Identify sensitive data and how they should be handled

Data are critical to organizations and to users. One of the first tasks that systems designers must do is identify sensitive data and determine how to protect it appropriately. Many deployed systems over the years have failed to protect data appropriately. This can happen when designers fail to identify data as sensitive, or when designers do not identify all the ways in which data could be manipulated or exposed.

Data sensitivity is context-sensitive. It depends on many factors, including regulation (which is often mandatory), company policy, contractual obligations, and user expectation. Note that sensitive data are not always user-generated input. Rather, they include data computed from scratch, data coming from external sensors (e.g., geolocation and accelerometer data on mobile devices), cryptographic material, and Personally Identifiable Information (PII). Creating a policy that explicitly identifies different levels of classification is the first step in handling data appropriately.

It is important to factor all relevant considerations into the design of a data sensitivity policy. For example, there are numerous regulations that system designers must consider, ultimately creating a unified approach that consistently addresses them all. A number of examples may help to flesh this out: various jurisdictions impose regulations on how personal data should be handled (e.g., medical records); the EU Data Protection Directive differs from the regulations in the United States; and PCI compliance issues, though not regulatory, directly affect data protection requirements.

Not all data protection requirements are the same. For some data, confidentiality is critical. Examples include financial records and corporate intellectual property. For data on which business continuity or life depends (for example, medical data), availability is critical. In other cases, integrity is most important. Spoofing or substituting data to cause a system to misbehave intentionally are examples of failures to ensure data integrity. Do not

conflate confidentiality alone with data protection.

Technical data sensitivity controls that a designer might consider include access control mechanisms (including file protection mechanisms, memory protection mechanisms, and database protection mechanisms), cryptography to preserve data confidentiality or integrity, and redundancy and backups to preserve data availability.

Data sets do not exist only at rest, but in transit between components within a single system and between organizations. As data sets transit between systems, they may cross multiple trust boundaries. Identifying these boundaries and rectifying them with data protection policies is an essential design activity. Trust is just as tricky as data sensitivity, and the notion of trust enclaves is likely to dominate security conversations in the next decade.

Policy requirements and data sensitivity can change over time as the business climate evolves, as regulatory regimes change, as systems become increasingly interconnected, and as new data sources are incorporated into a system. Regularly revisiting and revising data protection policies and their design implications is essential.