On Technology and Internet Privacy

John Linn, Sr. Technologist, Office of the CTO, RSA, The Security Division of EMC, Bedford, MA, US

john.linn@rsa.com

Position paper for IAB/W3C/ISOC/CSAIL Internet Privacy Workshop, Cambridge, MA, December 2010

29 October 2010

When considering Internet privacy from a technical perspective, it's a natural reaction to evaluate the problem and anticipate its solution in terms of the Internet's defining characteristic: its protocols. A premise[1] of this paper is that, while this is a necessary component, it isn't sufficient in itself. Rather, I contend that a comprehensive solution to Internet privacy must be data-centric, structured around the information objects whose privacy must be maintained. Privacy properties must persist throughout the information's lifetime, not only within the ephemeral scope of a particular Internet communication. A comprehensive architecture must combine several elements, including:

- Protection facilities associated with data objects as they traverse the Internet and are stored at endpoints
- Mediation functions within or associated with the systems processing the data objects
- Secure channels within (or encapsulating) the distributed protocols within which the data objects are exchanged
- Distributed authentication and identity management methods, including pseudonym support
- Annotation facilities describing privacy-relevant attributes of principals, processing entities, and objects

Corresponding management interfaces are also needed, with interaction characteristics and rights appropriate to the users and administrators that will employ them. When constructed and operated properly, these can enable essential linkages among preferences, policies, and the technologies deployed to support them.

Conventional security technology focuses on means to ensure that data is available only to those entities that are authorized to obtain it. For privacy purposes, this is necessary but insufficient. Privacy goals also require that authorized holders be constrained in terms of how they may use data. For example, it may be acceptable to use a provided phone number to confirm delivery of a parcel, but not for subsequent telemarketing. This introduces significant new challenges. Approaches to such problems fall into two general categories:

[1] Further related commentary and background is available within a prior article: J. Linn, "Technology and Web User Data Privacy: A Survey of Risks and Countermeasures", IEEE Security and Privacy, January-February 2005, pp. 52-58.

- Reliance on trusted data holders: here, the system receiving sensitive data is relied upon to act according to understood and accepted policy, possibly informed by user inputs and/or protocol-based indicators. Once data is shared, it becomes part of a system operated on behalf of its recipient. Privacy assurance is based on trust in that system. This model represents today's common mode of usage.
- Mediated privacy enforcement: here, users are assumed to remain suspicious of systems they access across the Internet, and unwilling to share sensitive data with them fully and directly. Instead, access is mediated for particular usages, perhaps allowing filtered queries or selective transfer of data to other destinations. In this model, data sharing relies on technology operating on behalf of the data's owner or a neutral third party. Possible realizations include use of brokering services to maintain sensitive data, or transfer of data in a protected representation whose contents are accessed indirectly via a form of rights management technology.

Within both models, data holder systems must be implemented and managed securely in order to promote confidence in trustworthy processing; transparency and auditability are important measures and warrant architectural support. Nonetheless, the first model (and current practice) suggests a possible conflict at data holders: simultaneous operation both on behalf of users and of the holders' potentially divergent interests. Technical approaches, per the second model, may ameliorate this concern and reduce the level of trust that users must place in peers and their operations. As privacy protection is enhanced within applications, I encourage designers to evaluate and support approaches where data access may be mediated outside the direct control of a single application peer. Although this prospect may transform some pairwise exchanges into multi-party transactions, it represents an important direction towards more trustworthy privacy control.

From a protocol perspective, it's important to accommodate privacy management of sensitive data that is present and used within different distributed applications. This doesn't necessarily mean, however, that most aspects of privacy management must or should be application-specific. As a goal, it seems desirable to unify processing of sensitive data in a consistent fashion, potentially through development of a cross-protocol framework that can be included by reference by particular applications and managed through common means.

Internet privacy is a critical topic, raising challenging problems that reside and must be addressed on both sides of the interface between people and technology. Today, privacy assurance relies on trust that is hard to avoid but that may not always be well-placed. The balance of power and control resides in the hands of data holders, but technical means can be applied to adjust this balance. Standards can provide important components for this process, but should evolve in a consistent fashion and broader context rather than in fragmented isolation from one another.