Summary

Introduction

Paper-based communication and transaction mechanisms are being replaced by automated transaction mechanisms at a breath-taking pace. Traditional security mechanisms such as photographs, paper-based certificates, and handwritten signatures are rapidly becoming outdated: they require physical transport or proximity, are increasingly vulnerable to counterfeiting and unauthorized duplication, and can be stolen, extorted, or irreversibly destroyed. The enormous potential of communicating and transacting in cyberspace (including the Internet, e-mail, cable TV, and mobile phone networks such as GSM) and in the physical world (by means of smartcards and handheld computers) can only be unlocked if the new communication and transaction mechanisms are adequately safeguarded.

Digital certificates are by far the most promising technique for safeguarding electronic communications and transactions. Just like passports, diplomas, drivers' licenses, and other traditional certificates, they can specify any kind of data. Digital certificates are no more than cryptographically protected sequences of zeros and ones, and so they can be transferred electronically to any place on earth (and in space) without noticeable loss in time or costly human intervention. Digital certificates offer unprecedented security: even if all the computing power on earth could be tapped, it would take millions of years to forge a digital certificate.

Digital certificates are already widely used on the Internet, to authenticate e-mail, Web servers, and software. The most popular Web browsers have built-in capabilities for storing, sending, and verifying digital certificates. Digital certificates are also playing an increasingly important role in electronic payments, access control (to Web sites, databases, etcetera), digital copyright protection, electronic voting, electronic patient files, and so on. Around the world, transport organizations, municipalities, health care providers, financial institutions, and other influential organizations are planning to provide their customers with digital certificates that will be the sole means of participating in their systems. In the near future, digital certificates may be built into any device or piece of software that must be able to communicate securely with other devices or with individuals. This includes mobile phones, watches, televisions, cars, and conceivably even computerized household appliances.

The problem

While their prospects look bright and shiny, digital certificates have a dark side that has received surprisingly little attention thus far. Unless drastic measures are taken, it will not take long before everyone is forced to communicate and transact in what will be the most pervasive electronic surveillance tool ever built. Each digital certificate can be traced uniquely to the person to whom it has been issued (or to the device in which it has been incorporated) and can be followed around instantaneously and automatically as it moves through the system. Even digital certificates that do not explicitly specify the identity of their holder can be traced in a trivial manner, because the string of zeros and ones that makes up a digital certificate for security reasons must be unique; digital certificates in this respect offer no more privacy than Social Security numbers, credit card numbers, and health registration numbers. On the basis of these unique serial numbers, which will travel along whenever an individual engages in a communication or a transaction, organizations and individuals can compile extremely detailed personal dossiers. The dossiers can be compiled and linked without human intervention, can be dynamically updated in near real time, and will contain minute information about a person's financial situation, medical history and constitution, lifestyle, habits, preferences, movements, and so on. Any digital signatures made by certificate holders can be added to their dossiers; they form self-signed statements that cannot be repudiated. With the cost of digital storage space dropping all the time, all dossiers will be stored potentially forever.

Furthermore, digital certificates can be misused to deny a certificate holder access to services, and to block his or her communication attempts in real time. For example, certificate blacklists can be built into Internet routers. Also, transaction-generated data conducted with target certificates can be filtered out by surveillance tools, and delivered electronically to law enforcement and other third parties for examination or immediate action. Online certificate validation services even enable central authorities to learn in real time who communicates with whom and to falsely deny access.

These exceptional surveillance powers will be enjoyed not only by all the organizations that a person directly communicates or transacts with, but also by a myriad of private and public organizations that routinely acquire dossiers, by unscrupulous employees, by hackers, by law enforcement and intelligence agencies, and by all organizations that issue digital certificates. Typical representatives of the latter group will be financial institutions, governments (local, state, and federal), insurance companies, health care providers, post offices, public transport organizations, and consumer credit bureaus.

Smartcards exacerbate the privacy problems. As Moreno, the inventor of the first generation of smartcards, remarked, smartcards have the potential to become "Big Brother's little helper." It is almost impossible to verify that a smartcard does not leak personal data stored inside the card, and different applications can all share the same card data without the consent of the cardholder.

The solution

This book analyzes and documents the privacy dangers of digital certificates. On the basis of the findings, practical digital certificates are constructed that preserve privacy without sacrificing security. The new certificates function in much the same way as cash, stamps, cinema tickets, subway tokens, and so on: anyone can establish the validity of these certificates and the data they overtly specify, but no more than just that. A "demographic" certificate, for instance, can specify its holder's age, income, marital status, and residence, all digitally tied together in an unforgeable manner.

The new certificates are not only much more secure and efficient than their non-electronic counterparts, but also much more powerful. For instance, each certificate holder can decide for him or herself, depending on the circumstances, which property to disclose of the data encoded into a digital certificate. This goes beyond the analogy of using a marking pen to cross out data fields on a paper-based certificate; a certificate holder can prove that he or she is either over 65 or under 18, for instance, without revealing which is the case. More generally, certificate holders can demonstrate any satisfiable proposition from proposition logic, where the atomic propositions are linear relations in the encoded data; any other information remains unconditionally hidden.

Also, a certificate can be presented in such a manner that no evidence is left at all of the transaction; this is much like waving a passport when passing customs. Alternatively, it can be presented in such a manner that the only information left is self-authenticating evidence of a message or a part of the disclosed property; this is much like presenting a paper-based certificate with crossed-out data fields so that a photocopy can be made. Furthermore, the self-authenticating evidence can be limited to designated parties.

The new techniques enable certificate issuers to discourage lending of personal certificates. An issuer of gender certificates (needed to gain access to gender-specific online forums, say) could encode into each certificate not only a bit indicating the gender of the designated receiver, but also the credit card number or some other secret of the receiver. While certificate holders can hide their built-in secrets when they show their certificates, it is not possible to show a certificate without knowing the built-in secret. Therefore, certificate holders cannot lend their certificates without revealing their secrets.

Another useful technique makes it possible for a central authority to compute all the data that have been encoded into a certificate once that certificate is shown more than a predetermined number of times. In particular, a built-in identifier can be computed even if the certificate holder never discloses any of the built-in data when showing his or her certificates. This magical security property holds even when the certificate holder is free at each occasion to choose the property that he or she demonstrates when presenting the certificate. It allows the certificate issuer to trace and contain fraud with limited-show certificates (such as subway tokens and electronic coins), to (further) discourage unauthorized lending and copying of personal certificates, and to discourage the destruction of unfavorable certificates (such as a mark for drunk driving or late payment).

Yet another technique enables a certificate issuer to refresh a previously issued certificate without knowing the encoded data. The unknown encoded data can even be updated before it is recertified. By way of example, a doctor could issue a prescription to a patient for 20 doses of a penicillin cure. Each time the patient visits a drugstore to collect some of the doses, the drugstore can verify that the patient is still eligible and can decrement the number of remaining penicillin doses. On the other hand, no drugstore can determine the total number of doses prescribed or the number remaining at the time of a visit, nor can different visits by the same patient be linked. Using our certification techniques, the patient could even pay for each dose in untraceable electronic cash and receive a digital receipt that could be used to get reimbursed by his or her health insurance company.

We also describe techniques to improve the privacy of organizations. In particular, we show how an organization can verify a certificate in such a manner that it receives self-authenticating evidence that proves that the certificate has been shown but unconditionally hides all or an arbitrary part of the property that has been demonstrated. In applications where organizations submit the certificates they receive to a central authority, to enable the latter to compute statistics or to combat fraud, this property prevents the central authority from learning which information an organization's customers disclosed. Organizations cannot provide false information to the central authority, but in the case of disputes they can always reveal additional information about the demonstrated properties.

All these and other software-only techniques can be implemented in tamper-resistant smartcards. The smartcard offers strong protection against loss, theft, extortion, lending, copying, and discarding of certificates, and can restrain its holder from other undesired behavior. More generally, the smartcard can be used either to strengthen our software-only security provisions or to add security features that software-only techniques cannot enable at all. Also, the presence of the smartcard removes the need for online authorization or frequent distribution of certificate revocation lists. At the same time, the smartcard can be prevented from learning the certificates of its holder, the information encoded into the certificates, and even the properties that are demonstrated when showing digital certificates. In addition, any data leakage by or to the smartcard can be blocked. The cardholder can even prevent his or her smartcard from developing information that would help the card issuer to retroactively trace the cardholder's transactions should the card's contents become available to the card issuer. Transactions can be completed within as little as $1/20$-th of a second by a standard 8-bit smartcard processor, so that road-toll pricing and other demanding applications are entirely feasible.

Our techniques protect privacy in the strongest possible sense: even if all organizations (including those that issue certificates and those that verify them) conspire and have infinite computing resources, and issue smartcards that are programmed in adverse manners, they cannot learn more about (honest) certificate holders than the assertions they voluntarily demonstrate. Different actions by the same certificate holder cannot be linked, unless the certificate holder consents and cooperates.

Our certification techniques are advantageous not only to individuals, but also to organizations: they prevent certificate issuers and other central parties from competing unfairly; they minimize the need to consult certificate revocation lists or online certificate validation services; they minimize the scope for law enforcement intrusions on databases; they minimize the need to protect online databases against intrusions by hackers and insiders; they reduce the scope for discrimination and identity fraud; they foster fair competition with respect to the collection and use of personal data; they are the cheapest and most effective way to comply with most of the fair information principles of privacy legislation and codes of conduct; they improve transaction finality; and, they cultivate goodwill among customers.

The presented techniques could help stimulate the public acceptance of smartcards, because low-cost smartcards without cryptographic coprocessors can be used and smartcards cannot be misused for the purpose of surveillance. They could even stimulate the growth of electronic commerce by providing a firm grounding for upcoming digital signature legislation. Namely, secret keys protected by smartcards (or other tamper-resistant devices) with biometric protection are not vulnerable to theft, extortion in cyberspace, and unauthorized use, and can therefore be reliably associated with a particular individual.