Access Management in Critical Information Infrastructures

Dr. Stefan Brands
Credentica, Inc.
Montreal, Quebec (Canada)
E-mail: brands@credentica.com
April 2003

Abstract: This paper examines two different approaches towards secure electronic access management in critical information infrastructures. The first approach, X.509-style PKI, is widely believed to be the strongest solution available in this context, but nothing could be further from the truth. X.509-style PKI was invented in 1978 for the purpose of facilitating message encryption, with entity authentication serving to prevent man-in-the-middle attacks. Authentication as used for message encryption, however, is a far cry from managing the access of authorized parties to resources, with all its privacy, security, and performance sensitivities. We show how X.509-style PKI fails in the context of critical information infrastructures, and present a superior approach based on Digital Credentials.

Technical level: 2 (on a scale of 1 to 5).

See also the accompanying PowerPoint presentation.

Access Management in Critical Information Infrastructures

Dr. Stefan Brands
Credentica, Inc.
Montreal, Quebec (Canada)
E-mail: brands@credentica.com
April 2003

1. INTRODUCTION

Critical information infrastructures are information-centric infrastructures that are considered essential to the defense and economic prosperity of a society, and to the well-being of its people. Telecommunications, the supply of utilities, banking and finance networks, public transportation, national defense, and other critical information infrastructures all crucially rely on the ability to effectively manage and share sensitive information. In order to facilitate the sharing of information across traditional organizational boundaries, critical information infrastructures are increasingly being made accessible over open networks. This transition enhances productivity, minimizes costs, reduces errors, and opens up a range of new opportunities.

To protect information that is shared in electronic form, adequate security mechanisms are needed that limit access to authorized parties. Firewalls, anti-virus software, intrusion detection applications, and vulnerability assessment products and services provide little security in this context: critical information infrastructures typically involve many interactions between individuals operating in different administrative domains. With trust domains becoming logical rather than physical, security must be tied directly to the information itself instead of to the perimeter of its repository. Indeed, Wedbush Morgan Securities in a February 2002 industry report state that "the government and especially the military are beginning to show a strong preference for conducting business with companies that have implemented access management solutions and strong authentication."

By far the best way to provide secure access management in critical information infrastructures is the application of modern cryptography, preferably implemented in tamper-resistant user devices such as smart cards (for two-factor authentication) and possibly augmented with biometrics (for three-factor authentication). Through the application of cryptography, information can be protected on the basis of secret keys that (contrary to simple passwords) never leave the physical confines of devices held by their legitimate users. The trend towards the use of strong cryptography in critical information infrastructures is undeniable. According to the U.S. Federal Chief Information Officers Council in June 2001, there is a "fundamental need for strong authentication over a wide range of applications that use electronic transactions. […] public key technology meets this need better than any other single technology […] Federal agency use of public key technology is growing quickly both vertically (within organizations) and horizontally (across organizations)."

Authentication and authorization are not the same tasks, however. Indeed, as Giga Information Group observe in a report in June 2002: "Authentication — generally by user name and password for citizens, but increasingly by digital certificates combined with smart cards and/or biometrics for employees and government contractors — is the current visible battle. However, the real war is authorization through the use of directories that describe what access privileges an authenticated user will have to which systems and databases."

This paper examines two different approaches towards secure access management in critical information infrastructures. The first approach, X.509-style PKI, is widely believed to be the strongest solution available in this context, but nothing could be further from the truth. We show how X.509-style PKI fails in the context of critical information infrastructures, and present a superior approach based on Digital Credentials.

In: Proceedings of the 15th annual Canadian Information Technology Security Symposium, Ottawa (Canada), May 12 -15, 2003.

2. WHY X.509-STYLE PKI IS A POOR SOLUTION

A Public Key Infrastructure (PKI) is an information security infrastructure that revolves around the distribution and management of public keys and digital identity certificates (which bind an individual's public key to his identity). Digital identity certificates were invented in 1978 for the purpose of facilitating message encryption, and are at the heart of the X.509 digital certificate framework. Today's leading PKI implementations are all based on the X.509 standard. In line with the original objective of digital identity certificates, these implementations provide confidentiality of data in transit (through encryption), user authentication (to ensure that a message is encrypted under the right public key), data integrity (to prevent tampering with data in transit), and non-repudiation (proof of the sender's identity).

Now that it is becoming clear that message encryption is not where the information security market needs are heading, PKI vendors are reinventing themselves to address the problem of access management. For example, the industry leader in PKI software, Entrust, in its April 2001 strategy document states: "Authorization capabilities are emerging as the next major security requirement. […] A company not only needs to know that a customer is in fact who they say they are. It also needs to control what resources that customer has access to, what factors affect availability of the resources to that customer, and what auditing is required to ensure the non-repudiation of transactions with that customer."

Access control was never on the list of design requirements for digital identity certificates, however, because it is of little relevance in the context of creating a secure infrastructure for sending around encrypted messages. As a consequence, the application of X.509-style PKI is currently being distorted and stretched by the PKI industry into the realm of access management. By requiring individuals to provide their digital identity certificates whenever they request access, access providers can look up any information they want about them to make informed authorization decisions.

There is only so much distorting and stretching that can be done, however. Authentication as used for message encryption is a very far cry from controlling the access of authorized parties to information, with all its security, scalability, and privacy sensitivities. Applying PKI technology to the problem of electronic access management in increasingly open environments is like trying to propel a passenger plane using a steam engine; it simply will not fly. When used across multiple trust/administrative domains, the approach of using X.509-style certificates as strongly authenticated pointers to databases entries fundamentally provides poor security, suffers from bad performance, and leads to inescapable systemic identification:

- **Poor performance in smart cards:** Portable tamper-resistant devices, such as smart cards, are deemed crucial in an increasing number of critical information infrastructures, to protect against the compromise, loss, disclosure, modification, and unauthorized use of secret keys. However, in the words of the Aberdeen Group in an executive white paper of January 2001, "CPU drain prevents PKI from being a solo building block." Indeed, the computational requirements of processing an X.509-style certificate are well beyond today's popular smart cards and other low-cost computing devices. Addressing this problem by adding advanced circuitry (such as cryptographic co-processors) seriously increases the price of these devices. Moreover, the addition of sophisticated circuitry can easily lead to new weaknesses in the internal defense mechanisms, and adversely affects reliability. These problems worsen in the context of multi-application smart cards, which are believed to be the way to go for critical information infrastructures.

- **Unconditional trust required in smart cards:** Using X.509-style PKI in combination with smart cards (or other tamper-resistant devices) is also problematic from the viewpoint of security. The providers of critical information infrastructure services must place very strong trust in all the parties involved in the card manufacturing, masking, initialization, application loading, and personalization process. It is almost impossible to verify that none of the cards have been programmed to secretly leak keys, card identifiers, access control codes, data from other applications running on the same device, or other confidential data.[1] Also, it is almost impossible to guarantee the uniqueness,

[1] Information may be leaked via electromagnetic emanations or radio signals (a particular problem in the case of contact-less cards), by changing message formats in communication protocols (stop bits can be flipped, message fields can be set to values outside of specified ranges, and so on), by timing the delay before transmitting data, by causing response messages to be incorrect (e.g., it cannot be detected whether certain bits of a MAC have been flipped), or by encoding information in any data that may be chosen by the card.

randomness, and secrecy of the secret keys stored in the cards (e.g., rogue insiders could corrupt the process for generating randomness within cards, or keep track of seed values stored in memory). Furthermore, a variety of fake-terminal attacks become possible due to the lack of a trusted display and keyboard on the user's side: the verifier's terminal can store any passwords and other information provided to it, and can wrongly display information on its display to trick the user into accepting a transaction under modified conditions. National defense networks and other critical information infrastructures cannot reasonably place such trust in outsiders, but obtaining all the cards and terminals from highly trustworthy sources may not be feasible or cost-effective.

- **Access right cloning and lending:** Since most of the information exchanges in critical information infrastructures take place on the basis of the roles and privileges of those who need access, the need to prevent access right lending and cloning of privileges is of utmost importance. Lives could be lost due to attacks by terrorists, organized crime, foreign governments, insiders, and hackers. However, X.509-style PKI does not provide software-only protection to discourage certificate holders from providing (copies of) their access rights to other parties: a user's secret key is simply a random number, and so revealing it to someone else has no direct negative consequences for the legitimate certificate holder. Code transformation methods raise the barrier against reverse-engineering of the client software, but in critical information infrastructures it is too dangerous to rely exclusively on software obfuscation methods to deter unauthorized copying and lending. Locking the secret keys of certificates inside tamper-resistant devices does a much better job of raising the barrier, but introduces security and performance problems (as discussed above) and offers insufficient protection in the absence of a cryptographic solution; low-cost user-held devices cannot provide the kind of security that strong cryptography provides.

- **Denial-of-service attacks:** Critical information infrastructures relying on X.509-style access management are highly vulnerable to electronic and physical denial-of-service attacks, because of their strong reliance on the real-time availability of central parties (for providing online database access and for providing online certificate status validation). Notably, the approach of relying on on-line central databases that can be consulted by all access providers pushes the door wide open to devastating abuses of security holes. As the number of access points and the number of individuals with network access increase, it is becoming increasingly difficult to protect online databases against misuse by hackers and insiders. Indeed, a summer 2002 survey by Evans Data Corporation revealed that 20 percent of the 700 database specialists surveyed had experienced a direct breach in their database security, and a CSI/FBI computer crime survey conducted in 2000 found that 71 percent of unauthorized break-ins are by insiders.

- **Non-scalable:** The approach of using an X.509 certificate as an authenticated pointer does not scale beyond pre-established trust domains, since the actual authorization decisions are left to the access provider. When information is shared amongst different administrative domains, the parties that have to make authorization decisions may not be able to obtain all the up-to-date information they need for making an authorization decision: the missing data may reside in databases outside of their control or be otherwise unavailable, and it is difficult to guarantee its completeness and correctness. With increasing numbers of individuals and organizations seeking to share resources, it becomes infeasible to guarantee the availability and correctness of the data needed to make authorization decisions, even when switching to low-grained role-based access control.[2]

[2] A recent update of the X.509 standard describes the use of attribute certificates; these can specify attributes other than the identity of the key holder, such as access rights, capabilities, and preferences. More generally, the certificate of an access requestor could include all the attributes that the access provider needs to know in order to locally decide whether or not to grant access. It is not possible, however, to issue attribute certificates simply by replacing the identities in X.509 digital certificates by whatever attributes are deemed relevant, since that would enable certificate holders to copy, lend, discard, and pool their access privileges. For this reason, an X.509v3 "attribute certificate" is merely a digitally signed message that specifies the attributes and contains an embedded link to a standard X.509 identity certificate; it does not replace the X.509 identity certificate to which it points, but must be used in combination with it. As a consequence, most of the problems of identity certificates remain the same or become worse. In fact, X.509-style attribute certificate may not even improve scalability: to facilitate fine-grained control over which attributes are released to whom, users must carry a great many certificates that all have to be managed separately.

- **Inescapable systemic traceability:** With X.509-style PKI, the real name of the access requestor is systematically disclosed. The fundamental problem is that public keys are globally unique identification numbers (much like strongly authenticated super-SSNs) that unavoidably travel along with each and every action taken. Thus, PKI technology forces everyone to transact on the basis of unique identifiers, which are passed on not only to the intended access provider but (inherently to the PKI mechanism) also to third parties who have no business in knowing the details of the transaction (such as the providers of authorization databases and online certificate status verifiers). In other words, PKI technology roots inescapable systemic identification deeply into the information infrastructure. This enables terrorists, organized crime, foreign and domestic governments, and others who actively monitor the network (possibly with the help of insiders) to find out who is communicating with whom about what, and who is accessing which resources for what purpose. In an attempt to hide the huge privacy problems created by X.509-style PKI, the PKI vendors misleadingly define privacy as "communications are safe from eavesdropping." This works in the context of sending a message to an intended recipient, but in the context of access management, encryption has very little to do with privacy. Indeed, according to the Aberdeen Group in their white paper of January 2001, "as currently sold and implemented, PKI […] eliminates even the pretense of ensuring user privacy." Identity certificates that specify a "pseudonym" instead of a real name are not a valid solution, since they can be linked and traced as easily and in the same manner as identity certificates, on the basis of their public keys; indeed, few people would agree that their SSN is a pseudonym. For the same reason, role-based access through PKI (whereby the name field of the certificate specifies a role rather than its holder's identity) does not allow certificate holders to escape systemic identification.

- **Identity theft:** Technologies that root inescapable systemic identification deep into a critical information infrastructure can easily give rise to massive identity theft, whereby fraudsters assume the identities of their victims. According to the U.S. Federal Trade Commission (FTC), identity theft is the fastest growing crime in America, affecting approximately 900,000 new victims in 2001. The FTC expects that its cost will reach USD 8 billion by the year 2005. A recent study by the U.K. Department of Trade and Industry warns that in the not–too–distant future criminals will be as interested in stealing victims' identities as they are in stealing possessions. Notwithstanding the fact that X.509-style PKI provides for message encryption, it seriously increases the risk of identity theft, since its fundamental premise is that of inescapable system-wide identification. The problem is compounded by the strong reliance on central databases, with all their vulnerabilities.

- **Managed services are intrusive:** Managing PKI-based software requires specialized knowledge, a scarce resource that many companies would prefer to outsource. With an increasing number of incompatible authentication mechanisms available, and network identities becoming distributed ("federated") instead of centrally stored, access providers that need to make authority decisions will increasingly ask trusted authorities to verify the certificates presented by their clients. At present there are relatively few takers of managed security services. A survey released April 2002 by the McAfee security division of Network Associates showed that firms are holding back from outsourcing security primarily due to a strong reluctance to trust a third party. Indeed, there are serious reasons not to trust today's managed PKI services: the providers of online certificate validation services learn in real time the identities of their clients' customers, their peak hours, and other competitive information. Furthermore, Certificate Authorities must know the identities and any other attributes that go into the digital certificates they issue.

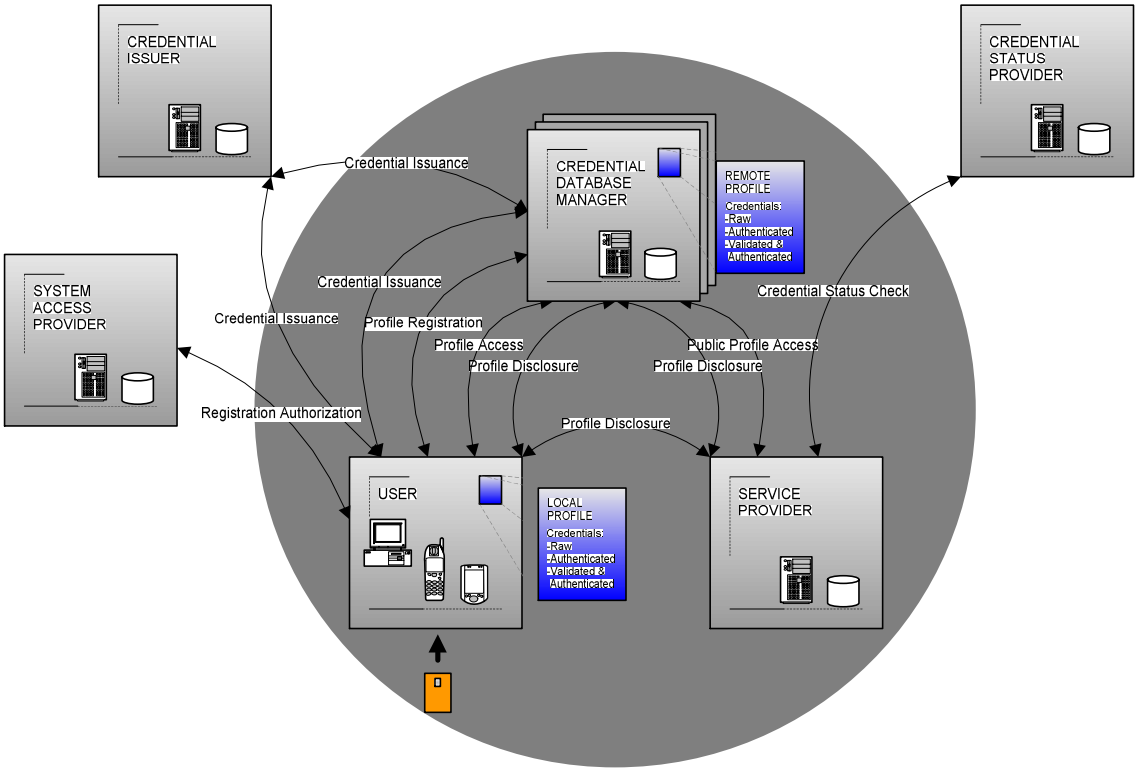
- **Violation of data protection laws:** In response to the growing security and privacy concerns, many governments have enacted data protection legislation that places stringent requirements on use, retention, and disclosure of information. Organizations that fail to comply may run into serious legal problems, ranging from fines to operational suspension. According to Wedbush Morgan Securities in their February 2002 industry report, "the need to comply with these government regulations and preferences will be a steady and driving force in the adoption of access management technologies." X.509-style PKI, however, is a poor match with fair information practices, since everything is systematically identifiable. In fact, it is quite possible that the unbridled use of PKI will be found unconstitutional when challenged in court. Some precedents: the Hungarian Constitutional Court in

1991 decided that multi-use personal identification numbers violate the constitutional right of privacy, the Portuguese Constitution states that "Citizens shall not be given an all-purpose national identity number," and SSN legislation in many countries prohibits the use of SSNs beyond very specific purposes.

In sum, X.509-style PKI is a poor solution for access control beyond a single administrative domain.

3. A SUPERIOR SOLUTION

Credentica is currently building its Credential Management Platform (CMP), a set of server and client components that provide authentication and authorization services that do not suffer from the abovementioned shortcomings. The figure below provides a schematic overview.



CMP holistically overcomes the shortcomings of access management systems based on X.509-style PKI:

- **Information can reside anywhere:** CMP allows for sensitive information to be held both locally and remotely. CMP also supports federation of remotely stored information. A record can be managed electronically as one logical entity, even though different parts may reside in different physical locations; those with legitimate access rights might not even realize the dispersed nature of the data they see. CMP facilitates automated sharing and synchronization of sensitive information across local and remote storage locations in accordance with application-specific rule sets. CMP also supports roaming.

- **Fine-grained multi-party ownership management:** CMP allows data records to be securely managed by multiple parties, in a manner that simultaneously protects their security interests. Different data "owners" can vouch for the authenticity of record entries by digitally authenticating them (through either role-based or identity-based digital signatures). 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) physically controlling the storage of a record can modify, delete, or add to it without authorization. Furthermore, access rights can be delegated to others (for example to over-ride protections in case of temporary absence or emergency situations) by means of delegation authorizations.

- **Negotiable privacy:** CMP accommodates for fully adaptable levels of privacy ranging from user-driven anonymity to mandated identification. In particular, CMP allows for pseudonymous access as well as role-based access (both server-driven for improved scalability and client-driven for privacy). CMP provides multiple protocols for gaining access to credential information, with varying levels of active participation from database managers and those who are deemed to "own" the information. CMP provides rule-based trust negotiation for the automated exchange of information, ensuring that only the minimum information needed to meet the requirements of the access provider is disclosed. In particular, personal information can be selectively disclosed in a manner that does not enable identification.

- **Limited-use access rights and credentials:** CMP provides for the issuance and management of limited-use access rights, and anyone can issue access credentials that are valid only a limited number of times. A built-in identifier, value token, or self-signed fraud confession will be exposed if and only the credential is shown more times than allowed, even for anonymous, pseudonymous, or role-based access. This functionality is particularly useful for delegation authorizations.

- **Strong security:** CMP offers data integrity, confidentiality of data at rest as well as data in transit (through encryption), secure time-stamping, secure audit trails, and digital receipts. It supports authentication strengths ranging from weak to military-grade two-factor and three-factor security. Organizations can strongly discourage credentials holders from lending or cloning their access rights (even for non-identifiable access) by embedding disincentives that will be disclosed if and only if the legitimate holder commits a fraud.

- **Efficient smart card implementations:** CMP allows for literally billions of digitally certified record entries (which may come from different parties who do not trust each other) to be securely managed using a single 8-bit smart card. The storage and computational burden for the smart card can be off-loaded almost entirely to a user-controlled device (such as a PC, a laptop, or a PDA), while preserving all the smart card's security benefits.

- **Secure multi-application smart cards:** Smart cards can be used as multi-application devices, without introducing any of the privacy and security problems caused by other technologies. Specifically, different application providers can all share the same secret key stored in a user's smart card to derive the security benefits of that smart card. The certificates will have uncorrelated secret keys which cannot be determined by anyone including the smart card supplier, and all certificates can be revoked separately. The application software on the user's trusted computer ensures that smart cards cannot leak or learn data they are not supposed to, that different applications using the same smart card remain fire-walled, and that the card issuer (or anyone else involved in the card manufacturing, masking, code loading, personalization, and issuing process) cannot compromise the legitimate privacy and security interests of parties that ride along on its added security. Smart cards are prevented from covertly leaking any information to the outside world, and outside parties cannot covertly send information to the cards.

- **Managed Services:** CMP enables outsourcing of all security functions to specialized third parties, in a privacy-preserving manner. Specifically, specialized Credential Issuers can digitally certify information on behalf of others without being able to learn attribute data that they have no business in knowing. Likewise, Credential Status Providers can validate certificates without being able to learn the identities of access requestors and access providers. In this manner, access providers can outsource core tasks related to digital authentication and authorization to security specialists, without having to provide them with sensitive information. Even the role of the tamper-resistant smart card can be outsourced, thereby circumventing the logistical problem of securely distributing tamper-resistant devices and making it impossible for card holders to bypass the security of their

cards; while this comes at the expense of involving the managed card service provider in the authorization of all transactions, the service provider will not be able to learn anything about the transaction other than the time it takes place.

- **Multi-purpose and multi-application digital certificates:** A single digital certificate can contain many attributes, while providing the certificate holder at all times with full control over which attributes are disclosed in a given interaction. This is analogous to having a paper certificate specifying a variety of attributes, and using a marker to cross out any attribute data one does not want to disclose. In contrast to paper-based certificates, after crossing out attributes on a digital certificate in one interaction they can still be disclosed when using the certificate another time. Also, certificate holders have much finer-grained control over the kinds of attribute data they can cross out than what can be done with a paper-based certificate and a marker.[3] Furthermore, a digital certificate can be presented for recertification or for updating to a Credential Issuer, without enabling the Credential Issuer to learn the current attributes in the digital certificate. Also, multiple Credential Issuers can certify attributes within the same Digital Credential without needing to know all of the attributes; this is particularly useful for multi-application certificates.

- **Peer-to-peer support:** Organizations can securely give individuals control over some or all of their own credential information by allowing them to store and manage the information locally on their own computer. This information is cryptographically protected to ensure that unauthorized users cannot modify, discard, lend, pool, or prevent the updating of information they hold. In the extreme, an organization can do away with its central databases, by securely distributing each database entry to the individual to whom it pertains. By basing authorization decisions directly on authenticated attributes shown by the requestor himself, trust can be established locally on first contact, with no need to consult on-line databases. This peer-to-peer approach provides a superior alternative from the perspective of administrative complexity: it circumvents scalability problems, makes it easier to dynamically add and remove resources, and minimizes the required set-up.

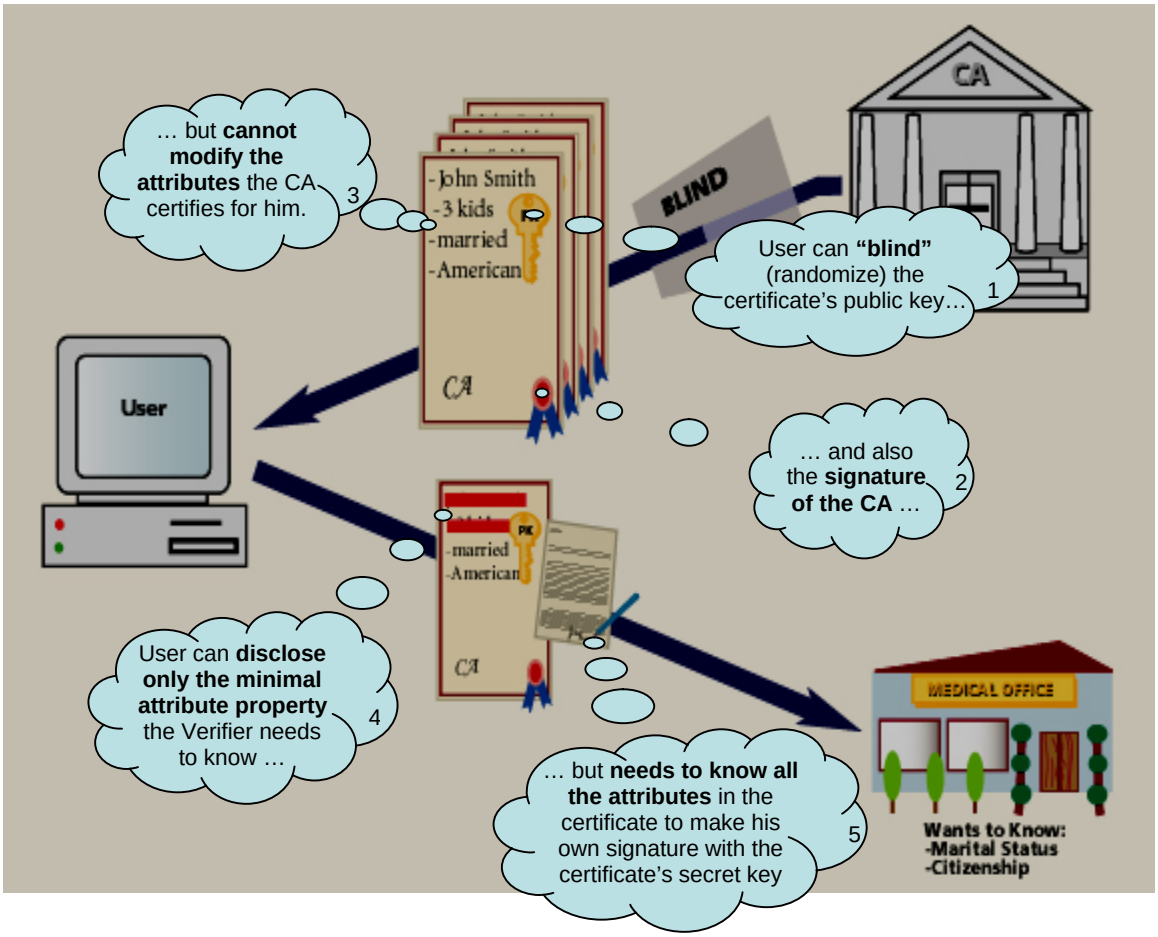
To achieve these properties, CMP leverages the unique powers of so-called Digital Credentials, a technology that has been designed specifically for the purpose of electronic access management. Digital Credentials are basic cryptographic constructs. They are much like digital identity certificates and digital attribute certificates, but offer a multitude of security, privacy, and performance benefits. As with an X.509-style digital certificate, two basic stages can be discerned in the life cycle of a Digital Credential: an issuing protocol and a showing protocol. The characteristics of both protocols are different, however:

- **(Issuing protocol)** In the issuing protocol, an individual obtains from a Credential Issuer a public key (for which the individual knows a secret key) and a digital signature of the Credential Issuer. The digital signature binds the individual's public key to one or more attributes. The whole package is called a Digital Credential. Although the sequences of zeroes and ones that make up the individual's public key and the Credential Issuer's signature are unique for each Digital Credential, the Credential Issuer cannot learn who obtains which sequences. At the same time, the individual cannot prevent the Credential Issuer from encoding the attributes into the Digital Credential, or more precisely, into his or her secret key.

- **(Showing protocol)** To show a Digital Credential to a verifier, the Digital Credential holder sends the public key and the Credential Issuer's digital signature to the verifier, and uses the secret key to authenticate a message. The message must include a nonce (such as a number unique to the verifier concatenated to a random number chosen by the verifier, an indication of time and date, or a sequence number); this prevents the verifier or a wire-tapper from "replaying" the Digital Credential, since in each showing protocol execution a new nonce must be signed, which requires knowledge of the secret key of the Digital Credential. At the same time, the Digital Credential holder is able to selectively disclose to the verifier any property of the attributes in the Digital Credential, while hiding all other information about the attributes. To convince the verifier that the

[3] To be technically precise, one can demonstrate attribute properties that combine linear relations by zero or more of the logical operators AND, OR, and NOT, with any other information about the attributes remaining unconditionally hidden. One can also prove that attributes are contained in a specific interval of set of intervals, without disclosing more.

claimed property is true, the signature on the verifier's nonce doubles up as a proof of correctness. In order to make the signature, the Digital Credential holder must know *all* the attributes in the Digital Credential, including those that are not disclosed to the verifier.



By carefully building on these four basic properties, it is possible to design Digital Credentials that are much more secure than X.509-style certificates, scale seamlessly across multiple trust domains, and offer fully adaptable levels of privacy. The interested reader is referred for full details to a book published by MIT Press [1], and for non-technical and technical overviews to [2] and [3], respectively. A more detailed discussion of CMP appears in [4].

4. REFERENCES

[1] "*Rethinking Public Key Infrastructures and Digital Certificates; Building in Privacy*," Stefan Brands, MIT Press, ISBN 0-262-02491-8, August 2000. With a foreword by prof. Ronald L. Rivest. Excerpts and endorsements available from http://www.credentica.com/technology/book.html

[2] "*Towards Digital Credentials*," Stefan Brands, ERCIM News #49, April 2002. Available for download from http://www.ercim.org/publication/Ercim_News/enw49/brands.html

[3] "*A Technical Overview of Digital Credentials*," Stefan Brands, invited publication to the International Journal on Information Security, to appear in the Q2 issue of 2003. Available for download from http://www.credentica.com/technology/overview.pdf

[4] "*Digital Identity Management based on Digital Credentials*," Stefan Brands and Frédéric Légaré, proceedings of the first workshop on Credential-Based Access Control In Open, Interoperable IT-Systems, Dortmund (Germany), October 2nd, 2002. Available for download from http://ls6-www.informatik.uni-dortmund.de/issi/cred_ws/papers/brands.pdf