Privacy Preserving Trust Agents

Stephen Crane
Trusted Systems Laboratory
HP Laboratories Bristol
HPL-2004-197
November 11, 2004*

A Trust Agent is an assembly of software components arranged to provide trusted remote entities access and control over certain aspects of a user's end system, in a privacy preserving manner. The Trust Agent recognising the end user as the platform owner, and consequently the owner of any personal information held on the platform. In this paper we describe two scenarios (one domestic, the other commercial) where Trust Agents can be deployed. The solution draws on some of the benefits that can be gained through the introduction of a TPM (Trusted Platform Module) to enhance platform security.

* Internal Accession Date Only
Approved for External Publication

Privacy Preserving Trust Agents

1 Abstract

A Trust Agent is an assembly of software components arranged to provide trusted remote entities access and control over certain aspects of a user's end system, in a privacy preserving manner. The Trust Agent recognising the end user as the platform owner, and consequently the owner of any personal information held on the platform.

In this paper we describe two scenarios (one domestic, the other commercial) where Trust Agents can be deployed. The solution draws on some of the benefits that can be gained through the introduction of a TPM (Trusted Platform Module) to enhance platform security.

2 Introduction

This paper provides an insight into the concept of a Privacy Preserving Trust Agent, an evolution of the Personal Trust Assistant (PTA) [Cra03], [CC03]. The previous work on the PTA revealed that computing trust is difficult without predefined metrics against which to measure. This is mainly due to the subjective nature of human trust.

Trust is an essential pre-requisite for on-line interactions. At some point during an interaction both parties form an opinion about how the other will behave. Establishing trust is hard and has traditionally relied on soft interpersonal factors like past performance and recommendation. In today's on-line world the indicators that would normally be used to establish trust are either not present or cannot be relied upon to the same degree as they would in a conventional face-to-face interaction.

Personal information, when shared with others, raises questions about trust. Given a choice, individuals will choose to share their information with only those that they trust. Individuals have greater trust in others when a) they retain control over the information they share, or b) they are provided with strong evidence that the receiving party will respect their privacy wishes.

So, how can an individual be sure that the facts they reveal about themselves to others will not be used inappropriately? Of course, individuals can chose to simply not share the information, but this is likely to disadvantage the individual since they will probably be denied access to the service. Individuals need assurance that they can share information with confidence.

This paper explains one approach to addressing some of these trust and privacy concerns. By providing technology, which we call a Trust Agent, individuals can still retain control of their shared information. Trust Agents allow users to express exactly how they wish their information to be used.

3 The Trust Agent

Consider a simple client-server model in which the client represents the individual and the server represents the organisation requesting the individual's personal information.

The Trust Agent consists of several software components that are provided by the service provider and reside on the client system. The role of the Trust Agent is to provide local management and local trusted functionality, but limit the rights that a remote server has when accessing the client. With the Privacy Preserving Trust Agent, the Trust Agent is extended to include the Personal Privacy Manager. These two components share overall responsibility for access to the individual's personal information. See Figure 1.

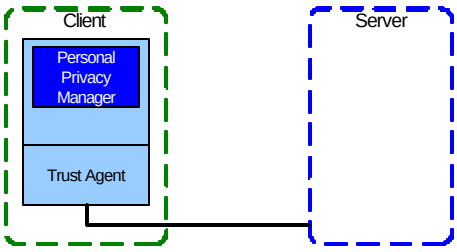

Figure 1

The suggestion of granting a remote party routine access to information stored on a client system raises obvious concerns. However, these concerns can be mitigated if the solution provides sound justification for each party to thrust the other. Within the Trust Agent, the model of trust is simply 1) that each party understands and trusts the functions that the Trust Agent is permitted to perform, and 2) for controlling access to personal information, the individual is provided with functionality that they trust and manage independently.

3.1 Operating modes

The Trust Agent provides several distinct modes of operation. In the most basic mode the trust agent simply observers the status of the client operating environment, e.g. hardware version, operating system version, security patch level, application(s) version, virus scanner status (last run, result), and reports the status back to the control centre so that decisions can be made regarding further actions.

In a more sophisticated model the Trust Agent adopts an autonomous role, still operating under the instruction of the controlling centre and reporting back status, but able to reconfigure the operating environment, disable high risk operations and respond to varying situations.

3.2 Establishing trust

For the solution to work effectively, the user and service provider must each trust the Trust Agent. The user will be concerned about granting a 'foreign' application access to their personal information. For the service provider the situation is slightly different. The concern here is whether the Trust Agent will operate correctly and as intended in the foreign environment.

Trust is established as follows:

- The server will want to check that the Trust Agent is intact and operating correctly. The preferred way to do this is by building on the *root of trust* present in the Trusted Processing Module (TPM) installed in the client. The TPM technology we have in mind is that being pursued as part of the TCPA/TCG initiative [Pea03], [TAO04]. This same integrity reporting process can be used by the client to check the integrity of the shared Trust Agent components and any component that the client alone has access to. Alternative approaches that do not use a TPM are possible but clearly not as robust.
- Protection of the client components is achieved through isolation, in our case using independent virtual machines running on the client platform. The same isolation provides the service provider with assurance that the Trust Agent will operated as expected, although it is recognised that with current platform technology is it difficult to guarantee no interference from the user. It is for this reason that the service provider and the user must be expected to behave honourably, which we believe is a reasonable assumption given that in the scenarios described they both have something to gain from so doing.
- Ultimately, users want to be assured that the information they store on their system can only be accessed with their authority. Therefore, they demand a high level of confidence in the processes used to control access to this information. Trust is assured by allowing the individual to source and manage this process completely independently, which in our examples is achieved using protected storage feature of the TPM.

- Trust in the communications between client and server will be achieved using conventional cryptographic techniques, e.g. SSL, IPSec, Mixes, etc.

Section 5 discusses the architecture of the Trust Agent in detail. One important point to note when reading the section is the modular, or component-based, approach we have adopted. This is intentional, and designed to instil further trust, particularly for the client. In theory each component can be sourced from a different provider, chosen by the client on the basis of trust. In fact, the client can establish multiple Trust Agents from a variety of sources, again dependent on the trust the user has in the supplier.

In the specific case of the Privacy Preserving Trust Agent, the following trust requirement exists:

- Both parties must agree a set of privacy enabling functions that meet their particular needs of the application the service provider is offering. Having done so, each party must be satisfied that the functions have been implemented correctly. Exactly how this is achieved will depend on the trust that each party has in the other. For example, the individual may be willing to accept the server's Trust Agent on face value or may want an independent third party to evaluate it.

4 Scenarios

To help understand the high level architecture presented in this paper, consider the following two scenarios:

4.1 Scenario 1 – The corporate operating scenario

A large multi-national financial institution operates a workforce of mobile users. These users work from home, customers' premises, the car or the coffee shop, using a laptop or PDA. The institution recognises the risks that these platforms present and defines a security standard that describes the state that an appliance must be in before it can connect to the corporate network. The standard goes beyond a simple state description to include dynamic monitoring and recording that must be undertaken for the duration of any connection.

The institution places a Trust Agent (read as 'trusted application') on the user's appliance to monitor and enforce compliance with the standard. The user is comfortable with this arrangement because he recognises (but does not necessarily understand) the implications of poor security and is willing to trust the institution's apparent good intentions.

Each time the user attempts to connect to the institution the Trust Agent checks the status of the appliance and continues to monitor and report back to the institution. Any deviation from the standard results in the connection being blocked and an exception report raised by the Trust Agent, which is passed back to the institution.

4.2 Scenario 2 – The domestic/personal operating scenario

A child has an account with a popular Internet Service Provider (ISP), which is used for email and browsing the Web. The child's parents have heard reports about pornography, child abuse and other 'undesirable' activities that are promoted over the Internet. However, while they understand the significance they have no idea how to deal with the problem and look to their ISP for help.

The ISP installs a Trust Agent on the child's computer. The Trust Agent carries out initial and on-going checks to assure the parents that their concerns are being addressed. The motivation for the ISP to provide this service is a mixture of financial, 'being seen to be doing the right thing' and as a differentiator from other ISPs to help recruit/retain customers.

The Trust Agent checks the state of the client environment and confirms that it is not vulnerable to known attacks. It also monitors connections and their content, blocking anything that appears to be undesirable. Metrics are described in policies that the ISP manages and deploys to the Trust Agents.

4.3 Scenario summary

Both scenarios discuss very different situations, but which have almost identical architectural solutions. One party, recognising the problem, employs a solution to manage and reduce their level of risk/exposure. This solution comes in the form of technology – the Trust Agent – that is developed, managed and owned by the service provider, but which is operated by the client. Understanding the client environment enables far greater control and predictability. The architecture can be simply represented as Figure 2.

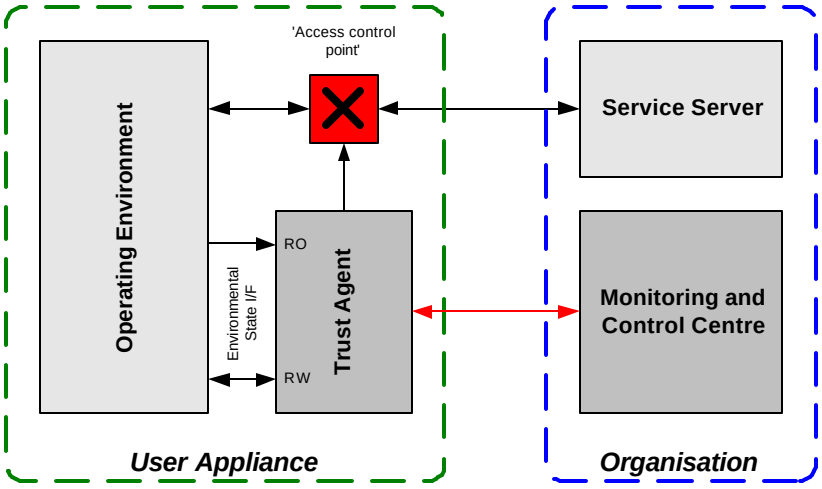

Figure 2

There is a strong analogy between this architecture and the 'Chip and PIN' initiative [CP04] being rolled out by the UK banks. The banks give customers trusted technology (smart cards) that the customers do not really understand but are happy to use because 1) they typically trust their bank and 2) recognise that if the bank is willing to invest in this technology then it must be worthwhile. (Of course, these changes may not represent the best option for customers, since the current limited customer liability model may disappear with the introduction of Chip and PIN!) For the banks, the initiative represents a way to manage exposure through the introduction of trusted technology, effectively allowing the banks to extend their domain of trust.

5 Architecture

The architecture of the Trust Agent is shown in Figure 3. The components are described in more detail in the following sections.

5.1 Components

The architecture consists of the following four components:

5.1.1 Trust Agent Core Component (TACC)

The Core Component provides the trusted functionality specific to the particular Trust Agent role. The functionality includes local policy interpretation (and possibly enforcement), control of access to private information, communications with servers and interaction with the TPM. (Note: This functionality is complementary to any privacy functionality provided by the PPM.)

5.1.2 Trust Agent Local Operating Environment (TALOE)

This component of the Trust Agent provides a local processing environment. It enables instructions received from the server to be executed under the constraint of the TACC (and PPM). The TALOE can operate as a trusted container or as a receptor for remote requests received from a server.

5.1.3 Trusted Platform Module (TPM)

The TPM provides environment and application integrity checking, and therefore performs the role of Trust Agent *root of trust*. (Note: Currently, as specified by TCG, the TPM provides only protected storage and hardware cryptographic functionality, but as the supporting infrastructure develops integrity reporting services will added.)

5.1.4 Personal Privacy Manager (PPM)

The PPM manages all requests for access to personal information. Within the PPM the Policy Manager allows the individual to specify access based on personal preferences. The Security Manager ensures that access from the Trust Agent (or other applications that need access to personal information) is controlled according to the rules set out in the policy and the nature of the agreed privacy functions.

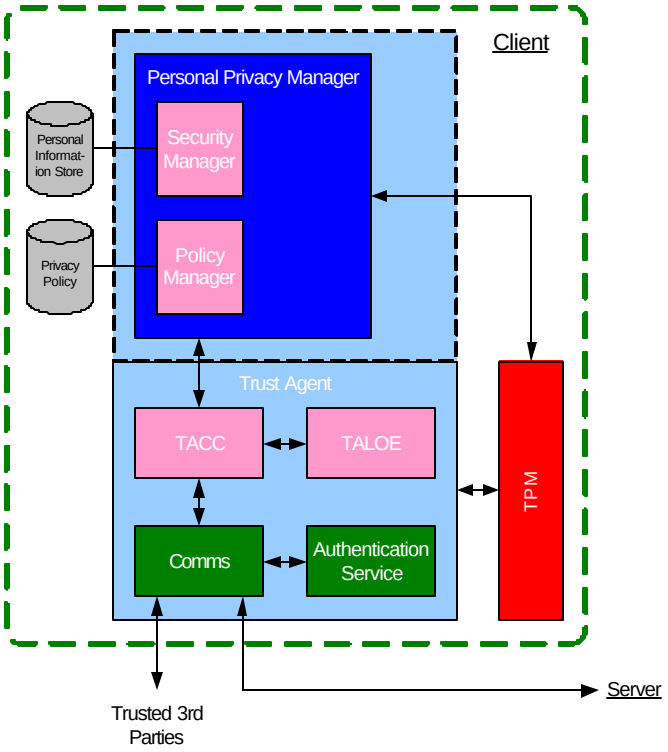

Figure 3

6 A typical implementation

The security requirements governing the service provider's access to the client platform are set out in policies that can be automatically understood by the Trust Agent.

Take Scenario 1: The users want to access the institution's web service using a standard browser. Before granting access the Trust Agent is instructed to carry out a range of checks on the client's environment. It does so by looking for specific characteristics as stated in the policy. The Trust Agent then activates a local proxy that monitors access requests and directs those not authorised by the policy to a holding site (or redirects them to an authorised site). The protocol implemented between the proxy and the host is proprietary and incorporates additional security controls, e.g. encryption. See Figures 4 & 5.

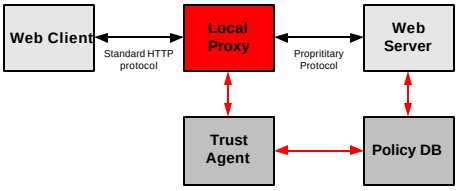

Figure 4

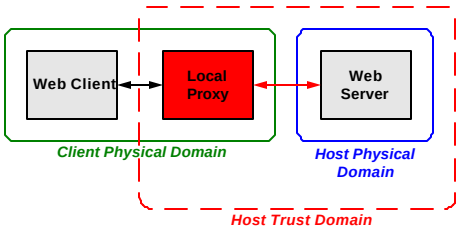

Figure 5

The appropriate policy is sent by the service provider to the Trust Agent over the logically separate secure channel. The result of the analysis performed by the Trust Agent is returned through the same logical channel to the service provider.

7 The Privacy Preserving Trust Agent

7.1 Personal Privacy Manager

As previously explained, the Personal Privacy Manager (PPM), a component of the Trust Agent, manages the controlled release of personal information held on the client platform. The PPM exists as an isolated application running in its own virtual environment under the direct control of the user.

7.2 Principles of privacy preservation

Individuals have four fundamental ways available to them to control the release of their personal information. They can:

1. Refuse to share it
2. Share it unreservedly (the current position for many users)
3. Only permit it to be used under their control (provide functionality/capability but not the underlying data)
4. Share it with conditions attached.

The first two options are perfectly reasonable but could either result in a service being denied or exposing the individual's personal information to potential misuse.

Options 3 and 4 place the individual in control, and it is these two options that the Trust Agent makes possible.

Note that additional privacy mechanisms like anonymity and pseudonymity are not discussed in this paper. Nevertheless, these are privacy controls that users are likely to want to use in certain situations, and for completeness are treated here as belonging to the 'refuse to share' category.

7.3 Operating modes

The Trust Agent PPM operates in one of two modes. It can obfuscate any personal facts that might otherwise be exposed to the service provider. Where this is not possible the Trust Agent initiates a process for enforcing terms and conditions on each item of personal information that the service provider requests. For example, consider the case where a server asks a client for the individual's date of birth. The individual is reluctant to reveal this information and queries why the information is required. Two situations could arise:

- The server explains that it simply wants to build a unique identifier for the individual based on name, address and DoB. This identifier is of the form Hash (Name | Address | DoB). The solution in this case is for the client to build this index locally and pass the result to the server.

This technique for preserving privacy lends itself to a range of scenarios. Functions other than the simple hash function can be used, and a list of suggested privacy preserving techniques can be found in the following section.

- Alternatively, the server may state that it really does need to know the individual's DoB, perhaps because the service offered can only legally be offered to adults. In this case the client associates a policy with the DoB data item that states that (say) it can only be used for this specific service, not retained and not passed on to third parties, and releases the combined policy and data as a single item.

A third possibility is where the client is unwilling to share personal information with the server, but is willing to share with a mutually Trusted Third Party (TTP). In this case the Trust Agent mediates on behalf of the client and server by passing the personal information directly to the TTP.

Note how all situations require a level of understanding and cooperation between the client and the server.

7.4 Privacy preserving functions

Several options exist for functions that would enable the Trust Agent to process a request for information in a way that preserves the privacy. However, in all cases the service provider must accept that the information they receive is unlikely to specifically identity the individual. If this is not sufficient for their needs then the individual must consider sharing the personal information subject to a usage policy.

Assuming that a privacy preserving function is acceptable, likely functions include:

- Cryptographic mechanisms, e.g. hash, zero knowledge
- Factually incorrect but statistically correct revelations
- Partial obscurity, e.g. MSB/LSB (Most Significant-/least Significant Bits) masking
- Processing by a Trusted Third Party
- Anonymisation techniques

8 Privacy Policies

8.1 Setting policy

The individual defines how items of personal information will be handled. Essentially, the individual states whether an item can be shared or not, and if not then whether it can be processed locally. The individual may also want to state who can access information, since in practice they are likely to trust some organisations more than others. For each item of information, and for each requesting organisation, the user categorises the sensitivity of the data. See Figure 6.

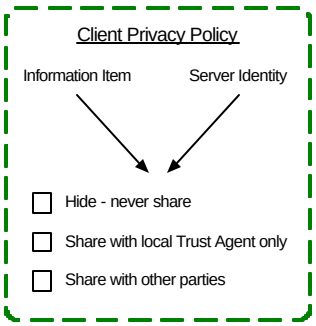

Figure 6

If only a few data items / service providers are being considered, setting policy manually at the client will not be onerous. However, it is appreciated that this situation could quickly develop into one that in unmanageable. Also, defining policies may be too much to expect of most users, so tools that ease the task will be required. For example, the individual may simply choose a 'preferred' policy specified by a third party that they trust. Alternatively, the user could select their preference using a sliding scale of default options. Both possibilities are likely to reduce the degree of control the user has, but are considered pragmatic ways to deal with the complexity.

More complex ways of expressing policy, e.g. scripting language, are possible, which would allow local decision-making processes to be implemented that go beyond a simple policy matching.

8.2 Enforcing policy

Strictly speaking, policy enforcement is an issue for the service provider's server. The process of 'attaching' a *use policy* to a piece of information clearly needs to be initiated by the client. Exactly how enforcement is achieved will depend on the capabilities of the server in ensuring a) that a policy cannot be detached and b) that information cannot be accessed in ways that contradict the wishes expressed in the policy. In certain situations if may not be necessary to provide any technical controls that specifically bind data and policy. Instead, methods that enforce accountability may be sufficient.

There is an expectant that the client and server can trust each other to some degree. Where the service provider acts dishonestly more robust protection mechanisms are required, along the line of Digital Rights Management (DRM) controls. This particular scenario is not addressed in this work.

9 Benefits of the Trust Agent architecture

Regardless of the technical implementation, this Trust Agent model addresses a wide range of situations where trust in a client must be determined, and where there is a sound business rationale for doing so. Specifically:

- The service provider is motivated to manage risks that arise from the operation of the remote client platform, because to not do so carries a direct/indirect financial penalty.
- It presents an open-ended framework that is scalable, and in which a range of technical security and privacy preserving defences can be implemented, including those not yet conceived.
- It is a security architecture that is designed and operates independent of any overarching applications. This is arguable the most effective approach to security since it avoids the potential conflict between application developer and security specialist.
- It requires no significant change to existing environments, e.g. applications, protocols.

10 Conclusion

The Trust Agent, along with the Privacy manager extension, provides an architecture that enables trust in a remote platform to be developed whilst at the same time respecting the privacy of personal information that the platform may hold. The Trust Agent addresses the case where two parties have reason to cooperate with each other but there is a limit to the trust that each has in the other. Equally, each believes that the other is essentially honest and motivated to behave honourably.

The TPM plays an important role in securing and providing baseline evidence that a Trust Agent can be trusted by the service provider and the user.

How we manage and separate client work/play environments, where (say) the client is less willing to entrust the Trust Agent with personal information, remain to be resolved. This limitation encapsulates the shortcoming that first generation trusted platform technologies like the TPM offer.

Other appliances, like mobile phones and intelligent network-enabled printer can host Trust Agents in exactly the same way, and the approach seems to be applicable to ad hoc peer-to-peer networking.

11 References

[Cra03] Crane, Stephen; Personal Trust Assistant (PTA) web site; http://w3.hpl.hp.com/tsl-pta/

[CC03] Cofta, Piotr; Crane, Stephen; Towards the Intimate Trust Advisor; First International Conference on Trust Management; May 2003.

[Pea03] Pearson, Siani; et al; Trusted Computing Platforms: TCPA Technology in Context; Prentice Hall; ISBN: 0-13-009220-7; 2003.

[TAO04] Trusted Computing Group (TCG) Architecture Overview. Available for download from the TCG website (https://www.trustedcomputinggroup.org/home) at https://www.trustedcomputinggroup.org/downloads/TCG_1_0_Architecture_Overview.pdf

[CP04] For Chip and PIN see http://www.chipandpin.co.uk