Chapter 2.36
Modeling Method for Assessing Privacy Technologies

Michael Weiss
Carleton University, Canada

Babak Esfandiari
Carleton University, Canada

INTRODUCTION

Late in 2004, Rice University researchers uncovered a flaw in Google's Desktop Search tool that could release private local data to an untrusted third party. Earlier in 2004, a flaw in Apple's Safari Web browser was reported that allowed an attacker to upload and execute an arbitrary program on the user's machine. These flaws spotlight some of the security and privacy risks users are exposed to when systems are composed from independently created components or services that interact in unexpected ways.

In both examples, the application designers had made incorrect assumptions about the identity of the initiators of service requests. For example, in the case of the Safari Web browser flaw, the help viewer would execute scripts referenced in an URL and not check who made the request. It was assumed that it would originate with the help viewer. This "feature" of the help viewer could be exploited by downloading and mounting a disk image with a malicious script and then asking the help viewer to execute it.

Such unexpected interactions are also known as *feature interactions*. Feature interactions occur when independently developed and separately tested components (also known as *features*) are combined, and the combination results in undesirable side effects. They are difficult to anticipate, in particular in an open system such as the Internet, due to the combinatorial number of ways features can interact with one another. These side effects are often non-functional in nature and in many cases are related to privacy or security.

As the examples show, failing to consider the actual identity of the requestor of a service may cause serious privacy and security breaches. The authenticity of a user or service provider on the Internet cannot be taken for granted. A variety of technique—smart cards, personal information devices, single sign-on services, to name but a few—have been developed to address this issue. However, the benefits and convenience of these

techniques must be weighed against the privacy and security issues their use may raise. The focus of this chapter is, thus, on privacy technologies and *assessing* them for privacy issues. For these to be accepted in the user community, we must ensure that there are no new privacy and security risks by using the very techniques intended to address such issues. We introduce a *modeling framework* for assessing privacy technologies and their possible side effects on overall system concerns such as privacy and security, and demonstrate its use for analyzing the privacy pitfalls of single sign-on services.

PRIVACY ASSESSMENT

Privacy concerns the collection, storage, use, and sharing of data about individuals (Cannon, 2005). It has several dimensions, including the protection of personal data, or *privacy protection*. Individuals do not want data about themselves to be shared with other parties without their consent, and when data is held by another party, they want to have control over the data that is held about them and its use (IPC, 2000).

Policy makers have defined *privacy protection principles*, or sets of practices for implementing and enforcing privacy. There have been several attempts to develop such sets of practices, such as the Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data developed by the OECD (OECD, 1999), or the Ontario Freedom of Information and Protection of Privacy Act (IPC, 1990).

Privacy (impact) assessment is the evaluation of a system (or service) for compliance with privacy protection principles (IPC, 2000; TBC, 2002). Aspects to consider during a privacy assessment include data flow, usage, security, user control, user access, disclosure, and dependency (Cannon, 2005). Data flow captures what data is collected and shared by a service. Usage refers to the purposes and parties using of the data.

Security relates to the use of encryption during data transfer and storage. User control refers to the control users have over what data is collected, whereas user access captures what data a user can view and change. By disclosure we mean how users are informed about the collection and use of their data. Dependency relates to dependencies between your system (or service) and other systems (upstream and downstream).

It should be noted that privacy and security issues are often intertwined. On one hand, for example, data encryption is considered an important component of privacy. On the other hand, security and privacy goals can often be achieved within the same technical solution. A good example is identity management technology, which can be used to achieve both authenticity and privacy. Privacy-enhanced identity management (Damiani, 2003) would, for example, allow the user to control the release of personal data.

PRIVACY ANALYSIS

One common approach to analyze a system for privacy weaknesses is to model them as dataflow diagrams (Cannon, 2005). They lend themselves to tracking the flow of data in a system. They provide a visual means for validating that we have captured all data that is collected, stored, used, and shared. Dataflow diagrams consist of processes (which manipulate data), data stores (where data is stored), entities (which consume or create data), and dataflows (which show the flow of data between the other types of elements).

During privacy analysis, a regular dataflow diagram is often annotated with attributes that apply to the data being sent and stored by the system. Common attributes are inclusion in privacy statement, retention policy, security, user control, and encryption. To represent complex systems, dataflow diagrams can be decomposed into subdiagrams. Another interesting feature is the notion of privacy boundaries that encapsulate