Access Management Based On Digital Credentials Part 2

Dr. Stefan Brands

Credentica and
McGill School of Computer Science

Abstract

This is the second installment of a two-part series on how to do digital access management in a highly secure yet non-intrusive manner that scales seamlessly across organizational boundaries. This installment describes a secure cross-domain access management architecture built around the Digital Credentials primitive introduced previously, and discusses the benefits of the new approach. Following this, we examine the use of the new approach in the context of four example applications: Electronic Health Record management, national ID chip-cards, e-government, and Digital Rights Management. We then discuss in greater detail why PKI and other conventional authentication techniques are a poor solution, and end with a brief outlook on the future of access management.

Cross-Domain Access Management

Most providers of access management products deal with security by way of password-only authentication. With the trend of moving from single-domain access management to cross-domain access management, however, the password-only approach rapidly breaks down. Passwords are increasingly being sent over public networks and can be misused by increasing numbers of access providers. Juggling multiple passwords causes serious user inconvenience and a dramatic overhead increase for administrators (not in the least due to people being more prone to forgetting their passwords). *Single sign-on* avoids having to remember multiple passwords, by giving each user a single password for all resources; however, this user convenience comes at the cost of a further reduction in security.

Secure electronic access management requires the application of cryptographic techniques (possibly implemented in tamper-resistant devices and augmented with biometrics). Both asymmetric and symmetric cryptography make it possible to electronically authenticate information on the basis of secret keys that (contrary to simple passwords) never leave the physical confines of devices held by their legitimate owners. At the same time, users no longer need to memorize a plurality of passwords. PKI is generally believed to be the most secure solution for access management. Through the secure distribution and management of public keys and digital certificates, trust can be established without prior acquaintance.

As briefly explained in the first installment [1] of this series, however, the use of PKI in the context of access management has a number of highly undesirable side effects related to performance and privacy. Privacy problems arise because PKI fundamentally relies on the transmission of unique inescapable identifiers in order to provide security. This may not be a problem in the context of single-domain access management, where scalability needs are low and privacy expectations typically are non-existent. (Think, for instance, of employees of an enterprise remotely accessing corporate network services over a VPN.) In single-domain settings, however, it may be equally secure yet more practical to use symmetric authentication (e.g., Kerberos) instead of PKI, or even simply password-only authentication.[1] In other words, while the negative side-effects of PKI may not be detrimental, the need for public-key cryptography for single-domain access management is highly unclear in the first place.

How different the situation is in the context of access management across administrative domains (let alone across trust domains). Access providers are now confronted with access requests originating from parties who are not (or only partially) registered in databases within their own administrative domain. By way of example, think of the kinds of data that government officials must verify when handling an application for a long-term work visa. For requests that cannot be handled completely within their own domain, access providers will need to access data residing in databases outside their own administrative domain in order to make an authorization decision. Security, privacy, and performance (including administration and scalability) needs all go up rapidly with each addition of an administrative domain, access policy requirement, back-end database, participant, access point, or resource. Why PKI, or any security mechanism for that matter that relies on unique identifiers pointing into back-end databases (including symmetric authentication, password-only authentication, and end-to-end biometrics), rapidly breaks down in this environment, will be explained later in the paper.

One might be tempted to rectify the situation by centralizing all the administrative data from different domains into a single trusted domain situated in the middle of everything. This approach maps the cross-domain context back to the single-domain context, which we know how to handle using password-only authentication, Kerberos, or perhaps PKI. However, centralizing systems of a decentralized nature brings its own administration, scalability, security, and privacy problems, which may be far worse than the original problems. In its original Passport architecture, for example, Microsoft relied on the centralization of all authorization data; the company had little choice but to switch to a “federated” approach following complaints from consumer groups, EU government officials, and organizations reluctant to entrust Microsoft with their customer data. Major industry efforts in access management, such as Liberty Alliance, WS-Security, and SAML, are driven by a growing awareness that centralization is often undesirable for individuals and organizations alike. These promising standardization initiatives do nothing to address the fundamental problems of authentication methods themselves, however. Instead, they merely assume the availability of suitable cryptographic authentication techniques, be they passwords, Kerberos tokens, digital identity certificates, or something else. Indeed, nothing stops one from plugging in *Digital Credentials*. As we have seen in the first installment [1] of this series, Digital Credentials enable fine-grained privacy control (both for users and verifiers) and have security and smartcard performance properties that go well beyond those of conventional cryptographic techniques.

Our cross-domain access management architecture in the next section deviates in several ways from the Liberty Alliance architecture, in order to accommodate a range of unique features brought by Digital Credentials. The resulting architecture provides for genuinely pseudonymous identity certificates, enabling access requestors as well as access providers to prevent linking of their activities across different spheres of activity. More interestingly, the entries in back-end databases can be wrapped efficiently into portable Digital Credentials in a manner that

- does not adversely affect the security of organizations,
- gives access requestors and providers full control over what others can learn about them as they access resources and back-end databases, and;
- scales seamlessly across disparate systems and trust domains.

Non-Intrusive Cross-Domain Access Management

We will refer to our access management architecture as the Credential Management Platform (CMP). Figure 1 represents a schematic overview of CMP.

Architecture Overview

CMP is characterized by three central notions: *records, participants, and protocols*.

A record is a logical collection of information. Records may be held in a central database, may be distributed across multiple databases, or may be held locally on a user device. In the first two cases the record is called a *Remote* record; in the latter case it is called a *Local* record. In general, Local records offer greater security and privacy to access requestors, but may be less convenient. Implementations of CMP could facilitate the automated sharing and synchronization of Local and Remote records in accordance with applica-

[1] Even in those cases where PKI’s non-repudiation feature is required, there is no need for PKI certificates. Instead, the validity of public keys can simply be checked in a central database, exactly as Diffie and Hellman originally proposed when they introduced public-key cryptography in the seventies.

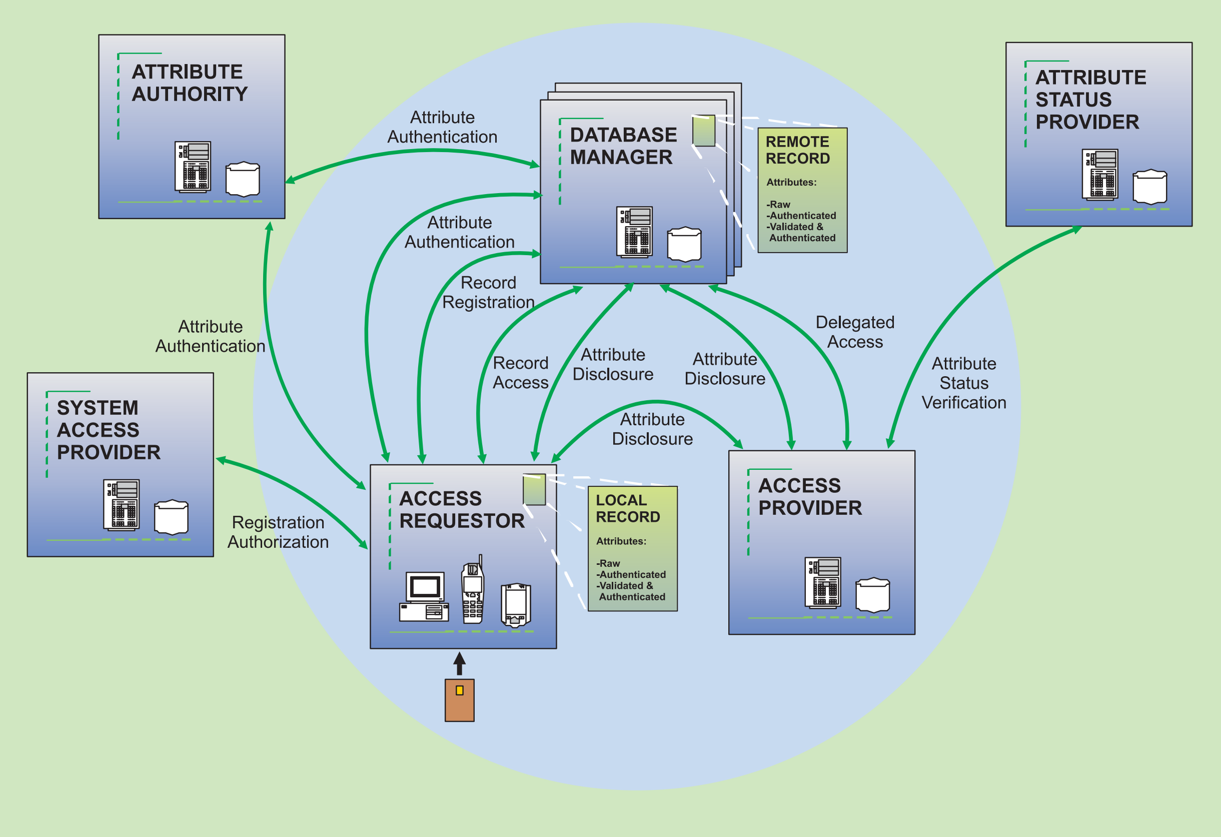

Figure 1 - **Credential Management Platform: Records, Participants, and Protocols**

tion-specific administrative data, to allow multiple records to be managed electronically as one logical entity.

A record contains two kinds of information:

- *Zero*[2] *or more attributes*: An *Attribute* is any personal data, corporate intelligence data, or otherwise sensitive information to which access must be guarded. Attributes may be encrypted by a key known only to a participant; this is useful for instance when attributes that are normally held in a Local record are temporarily stored on a public network to support roaming access by other devices.
- *Related administrative data*: The *Administrative Data* describes rules that specify by whom each attribute in the record may be read, written, modified, or otherwise accessed. Administrative data can include audit trails (possibly digitally signed) for access events.

CMP distinguishes between three kinds of attributes in a record:

- *Raw attributes:* These are attributes specified by any party without any guarantee as to their validity. Personalized display or content preferences for a Web site are an example. Modification or discarding of raw attributes by unauthorized participants might cause inconvenience to the party to whom the data pertains, but would not adversely affect the security of any other party.
- *Authenticated attributes:* Attributes that are digitally authenticated by a participant by means of a digital signature, but without prior verification of their validity. This prevents other participants from modifying the attribute. In an on-line chat group discussing gender-related issues, for example, a person might wrongly specify his own gender but would be stuck with it in future sessions.
- *Validated attributes:* Attributes that are digitally authenticated by an *Attribute Authority* only after the validity of the attribute has been verified by that Attribute Authority.[3]

[2] CMP includes the case of "empty" records, to encompass situations where resources other than information records are being accessed (e.g., downloading a document from a protected Web site, or accessing a site to store something). In such situations, not all the power of CMP is put to use. Notably, there is no need for functionality dealing with the digital authentication of entries in Remote records.

[3] Attribute information may be supplied to Attribute Authorities by "Registration Authorities" who are responsible for validation; we do not single out Registration Authorities in our architecture, however.

Authentication of authenticated and validated attributes takes place by wrapping one or more attributes into a Digital Credential, to offer unique security, privacy, and usability benefits. Different attributes may be packaged either into separate Digital Credentials or into the same Digital Credential. Furthermore, different Attribute Authorities may authenticate the same attribute by packaging it in different manners. By way of example, consider an electronic patient record: multiple doctors may digitally "sign off" on the same entries in a patient record. More generally, multiple Attribute Authorities may package the same or different overlapping subsets of attributes in a record in different ways into Digital Credentials. In this manner, access providers can be assured that the data entries on which they rely have been entered by authorized parties, and different parties can effectively maintain partial ownership of information in a record. Not even the party (or combination of parties) controlling the storage of a record can modify, delete, or add information, unless they are properly authorized.

A *participant* is a device or application (or a collection of devices or applications) that acts either autonomously or on behalf of an individual, a group, or an organization. For simplicity we will interchangeably refer to participants as both devices or applications and the parties they represent. CMP distinguishes between six types of participant:

- *Database Manager:* A party that controls the physical storage of records.
- *Attribute Authority:* A party that issues authenticated or validated attributes. These attributes may be valid only a limited number of times or only for a limited-time period.
- *System Access Provider:* A special Attribute Authority responsible for granting participants the right to *initialize* Remote records (and possibly to subsequently manage it in a co-owner role). The System Access Provider issues *Registration Tokens*, either one per participant until token expiry or a new one at regular time intervals or when requested.
- *Access Requestor:* A party interested in accessing a service that requires an authorization decision. The Access Requestor may be represented by a PC, a hand-held, a mobile phone, a smartcard, or any other device (or combination of devices) capable of computing and communicating.
- *Access Provider:* A party that relies on some or all of the attribute information in a record in order to make an authorization decision pertaining to an Access Requestor. Attributes in the record (more generally, properties about attributes in one or more records) are presented to the Access Provider either by the Access Requestor or by the Database Manager. In the latter case, either the Access Requestor's active involvement or prior explicit consent (in the form of a *Delegation Token*) is needed. The Access Provider may resort to an Attribute Status Provider to complete the verification of authenticated and validated attributes.
- *Attribute Status Provider:* A party that verifies the status of one or more attribute-related requests presented by an Access Provider. Its primary role is to verify the revocation status of validated Attributes, to manage and issue updates of revocation lists, and to keep track of the number of times a limited-time attribute has been used.

In a real-world application, there will normally be many instantiations of most types of participant. For instance, in an electronic health record management system, each doctor authorized to update patient records would be an Attribute Authority. Of course, the role of multiple participants from the same or from different systems may in practice all be performed by the same party.

Participants interact with each other by means of *protocols*. CMP distinguishes between seven protocols:

- *Registration Authorization:* A protocol between a System Access Provider and an Access Requestor whereby the Access Requestor obtains a Registration token allowing him to subsequently initiate a Remote record. The Registration token may be issued to a tamper-resistant device (e.g., a smartcard) of the Access Requestor for greater security.
- *Record Registration:* A protocol between an Access Requestor and a Database Manager whereby the Access Requestor presents a Registration token to initialize a record. As part of the protocol, the Access Requestor and the Database Manager will specify administrative data for the record.
- *Record Access:* A protocol between an Access Requestor and a Database Manager whereby the Access Requestor accesses a record stored by the Database Manager in order to read, write, or modify attribute information. The Access Requestor must show an Authorization credential (which may be the Registration token) to demonstrate proper access rights.
- *Attribute Authentication:* A protocol whereby an Attribute Authority issues an authenticated or validated attribute for entry into a Local or Remote record. The Attribute Authority issues the authenticated attribute either upon receiving an Authorization credential or upon receiving authenticated attribute information issued by another Attribute Authority.
- *Attribute Disclosure:* A protocol whereby an Access Requestor discloses attribute information to an Access Provider. The protocol can be

conducted either with or without the assistance of the Database Manager. For Local records, there is no need to involve a third party in order to disclose attribute information to the Access Provider. For Remote Records the Access Requestor can disclose that information to the Access Provider either by directly retrieving it online and forwarding it to the Access Provider, or by routing its own access request through the Access Provider to the Database Manager.

- *Delegated Access:* CMP allows the Access Requestor to provide the Access Provider with a digitally authenticated Delegation token specifying the latter's access rights, so that the latter can later access a record (perhaps for a limited period of time or a limited number of times) without further involvement from the Access Requestor's side.

- *Attribute Status Verification:* A protocol between an Access Provider and an Attribute Status Provider whereby the Access Provider requests and obtains information on the status of an attribute beyond what it can infer from the attribute itself. Attribute Status Verification may take place either on-line (in conjunction with an Attribute Disclosure protocol) or off-line. For short-lived authenticated and validated attributes, the Attribute Status Verification protocol may not be needed.

All tokens, access requests, and other forms of authentication in these protocols are implemented using Digital Credentials. Specifically, as mentioned in the first installment [1] of this series, our access management architecture relies on Digital Credentials in four basic manners:

- To directly implement access privileges, entitlements, delegations, and any other attributes that access requestors show to access providers to allow them to make local authorization decisions;
- To implement privacy-enhanced digital identity certificates (usable as digital pseudonyms where identification is not required) that allow the separation of different spheres of activity;
- To authenticate the entries of electronic records stored in central or distributed databases; and
- To implement digital audit trails and digital receipts that witness details of access requests.

Unique benefits of CMP

As a direct consequence of using Digital Credentials throughout the CMP architecture, a number of unique benefits arise (cf. [1]), including the following:

- The Registration token can be presented in a manner that does not enable identification of the Access Requestor. (Recall that Digital Credentials encompass identity certificates as a special case: an identifier is just one of infinitely many attributes that can be encoded into a Digital Credential, and the Digital Credential holder can disclose it whenever desired.)

- For Remote records, the Access Requestor can choose to be identified or to remain pseudonymous. The ability to pseudonymously hold a Remote record reduces the risk of identity fraud, and minimizes the damage that can be done by malicious insiders and outside attackers. In the case of a dispute a pseudonymous Access Requestor will not be able to deny having accessed the record; only pseudonymous Access Requestors who did not access the record can prove they did not do so.
- In the case of Local records, CMP allows the Access Requestor to be fully anonymous. The authenticators of attributes in the record can strongly discourage the Access Requestor from cloning or lending his attributes. Furthermore, the Access Requestor can present the Attribute Authority with a previously issued authenticated or validated attribute to have it re-authenticated or updated, without enabling the Attribute Authority to learn more than it strictly needs to. In the case of a limited-show attribute, a built-in identifier, value token, or self-signed fraud confession will be exposed if the attribute is used more times than allowed.
- The Access Requestor can disclose only the minimum attribute information (such as a particular property of multiple attributes) needed to meet the authorization requirements of the Access Provider. (In case the attribute is stored in a Remote record, this requires the Access Requestor to have some trust in the Manager.)
- In case Digital Credentials are issued to smartcards, all computationally expensive operations for the smartcard can be off-loaded to a more powerful device; virtually no smartcard storage space is required in that case, so that plenty of room is left for a software solution to protect against clever attacks such as differential power analysis. Also, CMP can offer protection against fake-terminal attacks and smartcard data leakage by routing communications from and to the smartcard through a device trusted by the card holder.
- Attribute Authorities can digitally authenticate information on behalf of others without being able to learn attribute data that they have no need to know. Likewise, Attribute Status Providers can validate certificates without being able to learn the identities of access requestors and access providers. In this manner, Access Providers and Database Managers can outsource core tasks related to digital authentication and authorization to security specialists, without having to provide them with sensitive information.
- Attribute information can be presented to the Access Provider in such a manner that the Access Provider is left with self-authenticating evidence that digitally proves only a part of the Attribute property disclosed by the Access Requestor; this enables the Access Provider to pass on the evidence to third parties (such as the Attribute Status Provider), while being able to protect its own privacy, comply with privacy legislation, and avoid leakage of competitive data.
- Digitally signed audit trails can be kept, which may be identified, anonymous, or role-based; in the case of a dispute only Access Requestors who did not access the record can prove they did not do so.
- For Access Providers, Record Access can be identified, pseudonymous, or anonymous. The latter two cases prevent the Database Manager or the Attribute Status Provider from gaining competitive intelligence on Access Providers or from improperly rejecting valid requests for access on the basis of the identity of the Access Provider. At the same time, the Access Provider can disclose exactly that which is required to enable the Database Manager to make its own authorization decision: CMP provides for role-based access. The Database Manager and other parties can strongly discourage the Access Requestor from reusing, lending, pooling, discarding, or cloning his access rights, even for pseudonymous access.

Example Applications

We now discuss the benefits of using CMP in the context of several emerging applications that fundamentally rely on cross-domain access management.

Electronic Health Record Management

An *Electronic Health Record* (EHR) is defined as the health record of an individual that is accessible online from many separate, interoperable automated systems within an electronic network.[4] EHRs can contain a variety of data and be used for different purposes by different parties involved in health care. They enable highly efficient and remote exchange of health information, remove administrative burdens, enhance productivity, open up new opportunities (e.g., remote health care services), reduce costs, and greatly reduce medical errors.

[4] A literature search reveals an abundance of related terms, including Computer-based Patient Records, Patient-carried Patient Records, Computerized Medical Records, Electronic Patient Records, Electronic Medical Records, Personal Health Records, Integrated Care Records, and Card-Based Patient Records. EHR has become most widely used term, encompassing all reasonable notions of health records in electronic form.

The grand vision of EHR infrastructures is the interconnection and reusability of all recorded health information, regardless of where it is stored, so that all relevant health information can electronically flow to wherever it is needed. Nothing will become of this vision, however, unless critical privacy and security problems are overcome.

Studies reveal that most patients do not trust the administrators of national health services and other insiders in the health care system with the control over their personal health information. Their trust often does not extend beyond their own care providers, and for good reasons: the opportunities for privacy invasions due to secondary use of health record information are enormous. Organizations with a justified need (according to current widespread regulations) to access health information include government and private health plans, insurance companies, administrators, hospitals, doctors, pharmacies, employers, schools, researchers, data clearinghouses, accreditation and standard-setting organizations, laboratories, pharmaceutical companies, practice management system vendors, and billing agents. Other parties that may seek to obtain patient information include relatives, public health officials, drug marketers, public assistance programs, credit bureaus, and law enforcement agencies. Furthermore, health information custodians may enlist the services of lawyers, accountants, auditors, consultants, volunteers, and a variety of subcontractors (who may be in different continents with inferior privacy standards).

Privacy is also sought by medical practitioners. Many doctors do not like the idea of central parties (such as health insurance organizations) being able to monitor all their actions, since they feel this negatively impacts their autonomy; in many situations, they would prefer to be able to access information on the basis of their role rather than their identity, and they certainly do not want identifiable digital evidence of all their interactions with patients to automatically flow to third parties. Role-based access is also preferred by medical researchers, for accessing online disease registers and other medical databases.

With CMP, an EHR is simply a Local or Remote record, or the logical combination of several such records. Attribute Authorities are health professionals and possible other entities (including the patient himself) who add digitally authenticated statements to EHRs. Due to the unique features of CMP, EHRs can be securely managed by both the data subject and his health care professionals, in a manner that simultaneously protects the data subject's privacy interests, the professional's liability interests, and the legitimate interests of researchers and other third parties:

- Each patient can jointly manage his health information together with selected physicians. A record can be managed electronically as one logical entity, even though different parts may reside in different physical locations. Each party with access rights can be assured that the data entries on which it relies have been entered by authorized parties, through either role-based or identity-based digital signatures. In this manner, health care service providers can effectively maintain partial ownership of a data subject's health information.
- By providing patients with a single tamper-resistant smartcard, health care providers can maintain even greater control over their own contributions to a the patient's EHR, since the card can further limit the patient's ability to manipulate entries. Literally billions of authenticated EHR entries (possibly from different health professionals) can be securely managed using a single 8-bit smartcard. Cards can be issued to patients by a central entity that cannot compromise the legitimate privacy and security interests of patients and health care providers that ride along on its added security; amongst others, different applications using the same smartcard can be fire-walled through the application software running on the patient's trusted computer, rather than having to trust the smartcard issuer.
- At the same time, patients as well as health professionals are able to selectively disclose authenticated health data in anonymous or pseudonymous form (with or without certifications). Patients can also delegate the right to do so to their doctors (e.g., to over-ride protections in emergency situations) or to third parties (e.g., for research purposes) in the form of Delegation tokens.

Naturally, EHRs can be stored on patients' home computers, on portable devices (such as laptops, smartcards, or PDAs), or on a server on the Internet (possibly distributed across multiple trusted parties, and likely encrypted). The shift back towards user-involvement is already taking place, driven both by advances in information technology and by health regulations such as HIPAA. However, health regulations do not promote in any way the idea of transferring full ownership over health information to its data subject; they typically only provide data subjects with the right to review information and to withhold consent over certain secondary uses. Indeed, from the perspective of the health care professional, it is one thing for a patient to be able to view his own health information, in order to be informed and to check for errors and out-of-date information; it is quite another thing for the patient to be able to add, delete, modify, or prevent updating of arbitrary data in the EHR. With medical liabilities being as they are, and most patients not being professionally qualified to make informed modifications to health data, medical professionals are reluctant to rely on patient-owned health records; they need to be able to verify the credibility of the source(s) of the health data they rely on.

This brings us to the use of CMP Local records. In the context of EHRs, these would be called *Electronic Personal Health Records* (EPHRs). EPHRs allow patients to make all the relevant medical information available at the point of care, and give them even greater control over the privacy of their own health information. The idea of giving patients electronic control over their medical information is not new. However, most Internet-based EPHR systems currently are little more than the equivalent of what software like Quicken and Microsoft Money allows consumers to do with their financial data; they help patients track their medical conditions, allow them to review and print information about them, and so on.

Through this fine-grained multi-party rights management approach to electronic health information, health professionals can ensure that the data they rely on is of high quality. CMP in effect creates a continuum between health records maintained by health professionals and health records maintained by data subjects, seamlessly unifying the two approaches and covering the entire spectrum of possible rights management settings. In the CMP approach, the issue of where the health data resides hardly matters anymore; it is all about who has electronic access to which parts of the record.

Let us end with a digital prescription scenario, to show how far the functionality of CMP can take us:

- Consider a hypothetical patient, Bob, who is issued a prescription Digital Credential. The attributes in the Digital Credential are the prescriptions and Bob's name (or some other identifier). The medical doctor would have a Digital Credential that certifies him as a professional M.D. qualified to write prescriptions. With that Digital Credential he is able to write a prescription for 10 doses of penicillin to the prescription Digital Credential kept by Bob on his PDA. Password encryption prevents others from seeing the information in case his PDA is misplaced or stolen. For security and privacy reasons, the prescription Digital Credential would be implemented as a one-show Digital Credential.

- At a pharmacy, Bob presents his PDA with the prescription Digital Credential and discloses only that he is still eligible for a dose of penicillin prescribed by a qualified physician. He can do this without revealing his own name, his physician's name, and the number of valid doses remaining on his prescription Digital Credential; namely, by means of the Attribute Disclosure protocol Bob can prove to the pharmacy that the number of refills remaining is greater than zero without revealing more. The pharmacy gives Bob his doses and decrements the remaining number of refills by reissuing a new prescription Digital Credential that clones all the attributes in the previous one, decrementing the dose counter in doing so; hereto the pharmacy either acts as its own Attribute Authority or as an intermediary between Bob and a central authority. Bob is able to go to different pharmacies at different times; all that pharmacies can do is to confirm his eligibility and adjust the number of doses left.

- To combat fraud by the pharmacist, the pharmacies would be required to submit the information they received from patients to a central authority that would verify that all dispensed drugs were given to valid prescriptions. The approach would also combat fraud by Bob, while at the same time perfectly protecting his privacy if he does not misuse his prescription. Specifically, suppose that Bob hacked his PDA to make an exact copy of his original prescription Digital Credential which he already used, with the goal of filling his prescription more than the allowed number of times. Using CMP, this fraud can be prevented without the pharmacy needing to rely on a real-time on-line verification by an Attribute Status Provider: the one-show feature allows the pharmacy to fill each prescription on the spot, and to send prescription requests in batch for overnight verification by the Attribute Status Provider at a later stage. Furthermore, using the lending discouragement technique, Bob can be financially discouraged from giving his prescription Digital Credentials to a friend; in fact, by lending his Digital Credential, Bob would not only enable his friend to use any hidden attribute information within his prescription Digital Credential (e.g., his credit card information), but he would also expose himself to punishment if his friend uses up more than the agreed upon portion of doses. For greater security, Bob's Credential key could be locked up in a smartcard.

- If Bob lost his prescription Digital Credential, he would be able to use a backup copy made on his personal computer to return the prescription and be issued a new one for the remaining doses. In doing so he would be identifying himself. In those cases where identification is undesirable, the backup software could be allowed to make a copy of the prescription Digital Credential with enough information to allow it to be shown again. To help an absent-minded Bob not to over-fill his prescription if he loses his PDA and uses the backup copy, the software restoring the backup on a new PDA would warn Bob of that possibility as it restores its prescription. Obviously the backup and restoration process would be password-protected so only Bob would have access to the process. If Bob has no idea how many doses are left, his computer could pseudonymously retrieve the information over a secure internet connection from the central authority. In retrieving that information he would only need to remember the last time he filled a prescription so that he could confirm that the central authority had updated its data since then.

National Identity Chip-Cards

National identity chip-cards, as envisioned by many governments and already used in Hong Kong, poses grave threats to the privacy of citizens. A Gartner survey published in March 2002 found that *the public supports a national ID only for very specific purposes, and people are quite suspicious of what governmental agencies might do with it.* National identity chip-cards allow the actions of all card holders to be linked and traced automatically and in real time on the basis of their digital identity certificates and other uniquely identifying information, not only by the parties directly involved in verifying entitlements but also by a multitude of other parties that users may not even be aware of (and that organizations and other verifiers may find highly undesirable). Most envisioned card uses, however, do not require such systemic identification at all. Consider proving eligibility to access services, establishing whether a person has the right to work in the country, allowing people to prove their age when purchasing age-restricted items, and supporting telephone or on-line voting; with all of these, identification is necessary only, if at all, at registration time.

National identity chip-cards that rely on conventional security mechanisms (such as PKI or symmetric authentication) are not only privacy-invasive, they also do nothing to discourage participants from using each other's credentials. As well, they encourage large-scale identity fraud and other devastating abuses of security holes that are inevitably caused by heavily relying on the central storage of information. Furthermore, digital identity certificates cannot be implemented efficiently and securely in low-cost chips.

Using CMP as the underlying architecture, one can build a national "identity" chip-card system that simultaneously addressed the security needs of government and the legitimate privacy needs of individuals. Due to the unique smartcards features of CMP, the resulting system can be implemented more efficiently and more securely than with conventional cryptographic techniques.

E-Government

E-government refers to the electronic delivery of government services to citizens. The primary objective is to simplify the interaction with citizens and institutions. In the past two years, many governments around the world have established some degree of on-line presence. Among the leading countries to bring government online are the United Kingdom, Canada, and the United States.

Market analysts distinguish between five phases of e-government:

- providing information via Web sites;
- electronic service delivery;
- improving operations through Web interfaces and electronic data exchanges;
- moving toward more personalized electronic service delivery ("e-CRM"); and
- introducing Web-based collaborative technologies.

Implementing CRM initiatives is widely considered a key priority to provide personalized citizen self-service.

In most cases, government organizations will need to be able to securely make authentication and authorization decisions about citizens who request electronic access to their services. Liberty Alliance is already being viewed with increasing interest by e-government architects. Indeed, the considerations of government for managing identity-related information in part match those of industry. Many governments in fact have an even stronger interest in protecting the security and privacy of information of individuals and private sector organizations, and for good reasons. For instance, an August 2000 survey by Hart-Teeter about U.S. citizens' view of e-government services found that 53% of respondents were extremely concerned with the potential loss of privacy, and in a Gartner survey in 2001, nearly 70% of consumers cited privacy concerns as one reason that could make them stop using e-government services.

On the security side, progress is being made in the right direction. Indeed, according to the Giga Information Group in June 2002, *in some technologies, like smartcards, biometrics and electronic records management, the government is ahead of business.* Many governments are keen on access management systems based on smartcards, not only for citizens but also for its own employees and to replace driver's licenses, airport security documents, passports, and so on. On the privacy side, however, governments are struggling. Consumer outcry, trade group complaints, potential violation of privacy laws, and complaints by data protection commissioners have already lead to the suspension of several national PKI e-government initiatives.

Using CMP, it is not hard to see how security, scalability, privacy, and general performance requirements can be reconciled.

Digital Rights Management

Digital Rights Management (DRM) is generally defined as the collection of tools and technologies for protecting copyrights and other rights on digital media. DRM is an umbrella term: no single tool or technology suffices to guarantee access and content usage controls throughout a digital content distribution infrastructure.

DRM deals with authorization decisions about access to resources, and as such it is an application of access management. However, DRM places stronger requirements on fraud prevention than general access management. Namely, access management in general does not deal with long-lived access, while DRM also seeks to control usage by authorized users after they gained access to a resource.

All modern DRM systems have at their core the notion of a digital license, and most deal with content and licenses in a separate manner, along the following lines:

- Licenses are issued when access is requested, while content is made freely available in encrypted form to prevent access by unauthorized parties. To access protected content, the client must obtain a digital license that specifies how the content may be used.
- To consume protected content, the client connects to a clearing house and requests a digital license for the content. The request requires the client to send a unique identifier that identifies him and/or a specific client device that will play the content. The request is typically initiated by the client's software application or hardware device upon the client's first attempted access.
- Assuming the clearing house makes a favourable authorization decision for the client, it sends the requested digital license to the client. The client's device or application, which is presumed to be secure against tampering, then decrypts the license and displays or otherwise makes available the content to its user in accordance with the usage rules.

By separating content from licenses, content providers can issue one license for multiple sources of content, can issue different licenses for the same content, and can support business models that cleanly separate the interests of copyright holders, content distributors, service provider networks, and others.

Of course, this basic DRM architecture inherits all the security, privacy, and performance problems of general cross-domain access management. Several authors have already noted the unique security and privacy benefits that Digital Credentials could bring to DRM; see, for instance [2]. By building DRM on top of CMP, consumers can control and limit the correlations that content distributors can establish about their consuming habits and identity, and content distributors can protect their intellectual properties more securely without infringing fair use rights.

Let us walk through a simple CMP-based DRM scenario:

- Bob visits MusicPortal, an Internet portal where the latest album of his favorite band is available via download. MusicPortal groups content distributors together so that customers can buy their music from a single point-of-sale. The portal offers various subscription packages, such as monthly fees for unlimited downloads, prepaid number of downloads, and so on. It also offers Web site personalization and can make recommendations to Bob by keeping track of his musical preferences.
- Bob chooses to purchase a subscription which entitles him to a limited number of downloads every month for a fixed monthly fee. He pays for the subscription with his credit card in a special section of the portal. To protect his privacy, his subscription is delivered in the form of a Digital Credential. This ensures that the subscription cannot be forged, while at the same time the portal will not be able to trace which credit card was used to buy the subscription.
- After making his music selection and the usage rights he wishes to acquire, Bob goes to the checkout section of the portal. To acquire the rights on a specific album Bob presents his subscription to the portal. The portal processes the payment through the clearing house and in exchange emits a digital license describing the rights and privileges associated with the music file. The music file is encrypted specifically for Bob using an encryption key that can be found in the digital license. The digital license is packaged into a Digital Credential as well, to provide lending protection and possibly other protections.
- To play the music, Bob needs a player that understands and enforces the license. Bob's player can permit him to copy the file from one player to another, so that he can play the file from many places. A lending disincentive placed in the licenses (e.g. the credit card information that Bob used to purchase the subscription) would strongly discourage Bob from copying the music to his friends even if he could bypass the hardware protections of his player.
- For extra security, all subscriptions and digital licenses could be managed using a simple 8-bit smartcard, by off-loading all expensive computations and storage to the user's PC, a laptop, or PDA, while preserving all the smartcard's security benefits. Multiple license issuers could

[5] Even digital certificates that do not explicitly specify the identity of their holder can be traced in a trivial manner, because each digital certificate is a unique bit string that at least its issuer has seen as issuing time; digital certificates in this respect offer no more privacy than Social Security numbers, credit card numbers, and health registration numbers.

all ride along on the security of the same card, without needing to trust each other.

Why PKI Breaks Down In Cross-Domain Contexts

We already gave an overview of why PKI quickly breaks down in the context of cross-domain access management. The following list describes the problems in more detail:

- Requests by access providers to access databases held outside their own domain may be dishonored for any reason, may be expensive, or may simply not be an option because of the absence of a network connection.
- Even if the access requests of access providers to back-end databases in other domains are accommodated, the response time may be very high or the response information itself may be incorrect.
- Access providers in turn may erroneously reject access requests from identity certificate holders on the basis of false or irrelevant data in back-end databases, or simply because the online connection to the databases in other administrative domains fails (due to peak load, a natural disaster, or otherwise).
- In order to enable access providers to access databases in other administrative domains, the databases must be made accessible on-line via public networks. This renders these databases more vulnerable against intrusions by hackers and insiders.
- The information that provides the basis for authorization decisions is typically personal information of the access requestor. To the extent that legislation requires organizations that handle personal information to adhere to privacy standards, organizations incur legal liability by sharing their databases with outsiders.
- Each access request can readily be traced back to its originator, simply by verifying to whom the disclosed identity certificate was issued.[5] In this manner, the access requests of a party can all be traced and linked instantaneously and automatically. This enormous surveillance capability is enjoyed not only by access providers, but also by the providers of the back-end databases that contain authorization information (queries from access providers are accompanied by access requestor identifiers), by certificate issuers, by revocation authorities, and possibly by outsiders such as wiretappers and third-party service providers.
- Similarly, central parties can learn in real time the identities of all the customers of access providers, and the providers of back-end database services may learn specific details about the nature of customer requests. Information such as customer identities and peak hours is typically considered sensitive corporate intelligence that for competitive reasons should remain confidential.
- Any access request, be it by an access requestor or from an access provider to a database service provider, is a digitally self-signed statement that cannot be repudiated.
- The reliance of identity certificates across many systems exposes the issuers of identity certificates to potentially unlimited liability, as it invites abuse through identity theft.
- If the secret key of a digital identity certificate is generated and stored on a personal computer or the like, it is virtually impossible to prevent its compromise, loss, disclosure, modification, or unauthorized use (e.g., lending or copying certificates). The integration of smartcards or other tamper-resistant form factors, however, exacerbates the privacy problem (cards can overtly or covertly leak any information they hold) and may not be practical due to the computational complexity of public-key cryptography.

Note that the majority of these drawbacks apply not only to PKI, but to any security mechanism for access management that relies on unique identifiers pointing into back-end databases. In particular, symmetric authentication, password-only authentication, and end-to-end biometrics all suffer from virtually the same problems, and add new ones of their own.

Attribute certificates, such as those specified by X.509, are not a solution either. True, all the information that access providers need to know to make authorization decisions can be stored within them; in this manner, access providers can make authorization decisions without having to consult databases outside their domain (other than, possibly, revocation databases). For security reasons, however, any attribute certificates supplied by an access requestor must contain an embedded link to a digital identity certificate of the access requestor. This identity certificate must be sent along with the access request. Without this, access requestors could pool together their access privileges to gain access to services that they do not individually have access to; also, nothing would stop or discourage access requestors from lending or copying their access privileges. As a consequence of this marriage of attribute and identity certificates, attribute certificates inherit almost all the problems of PKI. In fact, attribute certificates introduce new problems that may be more serious than the bottleneck problem they are intended to solve. Storing all of a person's attributes in a single attribute certificate quickly leads to a privacy nightmare, because all the attributes are systematically revealed each time when showing the certificate. The alternative of distributing the attributes across multiple certificates results in each user having to carry an enormous number of certificates (in particular in the case of fine-grained access management) that all have to be managed separately.

Closing Remarks

With the increased sharing of resources across applications, the boundaries between the enterprise, government, and consumer spaces will increasingly blur. In the next five years, all three spaces will develop highly similar needs for access management solutions: strong security based on public-key cryptography; strong privacy (for personal data as well as corporate data, the latter with an eye on competitive or national intelligence concerns); dynamic scalability across disparate systems; and, cost-effective implementations in low-cost portable devices. Some of the trends that drive and unify these needs include the following:

- *Smartcards:* At present, smartcards are far from being the norm for access management (with the exception of national security applications). Nevertheless, a steady growth of smartcards is expected, especially in enterprise security (for the purposes of both physical access to premises and network access) and in the public sector. The consumer space may see widespread adoption of smartcards as well in the next five years. In a recent white paper on smartcards, Microsoft states: *The smartcard will become an integral part of the Windows platform because smartcards provide new and desirable features as revolutionary to the computer industry as the introduction of the mouse or CD.*

- *Peer-to-peer:* Peer-to-peer file technologies will become increasingly useful in organizational environments. Peer-to-peer applications such as instant messaging and file sharing are highly useful, particularly in collaborative enterprise environments. By providing decentralized information storage, they circumvent scalability problems, make it easier to dynamically add and remove resources, and minimize the required set-up. At present, security concerns hamper the adoption of peer-to-peer solutions in open environments. Since users in a peer-to-peer application hold and manage their own information, corporate firewalls and other perimeter security measures are useless.

- *Wireless Mobile Devices:* Market analysts estimate that by 2006 almost half of all professional PCs will have wireless support. Rapidly increasing numbers of Bluetooth chips (a short-range radio technology for low-cost devices) are shipping, and major corporations have begun to introduce mobile handsets that operate at many megabits per second. Wireless mobile devices are expected to flourish in the corporate enterprise space. A survey by Evans Data Corporation in 2002 of over 700 database specialists across North-America found that almost half of respondents are either developing database applications that support wireless or handheld devices or plan to do so in the near term. In the longer term, wireless mobile devices are expected to represent their holders in all kinds of interactions, in both business and personal situations.

References

[1] *Access Management based on Digital Credentials (Part I)*, S. Brands, Information Security Bulletin, Volume 8, Issue 10 (December 2003), pp 369-379

[2] http://www.w3.org/2000/12/drm-ws/pp/hp-poorvi2.html

About The Author

Dr. Stefan Brands is an expert on the subjects of electronic authentication, digital identity management, payment systems, and information privacy. He is an adjunct professor in cryptography at McGill University in Montreal, and is affiliated with Credentica, an IT security company providing access management software. Previously, he was a principal researcher at DigiCash (1996-1998) and at Zeroknowledge Systems (2000-2001). Dr. Brands conducted his PhD research from 1992 until early 1996 at the *Center for Mathematics and Computer Science* (CWI) in Amsterdam. His dissertation was approved by professors Adi Shamir, Ron Rivest, and Claus Schnorr, and was published in 2000 by MIT Press with a foreword by Rivest. Dr. Brands is a member of the CSIS Working Group on Authentication in Washington, provides consultancy from time to time, and is a frequently invited speaker at leading industry and academic forums.

He can be reached at brands@cs.mcgill.ca or at brands@credentica.com.