Position Paper

Principles for Managing Data Privacy

Introduction

Data privacy refers to the evolving relationship between technology and the legal right to, or public expectation of, privacy in the collection, storing, management and sharing of data. This paper specifically covers the privacy aspects of the data, not the monitoring of data by either the corporate or various state and government entities through which this data flows.

Problem

Privacy problems exist wherever uniquely identifiable data relating to a person or persons are collected and stored, in digital form or otherwise. Improper or non-existent disclosure control can be the root cause for privacy issues. The legal protection of the right to privacy in general and of data privacy in particular varies greatly around the world. However the Universal Declaration of Human Rights states in article 12 that:

> *"No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks."*

Privacy and security are related, but are not the same. An organisation can take many measures to secure information but still not protect privacy. In an increasingly de-perimeterised world then, where users interact with organisations, information privacy (and security of any asset of sufficiently high value) requires a segregation of duties (JFC#10[1]) preferably with the data subject having ultimate control of their Personally Identifiable Information (PII).

The fundamental problem is that data is generally protected at the "container" level (where for example a container could be a directory within a file system, a server or a database) and not at the granular data level (JFC#9), resulting in the control on ownership residing with the "container" owner and not the data subject. Thus when data is transferred from one container to another container, however that transfer occurs, the data subject loses track of that data.

Why should I care

Information / data has value, whether confidential corporate data, or personal information, the problem is the same; unless it is properly controlled information leakage occurs. Identity theft is perhaps the most glaring example of the loss of private information. If I am the subject of the information, then I should have the right to decide what happens to that data, who can see it, when they can see it and ideally I should be able to track who has accessed that data and if necessary manage that data retrospectively.

Unmanaged PII risks being accessed and abused without my knowledge, changed and modified and used in ways I neither approved of nor would permit given a choice. PII outside of my control risks losing its integrity either through PII that goes stale and is not updated, or PII that is incorrectly generated/modified; worst of all, PII that I do not manage properly risks the loss of any rights I might have had, or recourse over that information.

Reliance on a Subject Access Request (SAR), to request access to PII held about you depends on you knowing that a body held that information and you suspecting that it is incorrect. Typically this only happened when incorrect information triggers an event – such as being refused credit.

[1] The term JFC#n refers to the relevant Jericho Forum Commandment number. See www.jerichoforum.org

Recommended Solution/Response

To assume that adequate data security solves the need for privacy is misunderstanding the problem. This problem already existed, but is highlighted when trying to implement information privacy in a de-perimeterisation environment.

The privacy information associated with data must be bound to (or reside) with that data. The copying of that data must not lose the associated privacy information. and thus the rights to that information, including the rights to modify, update, restrict or even destroy that information must be retained by the individual it pertains to, even when a 3rd party holds that information (in effect an automated, electronic SAR – see appendix).

When such PII is used or accessed then the subject of the PII should be notified, or, at a minimum, enough data should be logged to enable the subject to understand who (and why) is using their information.

Background & Rationale

Much PII is continuously changing; I am single this year, and married next with a new address, phone number; however my data of birth is static. Most information becomes stale, or is sometimes incorrect through error, omission or confusion. Incorrect PII benefits no-one, neither the person involved, or the person trying to use the information (HR department, credit reference agency, Government department etc.). Furthermore cleansing information is an expensive and time consuming task.

The HR departments of many large corporations are moving to self-service PII management where the employees are able to update their own data; this follows a "what's in it for me" principle. If I want my pay to be correctly credited to my new bank account, or my new child to be covered as part of my company health care package then I update my information. The draw back of these systems is that users only update/check their information at points when it affects them (they probably never updated the "cell-phone" field), and it relies on them to know that their information resides on a system, *and* they have an account to change it, *and* they are motivated to do so. With many companies outsourcing their HR function, then the subject's data may be held on an external 3rd party system, for which they have no credentials to access and possibly in a different legal jurisdiction from which they are a citizen.

Many technologies have a negative impact on privacy. Data warehousing, data mining, data fusion and data meshing technologies have been developed in order to exploit information that has been aggregated from multiple sources. The subject of that information will rarely know that the company has their information, or how it will be used, or give permission for it to be sold on. If you are not able to monitor how your PII is being used (or abused) then who is?

Ultimately, PII should be under the control of the subject of that information; and that "information" should be able to ask for permission to be processed (or prohibit processing), and should be able to "phone home" to validate it is still correct and also log its usage.

This is more imperative as the de-perimeterised world moves to increased peer-to-peer networking, and social networking, where computers and user information is opened up for anyone to see, often through ignorance of the risks, but often deliberately (for example the amount of sensitive information that teenagers choose to publish on "myspace.com").

Challenges to the industry

Enterprise Protection and Control (DRM) of information

In a corporate environment Enterprise Protection & Control of information (see the Jericho Forum Position Paper) is being used to allow the dissemination of information that needs rights management into the de-perimeterisation environment. This information protection and control mechanism should also include the need to maintaining privacy.

Access and segregation of duties

Information (data) that is subject to privacy constraints should ideally use the same protection when at rest as when in transit (JFC#11) – However the ability to have super-user level (root or admin) access to the system holding that data should not breach the privacy of that data.

Fine grain data management is difficult, especially when not designed in up front, which is why most organisations use systems with "container" level data management and network security, despite the evidence of daily reports of data breaches from such systems.

The ability to operate in an un-trusted environment

Operating in an insecure environment (the default assumption for a de-perimeterisation environment) poses a different problem for releasing PII (or selected parts of PII) into that environment.

Here the use of trusted brokers can play a part, verifying or vouching for your identity. An example is the use of one-time credit card transactions, allowing the user to divulge their credit card number (which can only be used once) to an un-trusted vendor, with the credit card company acting as guarantee and broker for the transaction.

Permission to access / use your data when it is held outside of your control

1. An organisation must take reasonable steps to destroy or permanently de-identify personal information if it is no longer needed for any purpose for which the information may be used or disclosed. The last two issues are data security issues.
2. End users should demand full disclosure of the use of personal information as part of any transaction.
3. End users should demand that all PII should be inherently secure. (JFC#9)
4. If privacy protocols are to become adopted as standards then they must be open and interoperable (JFC#3).
5. Information privacy or information protection in this context is not about keeping personal information secret. Instead, it is about creating a trusted framework for collecting, exchanging and using personal data.
6. IT and Information Systems need to be adequately secured to prevent unauthorised access and disclosure of information. One problem with information privacy is inevitable mistakes, such as systems misconfiguration, that exposes personal information.

With the right combination of secure software, careful implementation and sound privacy policies, organisations can make the protection of their users', employees' and customers' privacy robust and cost-effective.

Protecting customers' personal information in an information-intense environment can be challenging because customer / personal data forms the basis for many business decisions. The de-perimeterisation challenge is how to enable information custodians to collect, manage and use private information responsibly.

The way forward

Today

For privacy to be effective it is essential that systems support privacy, and have privacy set as the default option. It is important then, that privacy and information protection issues should be considered from the outset for any computerised system that is to hold PII.

Systems must have simple and easy mechanisms for information subjects to validate the information and understand why and for what it is being used.

Corporate / business data classification should be expanded to include "personal information", as a specific category of data with separate handling requirements. Personal Information should be further sub-categorised to define (for example);

- Business Private Information (BPI) such as your name on business card
- Personal Private Information (PPI) such as home address, date of birth, bank details
- Sensitive Private Information (SPI) such as sexual orientation, medical records

Properly understanding the sensitivity of the data therefore allows the processor of that information to correctly handle it.

The near future

There is no easy solution or "silver bullet" to privacy in a de-perimeterised world. Research is needed into methods to allow personal information to be shared in a way that is consensual, with safeguards around ensuring that it does not become stale, and that the subject of the data is aware of its (continued) use.

The industry needs a suite of "standard" Privacy Enhancing Technologies (PETs – see appendix) to enable the subject of the data to manage, update, authenticate, and grant rights to the use of their data in any environment not controlled by them. Business needs a method to validate personal information and ensure its accuracy.

The development of a PII broker, holding a subject's data, and being a single point of reference for the subject's information is required. Users can grant "facets" of themselves (including anonymous facets) to be used with any organisation needing their information.

The challenge confronting developers of privacy enabling technologies is that the legal, organisational and technical protections need to be trustworthy (JFC#5). If the power to override them is in the hands of a person or organisation that flouts the conditions (JFC#7), then privacy protection collapses.

Appendix

Legislation

Privacy matters emphasise four basic privacy principles:

- Notice – data collectors must inform users of what personal information is collected and who else might share it.
- Choice – users can decide how personal information is used by choosing to opt in or opt out when usage may differ from the original intent.
- Access – users can view data collected about them, and they have control in correcting inaccuracies.
- Security – reasonable security must be in place to ensure accuracy and security of the data.

These four principles surface regularly in the context of information privacy. They figure into the most significant regulations today, including the European Commission's Directive on Data Protection, the UK Data Protection Act, the Safe Harbor Act, the Health Insurance Portability and Accountability Act (HIPAA), Gramm-Leach-Bliley Act (GLB), and similar national privacy regulations.

International legislation and privacy is well catalogued here;
See: http://www.epic.org/privacy/default.html

Platform for Privacy Preferences (P3P)

The Platform for Privacy Preferences (P3P) is the most widely known of the proposed standards, and is an XML standard produced by the World Wide Web Consortium (W3C). The P3P project enables Web sites to express their privacy practices in a standard format that can be retrieved automatically and interpreted easily by user agents.

Enterprise Privacy Authorization Language (EPAL),

Other privacy standards under development include the Enterprise Privacy Authorization Language (EPAL), a formal language designed to specify fine-grained enterprise privacy policies. Unlike P3P, EPAL defines the privacy-practices that are implemented inside an enterprise. The Customer Profile Exchange, or CPExchange, specification defines a data format for disclosing customer data from one party (customer/enterprise) to another. It enables the specification of privacy meta-information as an option, and associates privacy controls with subsets of profile information. The privacy meta-information includes the exchange partners, the applicable jurisdiction, and a privacy declaration (based on P3P).

RFC 2965 guidelines on state management mechanisms for security and privacy protection

For users connecting with organisations over the Internet, informed consent should guide the design of systems that use cookies. A user should be able to find out how a web site plans to use information in a cookie and should be able to choose whether or not those policies are acceptable. In this area, RFC 2965 provides guidelines on state management mechanisms for security and privacy protection when creating stateful sessions with the Hypertext Transfer Protocol (HTTP) i.e. cookies. For privacy, both the user agency and the origin server must assist informed consent**.**

Technologies and countermeasures for privacy

Two general technology categories have emerged – Privacy-Invasive Technologies (PITs) and Privacy-Enhancing Technologies (PETs).

See: http://www.anu.edu.au/people/Roger.Clarke/DV/PITsPETs.html