A System for Privacy-Aware Resource Allocation and Data Processing in Dynamic Environments♦

Siani Pearson, Marco Casassa Mont
Trusted Systems Laboratory
HP Laboratories Bristol
HPL-2006-185
March 28, 2007*

privacy, privacy policies, trusted computing

In this paper we describe a system for allocating computational resources to distributed applications and services (within distributed data centres and utility computing systems) in order to perform operations on personal or confidential data in a way that is compliant with associated privacy policies. Relevant privacy policies are selected on the fly, based on related meta-policies, depending on contextual information (potentially including location) and properties of the resources. One or more Trusted Privacy Services are involved to mediate the access to the data, based on the satisfaction of pertinent policies. Resources might be equipped with trusted computing components (e.g. Trusted Platform Modules) to provide higher assurance and trust about the contextual statements or properties of these resources (such as their location, their status and integrity, etc.).

* Internal Accession Date Only
♦ I-NETSEC 06, May 2006, Karlstad University, Karlstad, Sweden

Approved for External Publication

A System for Privacy-Aware Resource Allocation and Data Processing in Dynamic Environments

Siani Pearson and Marco Casassa-Mont

Hewlett-Packard Research Labs, Filton Road, Stoke Gifford, Bristol, BS34 8QZ. UK.
{Siani.Pearson, Marco.Casassa-Mont}@hp.com

Abstract. In this paper we describe a system for allocating computational resources to distributed applications and services (within distributed data centres and utility computing systems) in order to perform operations on personal or confidential data in a way that is compliant with associated privacy policies. Relevant privacy policies are selected on the fly, based on related meta-policies, depending on contextual information (potentially including location) and properties of the resources. One or more Trusted Privacy Services are involved to mediate the access to the data, based on the satisfaction of pertinent policies. Resources might be equipped with trusted computing components (e.g. Trusted Platform Modules [1]) to provide higher assurance and trust about the contextual statements or properties of these resources (such as their location, their status and integrity, etc.).

1 Introduction

Enterprises store large amounts of confidential data about their employees, customers and partners. On the one hand, accessing and managing this data is fundamental for their business: confidential information is retrieved, analysed and exchanged between people (and applications) that have different roles within an organisation (or across organisations) to enable the provision of services and transactions. On the other hand, data protection and privacy laws, including [2,3,4], and data subjects' privacy preferences dictate increasingly strict constraints about how these data have to be protected, accessed and managed. Failure to comply with such privacy laws can have serious consequences for the reputation and brand of organisations and have negative financial impacts. There is therefore a need to reveal sensitive data but this must be done in a way that is legally compliant and consistent with data subjects' expectations.

A special case is where privacy management capabilities process confidential data in environments (such as dynamic and distributed enterprises, GRIDs, etc.) where IT resources are dynamically allocated. These environments can be subject to varying geographical, legal and organisational constraints. Because of the specific location of the resource different privacy policies could apply, and privacy management based on static assumptions is no longer valid.

This paper describes HP Labs' approach to addressing the problem above by providing an adaptive privacy management system in which relevant policies (governing

conditions to be satisfied by data requestors in order to access data) are dynamically determined based on the current context. Our solution consists of mechanisms for:

1. Specifying constraints for the dynamic allocation of resources based on privacy policies. This is achieved via Privacy Policy Packages strongly associated to confidential data. A Privacy Policy Package contains "localised" privacy policies (specific for a given context) along with meta-policies specifying the criteria for selection between the localised privacy policies;
2. Dynamically driving the selection of resources based on checking privacy constraints, specified in the above Policy Package, against the properties of available resources, including their localisation;
3. Enforcing the disclosure of confidential data, given a previously selected resource. This is based on a Trusted Privacy Service checking the relevant privacy constraints (specified in the Policy Package) against local credentials and contextual information;
4. (Optionally) providing trusted localisation of resources based on a Trusted Registration Service coupled with Trusted Localisation Providers leveraging trusted platform technologies.

In this paper we describe the main concepts underpinning our work and current results.

2 Addressed Problem

Dynamic, distributed and adaptive enterprises [5], utility data centers and grid systems allocate on-demand IT resources driven by business and computational needs. Resources could run applications and services that, amongst other things, might need to process personal and confidential data. These resources can be physically located in a variety of environments subject to different legislative and organisational rules and policies. Confidential and personal information might need to be transferred across organisational and geographical boundaries. In cases where this is legally allowable, this information might still be subject to different privacy policies or privacy guidelines depending on where it is processed. For example, data might be transferred between different data centres located in EU countries. Despite the fact that the same EU Data Protection Directive would apply, local privacy policies (dictated by the local government or organisation) or other types of constraint might require the data to be accessed and processed in different ways.

In addition, varying contextual information could influence choices for access control and data protection mechanisms relating to dynamic computational resources including personal and mobile resources such as laptops, mobile phones and PDAs. Employees (especially HR people, managers, doctors, etc.) need to process confidential data as part of their daily jobs, and as ubiquitous computing spreads, the resources used to do this need to be taken into account - different policies, settings and rules might apply if different computers, infrastructures, etc. were used at a given time.

Privacy management based on static assumptions is no longer valid as we move from a static processing model to a dynamic one: confidential data has to be processed adaptively depending on the context and the relevant policies and laws. Failure

to comply with privacy laws can have serious consequences for the reputation and brand of organisations and service providers and have negative financial impacts.

We address this problem and in particular provide a solution to the following key issues:

1. how to ensure that confidential data is processed only on resources (and in contexts) that satisfy privacy policies relevant for these data
2. how to increase assurance about the trustworthiness of properties of computational resources, including their physical location.

3 Our Solution

Our solution consists of a system to address the above problems. This section discusses some relevant scenarios we aim to address, introduces the model underpinning our privacy management solution and describes technical approaches for its implementation.

3.1 Addressed Scenarios

In this section we briefly describe some dynamic enterprise-based scenarios where our solution adds value.

Dynamic allocation of resources within data centres spread across geographic locations. In this scenario resources (e.g. servers) are dynamically allocated to run applications and services to process data, for example, in dynamic and distributed enterprises. Workloads are spread based on the availability of such resources, to optimise their usage. However, there are privacy issues because computational resources that belong to different geographical locations, organisational boundaries and administration domains, etc. can be subject to different privacy policies. So, the "location" of the resources is an essential input to decisions about resource allocation and privacy management.

Mobile employees. Employees can be dynamic, both in the sense of travelling around and using different mobile resources (devices and enterprise tools including laptops, PDAs, mobile phones, etc.) to process different types of confidential or private data used in daily work activities (such as confidential e-mails and documents, medical data, access private databases, etc). It could be desirable to ensure that such sensitive data would only be processed within well defined locations and potentially well defined types of devices (e.g. a certified laptop but not a cellular/smart phone or PDA). This is increasingly the case as ubiquitous computing spreads. Here, privacy policies could describe constraints not only on location but also type of device or resource.

3.2 Overall Model

The model underpinning our solution consists of mechanisms to:

- model and represent a set of "alternative" privacy policies associated with confidential data: one or more of these policies can be selected and enforced depending on the resource's context and location. Meta-policies describe the selection criteria. We refer to these aggregations of policies as the Policy Package;
- strongly associate a Policy Package to confidential data. Confidential data is obfuscated and can only be put 'in clear' if the constraints defined by the policy package are satisfied;
- constrain the dynamic selection of resources based on the content of the Policy Package;
- check and enforce privacy policies based on the current relevant set of privacy policies and trusted "localisation" information. We refer to this as the Trusted Privacy Service;
- provide trusted information about the "locality" of a computational resource. This mechanism involves a Registration Entity and Trusted Localisation Provider.

Figure 1 shows the high-level architecture of a system implementing our solution.

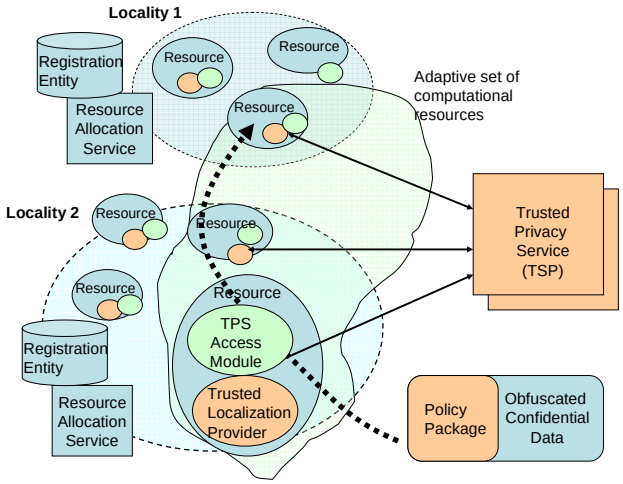

Fig. 1. High level architecture

A Trusted Localisation Provider (TLP), installed on resources, supplies trusted "location" information: it can be based on Trusted Computing Group (TCG)-compliant trusted components and a trusted software layer [6]. At the very least the TLP can be used, for each resource, to provide and retrieve a "trusted certification" of where the resource is. The resource administrator must be trusted and accountable for keeping these location-based certificates up-to-date. In a more complex scenario the TLP can leverage any "localisation techniques" (such as hardened GPS, GPRS triangulation, etc.) to provide trustworthy localisation information.

When confidential data needs to be moved (or a copy transmitted) from one resource to another, it is obfuscated and strictly associated to a Policy Package (by a resource controller or the resources themselves), by using traditional cryptographic techniques (RSA public key cryptography) or alternative cryptographic schemas [7]. The Policy Package dictates which privacy policies need to be enforced based on a variety of contextual information, including location of the resource. This drives the selection of the computational resource that will process a piece of confidential data.

Specifically, we focus on the concept of "resources" as the entities that require access to data instead of people. Resources need to interact with one or more Trusted Privacy Services (TPSs) (via their interaction module) in order to access the content of the obfuscated confidential data. The TPS is a secure Web Service that checks for policy compliance and audit interactions. Resources can be equipped with trusted computing components to provide higher assurance and trust about the contextual statements. The "third party" component, the TPS, mainly interacts with resources to grant or deny them access to data (via disclosing decryption keys) based on their compliance to policies associated to data. Resources' trusted components can be directly involved in this process.

We envisage two alternative mechanisms to dynamically allocate resources based on their "localisation" and the interpretation of Policy Packages:

1. A Registration Entity (RE) may be used during the resource allocation process – within a Resource Allocation Service – to mediate the provision of localisation information. This is a central (domain-based) mechanism for administering the localisation information associated with the resources it manages.
2. The allocation decision is made on-the-fly, by identifying a potential resource and checking if it is compliant with the policies defined in the Policy Package. The resource has its own "localisation information" that is provided by the TLP (either self-generated or injected by the RE).

This basic model can be extended and adapted to a variety of scenarios including enterprise and inter-enterprise contexts. In particular the TPS can be provided by an organisation for internal consumption or by one or more external trusted third parties, to enable multi-party interactions and at the same time increase the overall trust and accountability. For more details about the role of trusted third parties in such systems see [8].

Localisation is just one of the contextual aspects that we need to take into account during the policy verification and enforcement phase. Our approach to defining localisation of resources is mainly based on certificates (signed declarations) issued by resources, by relying on local trusted computing components; of course, other mechanisms could be used in order to provide contextual information within such a system.

Further details follow about the Policy Package, the TPS and the TLP.

3.3 The Policy Package

The Policy Package describes sets of context-related policies along with meta-policies to enable their selection. It is strongly associated to obfuscated confidential data and dictates terms and conditions under which this data can be disclosed. Figure 2 shows the high level elements of a Policy Package:

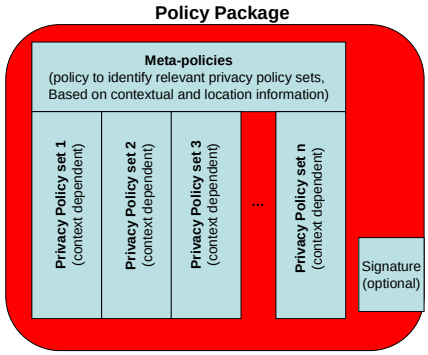

Fig. 2. Policy Package

Meta-policies and privacy policies can be expressed via logical expressions dictating constraints, which can be based on contextual information such as location, trust domain, type pf device, etc.:

- **Meta-policies**. These enable the selection of relevant privacy policy sets (there might be no policy set that satisfy specific conditions or more than one could be active), based on contextual information. For example a meta-policy could "activate" (select) a specific policy set, based on its relevance for a given location.
- **Privacy Policy sets**. These contain privacy policies dictating the conditions, obligations and requirements to be satisfied in order to allow a resource to access the obfuscated data. Constraints can refer to contextual information too.

The above policies can be represented by using standard formats, such as digitally signed XML. Suitable standards for expressing such rules include Extensible Access Control Markup Language (XACML) [9] and the Enterprise Privacy Authorisation Language (EPAL) [10]. The content of the Policy Package has a double function:

- To drive the selection of the computational resource that will process a piece of confidential data: the policy package can discriminate, via meta-policies, which resources can or cannot process its associated confidential data, for example based on the resource location;
- To designate the right set of privacy policies to be satisfied: given the "location" of a resource, the Policy Package can be used to determinate which privacy policies apply. In order to access confidential data, the resource will have to interact with the TPS that will interpret and enforce relevant policies.

3.4 Trusted Privacy Service (TPS)

Figure 3 describes the high-level architecture of the TPS and the resource's TPS access module. In our approach, resources are configured to host a TPS Access Module and a TLP module. The former can be considered as a locally installed "agent" and the second as a trusted computing component. Once resources receive obfuscated confidential data, they need to interact with one or more TPSs via their TPS Access Module in order to access the content of these data. The TPS, shown in figure 3, is a secure and trusted web service that checks for policies' compliance and audit interactions. The TPS Access Module exposes its functionalities via well defined APIs: applications/services running on resources can call these APIs either explicitly or via application plug-ins (for example for e-mail browsers or word processors).

Both the TPS and the Access Module have a policy engine to interpret policies. These engines can be implemented by using traditional rule-based systems. As shown in figure 3, the resource, via its TPS access module, sends the Policy Package (1) to the TPS in order to satisfy the relevant privacy policies and access to the associated confidential data.

The TPS contains a module to interact with the resource's TLP. It gathers trusted contextual information from the resource (2) and processes the relevant set of privacy policies, identified by the execution of the package's meta-policies.

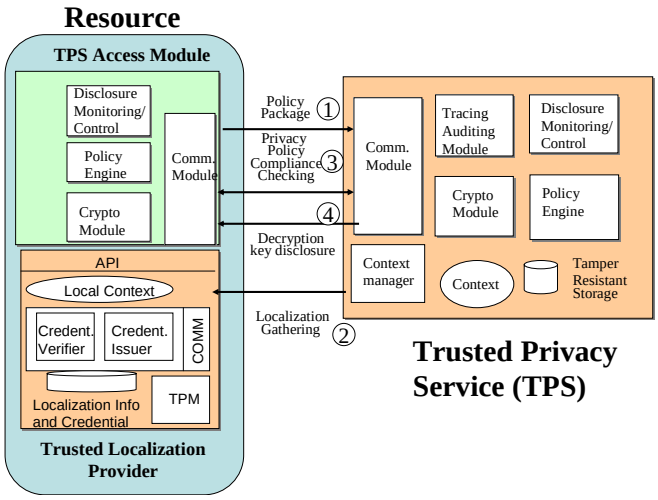

Fig. 3. Architectural detail of Trusted Privacy Service (TPS)

Multiple interactions between the TPS and the resource (3) might be required to check its compliance to the privacy policies (the resource might need to provide additional credentials, etc.). The exchanged information is audited and logged. If the resource satisfied the privacy policies, the TPS uses its cryptographic module to generate the keys to de-obfuscate confidential data (for example based on IBE [7] or traditional cryptography) and sends it to the resource (4).

3.5 Trusted Localisation Provider (TLP)

This subsection provides more details about the TLP and its interaction with the TPS.

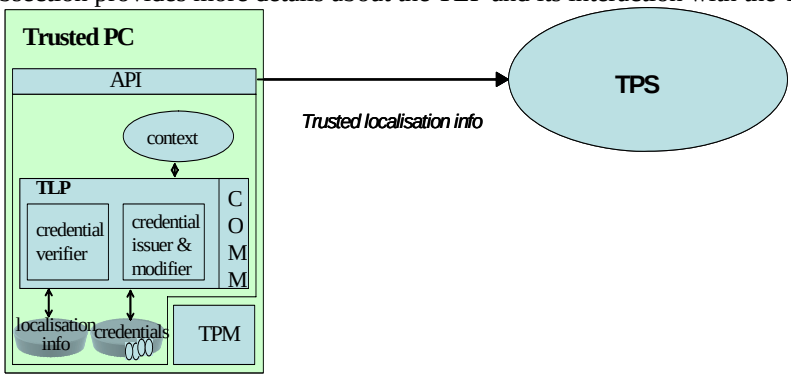

Fig. 4. Trusted Localisation Provider (TLP)

The TLP, located on resources, has two core components as shown in figure 4:

(a) *a trusted "localisation" software layer* within a platform that certifies and/or provides localisation information (e.g. MAC or IP address or system information) about that platform via an API. T

(b) *a trusted component*, such as a TCG-compliant TPM [1], to provide certified and trustworthy information so that a greater degree of trust may be achieved.

A TLP can be built only of the "localisation" software layer, with no trusted platform: in this case the degree of trust in the TLP is lower than the case where a TPM is leveraged. Localisation information may be certified by one or more TLPs; these could potentially form a hierarchy.

As shown in figures 3 and 4, a TPS can directly interact with the local TLP to gather the localisation information.

An alternative mechanism for how the TPS may receive "localisation" information about the platforms (resources) on which processing is to occur or on which sensitive information is to be stored, etc., is based on the Registration Entity (RE). As described previously, the RE is a trusted service (which could potentially be run by a trusted third party, but need not be) for registering machine information and its association with "localisation" information. This provides the benefits of a centralised service and being able to call upon a combination of further external trusted entities or knowledge and its own domain of expertise (or checks for which it takes responsibility) to provide a degree of trust in the information that it certifies that can be measurable, quotable and is commensurate to the type of checks applied (analogous to security checking).

In the case of a TLP leveraging a trusted platform (TP), normal software will operate in conjunction with the trusted hardware root (TPM) [1] within the TP, as follows. Whenever new localisation information is to be created on the client machine, the TLP instructs the TPM to create a new public key pair based on random sources comprising a new public key and a new private key. For security, the private key is

never revealed outside the TPM, and the TLP will request the TPM to form any operations involving it. Depending upon the circumstances, a RE (or other third party with enhanced CA functionality) can:

1. Add an association between the platform ID and localisation information in a database which may be queried by third parties.
2. Create an attribute certificate that certifies that the holder has certain "localisation" attributes. The RE will then need to send the attribute certificate to the TLP.
3. By analogous means to 2. above, use a previously certified identity to create another representation of that identity, possibly with additional attribute values, for use in different circumstances.

The TPM would protect this trusted mechanism; this involves third parties publishing integrity metrics of parts of the trusted mechanism (including the TLP) so that its correct operation could be checked as part of the TP boot integrity checking process, or in response to a challenge.

4 Deployment of Our Solution

We envisage the deployment of our solution in data centres whose IT resources may span across organisational and national boundaries and might be subject to different privacy policies. We assume that each resource will run the TPS Access Module.

To deploy our solution we require that administrators create and store a model of the managed types of data, along with the relevant privacy policies (and meta-policies), This can be done within the Resource Allocation Service. A model of managed applications and services is also required along with the specification of which types of data they will need to access.

The mechanisms provided by our solution can be leveraged directly by the Resource Allocation Service to dynamically allocate IT resources to applications and services. Based on the models mentioned above, the Resource Allocation Service can retrieve the relevant policies and check the suitability of potential IT resources against these policies, via the TPS service.

Personal data is stored in standard data repositories (e.g. relational databases, LDAP servers, etc.), hosted by specific data centres' resources. However, these data can now be stored in an encrypted form, along with the associated privacy policies. Copies of these data repositories can be made on IT resources and data will be protected because of this encryption.

The interaction of applications and services with data repositories still happens via standard protocols (e.g. JDBC, LDAP, etc.). However, we envisage the usage of proxies that are able to intercept attempts to access data and will transparently interact with the TPS service to ensure that privacy policies are enforced. Of course applications and services might need to be modified to be aware that part of the retrieved data is encrypted, along with associated policies. This is particularly true when they need to retrieve data not for local processing but to send it to other applications or

services running on remote resources. Further details of such an approach are provided in [16].

5 Comparison with Related Work

To the best of our knowledge, we are not aware of anything closely related to our approach. Most of the known approaches are about specific privacy management systems deployed within static environments i.e. subject to well defined (and static) privacy policies. This in particular applies for IBM's work on Enterprise Privacy Architecture (EPA), IBM Tivoli Manager [11] and EPAL (privacy language) [10] and SUN's user access and distributed identity techniques. Microsoft have carried out work in the area of context-aware policies [12]: polices are evaluated on the fly against the current context but there are not such concepts as adaptive set of privacy policies, dynamic enforcement, trusted localisation and accountability. IBM appears to have researched in the area of location-based, environmental and contextual controls to access resources [13]: the same comments as above apply.

Relevant work has been done in the area of protecting personal data by strongly associating privacy policies and managing the disclosure of this data based on the fulfillment of these policies. Related technical approaches include cryptographic schemas to protect sensitive data and allow its disclosure based on the fulfillment of associated policies: they are based on traditional public-key cryptography or alternative schemas, such as Identifier based Encryption [7]). Frameworks and services have also been implemented to leverages these cryptographic schemas and provide the required interaction mechanisms for a selective and conditional disclosure of data [8,14].

Part of this work can be leveraged to provide some of the basic functionalities necessary to build our solution.

It is important to notice that current systems for dynamic allocation of resources do not explicitly consider privacy requirements as a driver for the selection of computational resources. *Ad hoc* or specific approaches are deployed but these are not automated and the enforcement of privacy policies does not adapt to changing circumstances. Our approach explicitly addresses this issue by providing privacy enforcement mechanisms that are adaptable to different privacy contexts. Specifically, we provide mechanisms that use privacy policies for selecting suitable resources and dictating terms and conditions to be satisfied in order to access confidential data. The underlying infrastructure based on trusted privacy services provides mechanisms to enforce privacy policies in an accountable way. A Trusted Localisation Provider system provides further assurance about the location of resources by leveraging TCG technology coupled with a registration mechanism.

6 Current Status and Next Steps

We are currently taking steps towards the development of an integrated prototype of our solution. We have already implemented key sub-system modules and components that can underpin the construction of our overall solution: feasibility of the TPS and TPS Access Module components and the Policy Package mechanisms is demonstrated by our exploitation of Identifier-based Encryption (IBE) schemas and related interaction models [7,8,14]. In the same context, we have also demonstrated the feasibility of associating "sticky policies" to confidential data and using it to drive disclosure processes. A simple implementation of the TLP can be provided by leveraging HPL/TSL expertise on Trusted Computing and TPM technology: work in this direction is ongoing in the context of the EU PRIME project [15]. We anticipate that implementation of the RE component should be straightforward.

We still need to fully quantify the impact of our solution (including delay when performing typical operations) on the applications and services that need to use and access confidential data and must operate in accordance with privacy policies. This will be done once a first implementation of our prototype is available. In the meanwhile, we are exploring how to achieve this in a transparent way for applications and services by using proxy-based mechanisms that can preserve native application and service interactions with repositories where personal data is stored. In terms of dealing with a privacy-aware selection of computational resources, our solution can be seen as an "add-on" for enterprise middleware software or GRID software: further work has to be done to integrate it with a real system but we cannot see any major conceptual or technical problems in doing this.

7 Conclusions

This paper describes an innovative approach to deal with selection and allocation of computational resources in distributed and dynamic environments in order to process sensitive data in a privacy-compliant way. The discussed solution is based on privacy localisation provision and privacy management services and allows operations to be performed on personal and confidential data in a way that is compliant to associated dynamic privacy policies. Both allocation of computational processes to specific IT resources and data access are subject to the fulfilment of these policies. In the outlined approach relevant policies are dynamically determined based on the current context. In general, a set of (potentially quite different) policies can be associated to personal data along with meta-policies, which define criteria for selecting the relevant policies based on the context and resource properties. This allows the system to cope with heterogeneous and distributed environments that could be subject to different privacy policies based on their localisation and context.

These techniques allow management of the movement of private or confidential data throughout a dynamic grid of computing resources so that it is only moved to servers that are trusted as to their level of control for that sensitive data. The data is encrypted under control of a tightly bound agent that enforces the applicable privacy

policy and can dynamically qualify computing resources based on that policy and the other elements of the system that let it know which resources can be trusted. This is of value in distributed enterprise software environments in which sensitive data may be computed out in the dynamic virtual grid including trusted and not-so-trusted resources. It would also be of interest in highly secure entities, such as government, which would like to move to virtual utility models so long as they could be convinced that their security policies can be upheld.

Our research and development is work in progress. Part of this research may be carried out within the context of the PRIME project [15], an international project on identity and privacy management funded by the European Union.

References

1. Trusted Computing Group: TCG TPM Specification v1.2. Available via https://www.trustedcomputinggroup.org/home (2005)
2. Laurant, C.: Privacy International - Privacy and Human Rights 2003: an International Survey of Privacy Laws and Developments. Electronic Privacy Information Center (EPIC). Privacy International. http://www.privacyinternational.org/survey/phr2003/ (2003)
3. OECD: OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data http://www1.oecd.org/publications/e-book/9302011E.PDF (2001)
4. Online Privacy Alliance: Guidelines for Online Privacy Policies. Online Privacy Alliance http://www.privacyalliance.org/ (2004)
5. Hewlett-Packard Ltd (HP): Adaptive Enterprise - Overview, Technologies and HP Services http://www.hp.com/products1/promos/adaptive_enterprise/us/adaptive_enterprise.html (2005)
6. Pearson, S. (ed.): Trusted Computing Platforms. Prentice Hall (2002)
7. Cocks, C.: An Identity Based Encryption Scheme based on Quadratic Residues. Communications Electronics Security Group (CESG). UK. http://www.cesg.gov.uk/site/ast/idpkc/media/ciren.pdf (2001)
8. Casassa Mont, M., Pearson, S., Bramhall, P.: Towards Accountable Management of Privacy and Identity Management. Proc. ESORICS (2003)
9. OASIS: eXtensible Access Control Markup Language (XACML). http://www.oasis-open.org (2005)
10. IBM: The Enterprise Privacy Authorisation Language (EPAL). EPAL 1.2 specification http://www.zurich.ibm.com/security/enterprise-privacy/epal/ (2004)
11. IBM Tivoli Privacy Manager: Privacy manager main web page - http://www-306.ibm.com/software/tivoli/products/privacy-mgr-e-bus/
12. Microsoft Corp.: Methods and systems for context-aware policy determination and enforcement, patent no. EP1220510A2
13. IBM Corp: Protecting resources in a distributed computer system, patent no. US6658573B1
14. Casassa Mont, M., Harrison, K., Sadler, M.: The HP Time Vault Service: Exploiting IBE for Timed Release of Confidential Information. WWW2003 (2003)
15. PRIME Project: Privacy and Identity Management for Europe. European RTD Integrated Project under the FP6/IST Programme http://www.prime-project.eu.org/ (2005)
16. Casassa Mont, M, Pearson, S.:An Adaptive Privacy Management System for Data Repositories, Proc, TrustBus 2005 (2005)