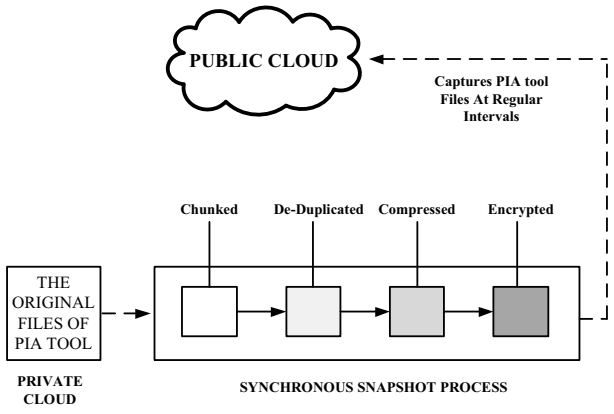

Figure 6. Synchronous Snapshot

Moreover, many cloud storage gateways facilitate the use of encryption techniques and frameworks (e.g. RSA, OpenPGP etc.), whereby the gateway has no access to customer data, as all encryption and decryption happens at the user site. Overall, the analysis of cloud storage gateways indicates that this approach may be suitable for our PIA tool.

4. Related Work

We believe that our tool is a novel approach as at present (i.e. August 2010) no such PIA application exists. Our approach offers several advantages over currently available solutions. For example, compared to manual or spreadsheet based solutions such as Microsoft's Excel and Apple's Numbers, our DSS tool has intelligence integrated into the solution so that it can efficiently tailor questionnaires about potential privacy risks and privacy law compliance checking to customer needs. More importantly, compared to decision tree-based solutions our approach is significant enough to express common privacy knowledge using natural language for cloud environments, while restricted enough to have provable determinism and termination. This is because decision tree-based solutions are not powerful enough to model common privacy knowledge, as they have difficulty in dealing simultaneously with multiple dependencies relevant for privacy decisions. We achieve this by using a rule-based system in a controlled way.

In the remainder of this section, we discuss related work in the areas of privacy and security in cloud computing, to evaluate whether these approaches are suitable to aid enhancement of our PIA tool.

Accountability as a way forward for privacy protection in the cloud is considered by Pearson and Charlesworth [10]. They propose the incorporation of complementary regulatory, procedural and technical provisions that demonstrate accountability into a flexible operational framework to address privacy issues within a cloud computing scenario. They believe that accountability is a useful basis for enhancing privacy in many cloud computing scenarios, as corporate management can quickly comprehend its links with the recognised concept of, and mechanisms for achieving, corporate responsibility. Accountability in this context is corporate data governance (i.e. the management of the availability, usability, integrity and security of the data used, stored, or processed within an organisation), and it refers to the

process by which a particular goal – the prevention of disproportionate (in the circumstances) harm to the subjects of PII – can be obtained via a combination of public law (legislation, regulation), private law (contract), self-regulation and the use of privacy technologies (system architectures, access controls, machine readable policies). The approach taken requires a combination of procedural and technical measures to be used and co-designed. In essence, this would use measures to link organisational obligations to machine readable policies, and mechanisms to ensure that these policies are adhered to by the parties that use, store or share that data, irrespective of the jurisdiction in which the information is processed. Companies providing cloud computing services would give a suitable level of contractual assurances, to the organisation that wishes to be accountable, that they are can meet the policies (i.e. obligations) that it has set, particularly PII protection requirements. Furthermore, technology can provide a stronger level of evidence of compliance, and audit capabilities.

However, while the approach appears to be a practical way forward, it has limitations. For example, while contracts provide a solution for an initial service provider to enforce its policies along the chain, risks that cannot be addressed contractually will remain, as data has to be unencrypted at the point of processing, creating a security risk and vulnerability due to the cloud's attractiveness to cybercriminals. Moreover, only large corporate users are likely to have the legal resources to replace generic SLAs with customised contracts.

Obfuscation, as a first line of defence is described by Pearson, Shen, and Mowbray [27]. This paper describes a tool called 'privacy manager', which they believe reduces the risk to the cloud computing user of their private data being stolen or misused, and also assists the cloud computing provider to conform to privacy law. The idea is that instead of being present unencrypted in the cloud, the user's private data is sent to the cloud in an encrypted form, and the processing is done on the encrypted data. The output of the processing is de-obfuscated by the privacy manager to reveal the correct result. The obfuscation method uses a key which is chosen by the user and known by the privacy manager, but is not communicated to the service provider. Thus the service provider is not able to de-obfuscate the user's data, and the un-obfuscated data is never present on the service provider's machines.

Although, some obfuscation methods are highly susceptible to known plaintext attacks [27], this does at least protect the data from opportunistic data thieves with access to cloud databases, because it ensures that the data is never present in the database in the clear.

Use of DSSs for cloud computing and PIAs is a very new field and there are few systems available. Those that are available for cloud computing are found in the areas of clinical decision applications [28], and life science enterprise solutions [29]. However, very recently there has been a step change in DSS for PIAs (such as privacy expert systems). Typically, a DSS has a KB that needs to be created and updated periodically by experts on an ongoing basis and a mechanism (for example a rules engine, decision tree, or dedicated queries to databases) by which output can be generated, based upon user input via questionnaires. Within this context, we will discuss briefly two DSSs that are at the cutting edge of research.

PRAIS [30] is a research project that has developed a prototype DSS tool for context-sensitive privacy-aware information sharing in children's social care . The DSS is based on the architecture developed for the Identity Governance Framework (IGF) [31], where information sharing is based on a pull model. This means that the recipients are alerted that information is being made available to them, after which it is retrieved from the source. PRAIS uses–the IGF architecture as its design choice because it allows the owner of the information to retain liability for the data and to audit each use by using the pull model. However, the scope of PRAIS is very narrow as it is not intended that the DSS will ever make decisions on behalf of properly trained personnel but instead will assist social care practitioners in making privacy-aware decisions where required.

Hewlett Packard's Privacy Advisor (HPPA) is an expert system that captures data about business processes to determine their privacy compliance [32, 33]. The tool helps organisations to ensure privacy concerns are met and supports enterprise accountability, supplying employees with sufficient information and guidance to ensure that they design and conduct their projects in compliance with organisational privacy policies. HPPA uses a rules engine for which rules are defined that are used both to generate questions that are customised to the employee's specific situation and also to codify HP's privacy rulebook and other information sources. Based on the employee's response to these questions, it automatically generates an output report that includes

analysis of possible privacy risks and a checklist of actions that the employee should take in order to mitigate these risks. This tool is currently being rolled out to employees within HP. Analysis, of this tool suggests that the methods and techniques used in HPPA are well suited for the PIA tool, although, it will be necessary to modify the approach of HPPA to fit our solution.

Sander and Pearson [34] outline a DSS for cloud computing that aids selection of appropriate cloud service providers (CSPs). Their approach is a semi-automated DSS tool that gathers context relating to CSPs and inputs to a rule-based system, to trigger decisions about whether or not to use that CSP and/or additional stipulations that would need to be made. The tool helps to determine appropriate actions that should be allowed and assesses risk before personal information is passed on through the cloud. For each customer enterprise, an administrator will set up the original questionnaire according to the policies that the customer (i.e. the enterprise) wishes to check, or uses the default setting offered by the assessment service. When a customer wishes to assess different CSPs offering a service, providers will use the tool via a Web interface in order to provide answers to the questionnaire, and the results will be sent back to the enterprise that wishes to choose between the service providers. These results include reports and automatically generated ratings, which allow the administrator to distinguish between them. This tool is similar to the HPPA tool, in that it is a form of expert system using a set of intermediate variables (IMs) to encode meaningful information and to drive the questionnaire generation.

Although there are some similarities between this tool, HPPA and our PIA tool (notably, use of the Drools inference engine), there are significant differences in architecture and deployment, the underlying mechanism for the knowledge representation and for generating questionnaires, and the rules, report structure and output.

The next section will discuss next steps associated with the development of our PIA tool.

5. Next Steps

Because the tool is only at the conceptual/design stage, our next planned steps include to:

1. Analyse how stakeholder analysis and workflow can be integrated into the tool, and whether there are any aspects of PIAs that cannot be captured by such an approach
2. Conduct empirical research to obtain the initial set of rules for the KB.
3. Consider different Artificial Intelligence (AI) methods for the display (i.e. the reports and the grading of privacy risks etc.)
4. Choose a cloud storage gateway provider for our tool. This will be measured by the services they provide and the costs that they charge for this service.
5. Develop the code using Java (i.e. Java Server Pages (JSP), JavaBeans etc.) technologies. This will involve employing a modular approach from the design phase, and includes building the KB.

6. Conclusions

We are currently developing a PIA tool that can be used in a cloud environment to identify potential privacy risks and compliance. We believe that this generic approach will prove of increasing benefit as cloud service adoption increases.

7. References

[1] B. Stewart. 1996. "Privacy Impact Assessments." *Privacy Law & Policy Reporter.* 3 (7), pp. 61-64. http://www.austii.edu/au/journals/PLPR/39.html.

[2] C. Bennett. R. Bayley. A. Charlesworth. R. Clarke. 2007. "Privacy Impact Assessments: International study of their application and effects." http://www.ico.gov.uk/upload/documents/library/corporate/research_and_reports/privacy_impact_assessment_international_study.011007.pdf#13.

[3] D. Tancock. S. Pearson. A. Charlesworth. "Analysis of Privacy Impact Assessments within Major Jurisdictions." *Proceedings Privacy Security and Trust (PST) 2010, Ottawa, Canada.*

[4] D. Tancock. S. Pearson. A. Charlesworth. 2010. "The Emergence of Privacy Impact Assessments." http://www.hpl.hp.com/techreports/2010/HPL-2010-63.pdf.

[5] A. Cavoukian. 2009. "Privacy by Design: The 7 Foundational Principles." http://www.ipc.on.ca/images/Resources/7foundationalprinciples.pdf.

[6] ServiceOntario. 2010. "Personal Health Information Protection Act 2004." http://www.elaws.gov.on.ca/html/statues/english/elaws_statues_04p03_e.htm.

[7] Canadian Treasury Board Secretariat. 2010. "Directive on Privacy Impact Assessment 2010." http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=18308&section=text#cha1.

[8] Information Commissioners Office. 2009. "Privacy Impact Assessment Handbook." http://www.ico.gov.uk/handbook/June.2009.

[9] E. Knorr. G. Gruman. 2008. "What cloud computing really means." *InfoWorld.* http://www.infoworld.com/d/cloud-computing/what-cloud- computing-really-means-031.

[10] S. Pearson. A. Charlesworth. "Accountability as a Way Forward for Privacy Protection in the Cloud." *Proceedings.1st CloudCom 2009,ed.M. G. Jaatun, G. Zhao, C. Rong, Beijing, Springer LNCS 5931,* pp. 131-144.

[11] S. Pearson. " Taking account of privacy when designing cloud computing services." *Proceedings. ICSE Workshop on Software engineering Challenges of Cloud Computing,* Vancouver, Canada, May 2009, pp. 44-52.

[12] European Network and Information Security Agency (enisa). 2009. "Cloud Computing: Benefits, risks and recommendations for information security." http://www.enisa.europa.eu/act/rm/files/deliverables/cloud-computing-risk-assessment.

[13] Cloud Security Alliance. 2009. "Security Guidance for Critical Areas of Focus in Cloud Computing V2.1." http://www.cloudsecurityalliance.org/guidance/csaguide.pdf.

[14] R. Gellman. 2009. "Privacy in the Clouds: Risks to Privacy and Confidentiality from Cloud Computing." http://worldprivacyforum.org/pdf/WPF_Cloud_Privacy_Report.pdf.

[15] K. Zeng. A. Cavoukian. 2010. "Modelling Cloud Computing Architecture without Compromising Privacy: A Privacy by Design Approach." http://www.ipc.on.ca/images/Resources/pbd-NEC-cloud.pdf.

[16] J. Giarratano. 2005. *"Expert Systems: Principles and Programming."* Canada. Thomson Learning.

[17] A. Andrenucci. E. Sneiders. "Automated question answering: review of the main approaches." *Proceedings, 3rd IEEE International Conference on Information Technology and Applications (ICITA), 2005,* pp. 514-519. volume 1.

[18] JBoss Drools. 2010. http://jboss.org/drools.

[19] D. Solove. 2008. *"Understanding Privacy."* Harvard, Harvard University Press, United States.

[20] OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data. 2010. http://www.oecd.org/document/18/0,2340,en_2649_34255_1815186_1_1_1_1,00.html.

[21] T. Karol. 2009. "A Guide to Cross-Border Privacy Impact Assessments." http://www.isaca.org/Knowledge-Center/Research/ResearchDeliverables/Pages/A-Guide-To-Cross-Border-Privacy-Impact-Assessments.aspx.

[22] R. Susskind. 1989. "The latent damage system: a jurisprudential analysis." *2nd International Conference on Artificial Intelligence and Law. Vancouver, British Columbia, Canada,* pp. 23-32. http://portal.acm.org/citation.cfm?doid=74014,74018.

[23] J. Rhoton. 2009. *"Cloud Computing Explained."* 2nd edition. United States. Recursive Press.

[24] Trusted Computing Group. 2010. "TCG Architecture Overview, Version 1.4." http://www.trustedcomputinggroup.org/resources/tcg_architecture_overview_version_14.

[25] Hewlett Packard. 2010. "Home Location Register (HLR)." http://h20208.www2.hp.com/opencall/products/mobility/ochlr/index.jsp.

[26] Nasuni. 2010. http://www.nasuni.com/.

[27] S. Pearson. Y. Shen. M. Mowbray. "A Privacy Manager for Cloud Computing." *Proceedings, 1st CloudCom 2009, ed. M. G. Jaatun. G. Zhao. C. Rong. Beijing, Springer LNCS 5931,* pp. 90-106.

[28] C. Preimesberger. 2010. "IBM, Aetna Join for New Cloud-Based Health Care Support System." http://www.eweek.com/c/a/Health-Care-IT/IBM-Aetna-Join-for-New-CloudBased-Health-Care-Support-System-667092/.

[29] CambridgeSoft. 2010. "ChemBioOffice Cloud - An Integrated Decision Support System for CHDI." http://chembionews.cambridgesoft.com/WhitePapers/Default.aspx?whitePaperID=43.

[30] R. Harbird. M. Ahmed. A. Finkelstein. E. McKinney. A. Burroughs. 2007. "Privacy Impact Assessment with PRAIS." http://www.cs.ucl.ac.uk/staff/A. Finkelstein/papers/hotpets.pdf.

[31] Liberty Alliance Project. 2007. “ID governance – identify privacy and access policy, marketing requirements document.” http://www.projectliberty.org/.

[32] S. Pearson, T. Sander and R. Sharma. “Privacy Management for Global Organisations.” *Data Privacy Management and Autonomous Spontaneous Security, J. Garcia et al (Eds),LNCS 5939,*pp. 9-17, Springer-Verlag Berlin Heidelburg, 2010. http://www.springerlink.com/content/a13142g81255p453/fulltext.pdf?page=1,March 2010 .

[33] S. Pearson. P. Rao. T. Sander. A. Parry. A. Paull. S. Patruni. V. Dandamudi-Ratnakar. P. Sharma. “Scalable, Accountable Privacy Management for Large Organizations." INSPEC 2009: 2nd International Workshop on Security and Privacy Distributed Computing, Enterprise Distributed Object Conference Workshops (EDOCW 2009), IEEE, pp. 168-175.

[34] T. Sander. S. Pearson. "Decision Support for Selection of Cloud Service Providers." To appear in International Journal on Computing (JoC), GTSF, 2010.