Medical Information Privacy Assurance: Cryptographic and System Aspects

Giuseppe Ateniese, Reza Curtmola, Breno de Medeiros, and Darren Davis

Department of Computer Science, The Johns Hopkins University
3400 North Charles Street, Baltimore, MD 21218, USA
{ateniese,crix,breno}@cs.jhu.edu, dd@jhu.edu

Abstract. It may be argued that medical information systems are subject to the same type of threats and compromises that plague general information systems, and that it does not require special attention from a research viewpoint. The firsthand experience of experts in information security and assurance who studied or worked with health applications has been of a different sort: While general principles of security still apply in the medical information field, a number of unique characteristics of the health care business environment suggest a more tailored approach. In this paper we describe some recent results of an on-going research on medical information privacy carried out at the Johns Hopkins University under the support of the National Science Foundation (NSF).

1 Introduction

Electronically managed information touches almost every aspect of daily life in modern society. This rising tide of important yet unsecured electronic data leaves us increasingly vulnerable to curious neighbors, industrial spies, rogue nations, organized crime, and terrorist organizations. The basic communication infrastructure of our society is becoming less secure, even as we use it for increasingly vital purposes. Cryptographic techniques more and more frequently will become the only viable approach to assuring the privacy and safety of sensitive information as these trends continue. Highly secure encryption can be deployed relatively cheaply, and it is widely believed that encryption will be broadly adopted and embedded in most electronic communications products and applications that handle potentially valuable data. Applications of cryptography include protecting files from theft or unauthorized access, securing communications from interception, and enabling secure business transactions. Other cryptographic techniques can be used to guarantee that the contents of a file or message have not been altered (integrity), to establish the identity of a party (authentication), or to make legal commitments (non-repudiation).

This should not lead to the conclusion that once the Internet is made secure, all privacy problems will disappear. While making the Internet as secure as possible is necessary for privacy, it is not sufficient in and of itself. Security is not synonymous with privacy. Privacy, as it relates to information, deals with the broader questions of the legitimate collection, use and disclosure of personal

S. Cimato et al. (Eds.): SCN 2002, LNCS 2576, pp. 199–218, 2003.

information, and the degree to which individuals are able to exercise control over the uses of their own information. One of the earliest definition of privacy was given by Warren and Brandeis in [1], who identified privacy as *the right to be left alone*. Westin [2] discusses the importance of privacy for free societies, defining privacy as *the claim of individuals, groups and institutions to determine for themselves, when, how and to what extent information about them is communicated to others*. Privacy brings a certain kind of freedom. Prejudices based on someone's background, social status, or previous actions are not possible. It enables people to speak freely and express themselves without fear of being persecuted. Privacy has been declared a fundamental human right in Art. 12 of the Universal Declaration of Human Rights of the United Nations [3].

2 Security and Privacy of Health Information

Unauthorized accesses to data and records in the intelligence and financial industries are likely to be used for criminal purposes, such as the sale of military secrets or fraud, respectively. With medical information such breaches and uses may be more insidious, and the damages less overt. Information systems administrators in both military and financial institutions are given strong mandates to curb criminal use of the housed data; breaches are often followed by inquiries to assign responsibility, and by punitive action. The loss of credibility following a finding of gross negligence can be as damaging to the institution as the event itself.

A different picture prevails in the medical field. Unlike banks, health care institutions have avoided public backslash after breaches by blaming unscrupulous insiders who violate sacred principles of ethical behavior. The public and even many health care professionals, perhaps out of a lack of understanding of security principles and practices, assume that the high ethical standards expected of health care personnel are enough of a deterrent to the misuse of information in all but exceptional cases. This view is contradicted by the fact that medical records are routinely available to non-medical personnel for essential business functions such as claim payment processing. Moreover, medical information has concrete monetary value to other stakeholders than the health care provider.

Until recently the prevalent view in the health care industry was that investing in security would hinder efficiency, decrease performance, and increase costs. It has been argued that we should learn as a society to accept some measure of risk to the security of our medical records as a better alternative to pricing health care beyond the reach of many. Embracing such stance shifts the cost of damages from the institution to the individuals who become victims of such breaches. Unfortunately, as medical information systems are deployed more widely, and made more interconnected, we must expect security violations, already common occurrences, to increase in number.

The consequences of security breaches for individual victims vary from inconvenience to ruin. (For examples, see [4,5,6,7,8].) Despite such incidents having occurred and gained public attention, market forces have failed to promote

changes in industry practices, mainly due to security not being a decision factor for those selecting a medical benefits plan. The population at large is not educated about security issues, and the medical plan choices available to them are limited to a few alternatives offered by employers. In addition, these plans do not disclose security practices in their informative materials. Finally, there are just too many variables to take into account when choosing a plan, not least the price-tag. Realizing another path was needed to promote change in the status quo, victims and other concerned citizens formed political pressure groups to achieve improvements through government mandate, and have been successful in having laws enacted to enforce health care standards for information security and privacy. One benchmark was the enactment of the 1996 Health Insurance Portability and Accountability Act (HIPAA) by the American Congress.

Other countries have legally addressed the protection of electronic health data; however, in the United States the issue acquired greater relevance for a number of reasons. Unlike European nations, the United States does not have comprehensive legislation protecting personal data. In some cases personal data gathered by the government falls under the jurisdiction of the Freedom of Information Act (FOIA), but in general private data gatherers are not federally regulated. In addition, the United States health system is fully privatized, very complex, highly competitive and dynamic. In such an environment, structural cost changes can upset the balance of power between stakeholders, while access to information is valued as a resource of business intelligence. Finally, the United States has advanced faster than some other nations in creating large health information systems and is converging to wide interoperability promoted by the adoption of a universal standard for health-related transactions and data: the Health Level Seven (HL7) standard [9].

Since changes to the security architecture of medical information systems are being driven by the requirements of law, and not by perceived market demand, security approaches are likely to be selected on the basis of cost, with technical merit as a secondary consideration. The patients whose information is to be protected do not decide on which mechanisms or resources shall be utilized for that protection. It is thus important to maintain a focus on efficiency when designing medical information security solutions.

Another aspect of medical information security is the importance of accessibility. It is well known that information accessibility and security are at odds. In the medical field, accessibility for certain authorized functions must overrule any other concerns: When a doctor needs to access the information about a patient in order to provide emergency treatment, it is imperative that the data become available without delay. Access control structures adopted in medical systems must accommodate such special situations without voiding all security guarantees along the way.

Technically speaking, security of medical information involves the adoption of access control and audit techniques. Access control can be easily implemented by using Access Control Lists (ACLs), since such tools are already available in at least some of the systems used by health organizations. On the other hand, ac-

cess control requirements are most easily expressed in terms of role-based access control concepts. Another dimension to the problem of implementing access control is that the required granularity (which can be at the data item level and not simply at the record or file level) may not enjoy native support in every system. The continuing operation of many legacy systems, some of which have no native support for personal accounts and/or different access control privileges are also a complicating factor. One recurring solution has been to hide data storage servers behind "information brokers." The latter are used to support interoperability, implement security intelligence at the application level, and maintain an encrypted and authenticated transport layer. For a good (somewhat outdated) review of approaches to medical information in the United States, see [10] which also includes recommendations for improvements.

From a theoretical standpoint, a formal model for medical information security, proposed by Ross Anderson [11], bears some similarities with the Clark-Wilson model for general commercial systems requiring data integrity [12]. Perhaps such finding is a reflection of the fact that medical data is not easily categorized in "security levels" – though some types of records are more sensitive than others – and that the foremost concern in security of medical information is to permit only authorized persons to access the data, and then only for well defined, allowed uses. From a practical perspective, it is necessary to model each business method and its accompanying information flow when considering the security of medical systems, as the extent of allowable uses and accesses to data may not be clear at first inspection.

While security of medical information involves issues of maintaining integrity and implementing access control, such measures are not sufficient to fully protect the privacy of patients. As with other industries, often the compromise of information involves the participation of insiders who have privileges to data or capability of improperly acquiring privileges, such as system administrators. Further technical obstacles must be put in place to reduce the opportunity for abuse. It is important again to remember that medical data is the most valuable source of business intelligence in the medical field, and worth more than would be suspected. Medical information is often stored and transmitted unencrypted, allowing technically capable unauthorized insiders or intruders to circumvent access controls. Fortunately, there is a growing appreciation for the need to maintain medical data in an encrypted format.

Patient and doctor advocacy groups have raised the issue of privacy attending the health care provider when choosing treatment options[13]. Such concerns stem from strong pressure from health care benefit administrators to curb medical costs by reducing the use of costly treatments and procedures. Doctors argue that treatment recommendations cannot be solely based on analysis of a medical chart, but must include subjective criteria derived from direct personal observation of the patient. As administrators push for standardized methods of review, and statistical analysis of doctor's prescribing patterns, there is a corresponding loss of a doctor's ability to provide treatment recommendations based on her or his own best judgment. As a result, the patient's interests can take second

place to a doctor's fear of not performing according to efficiency parameters set arbitrarily by administrators; ultimately the privacy of doctors is of interest to patients as well.

3 The MIPA Project

The 1996 Health Insurance Portability and Accountability Act (HIPAA) required the United States Department of Health and Human Services (HHS) to issue regulations protecting the privacy of health information [14]. Draft regulations were opened for some public comment periods and, on April 14 2001, the regulation went into effect. The HHS Office for Civil Rights (OCR) is responsible for implementing and enforcing the privacy regulation.

The regulation applies to health plans, health care clearinghouses, and to health care providers who transmit health information in electronic form in connection with specified financial and administrative transactions, such as claims for payment. These organizations are referred to as *covered entities* in the regulation. There are a number of elements that must be satisfied before health information is protected by the regulation. First, it must be health information as defined in the regulation. Second, the health information must be individually identifiable. Finally, it must be created or received by a covered entity. Health information is broadly defined as any oral or recorded information relating to the past, present or future physical or mental health of an individual, the provision of health care to the individual, or the payment for health care. This definition is broad enough to encompass not only the traditional medical record but also physicians personal notes and billing information. The regulation establishes a new federal legal right for individuals to see and obtain a copy of their own protected health information in a designated record set for as long as the information is maintained. It also establishes deadlines for covered entities to respond to requests for access and creates procedures for reviewing denials of those requests.

We believe that relying on supervision and, eventually, sanctions to enforce privacy protection may not be effective. It can be quite difficult or impossible to prove non-compliance with a policy, and law enforcement is usually expensive, slow, and complex. Moreover, legal protection can only be applied after the problem has occurred, when the damage has already been done and sensitive information has been leaked. Medical privacy should be addressed also from a technical point of view. In particular, techniques should be investigated that would make it technically impossible to violate the privacy of health care consumers.

The MIPA (Medical Information Privacy Assurance) Project seeks to develop privacy technology and privacy-protecting infrastructures to facilitate the development of a uniform health information system so that individuals can *actively* protect their personal information. The problem of data protection and privacy is addressed on a technical level, thus preventing any violation of a privacy policy in advance rather than correcting it after it has occurred. Supervision of compliance with privacy regulation can be substituted by a supervision of proper

usage of technical enforcement mechanisms. The latter can be more tractable if mechanisms are built to make improper usage easy to detect and to legally prove.

The main goal of our research is to design and then implement a system that allows users to interact anonymously with different organizations, using different pseudonyms, in a way that each user can prove a statement to an organization about his relationship with another organization, while still remaining anonymous to both, i.e., no information other than the statement is revealed even if the organizations involved cooperate. A major feature of our system is the possibility to revoke the anonymity of individuals in emergencies or cases required by law. In particular, section 164.512(j) of the federal privacy regulation states that a covered entity may use or disclose protected health information if the covered entity, in good faith, believes it is necessary to prevent or lessen a serious and imminent threat to the health or safety of a person or the public. Any disclosure must be to a person reasonably able to prevent or lessen the threat, which could include the target of the threat. Furthermore, section 164.512(2)(d) states that disclosures may be made to health oversight agencies for oversight activities authorized by law, including audits; civil, administrative, or criminal investigations; inspections; licensure or disciplinary actions; and other activities. The regulation allows a covered entity to disclose protected health information in response to an order of a court or administrative tribunal (section 164.512(e)) and to law enforcement officials (section 164.512(f)).

Finally, our framework incorporates the figure of the Privacy Officer in accordance with section 164.530(a) of the federal regulation which requires a covered entity to designate a privacy official for the development and implementation of its policies and procedures.

4 Ongoing Research Work

We now describe ongoing projects and research programs within the scope of the MIPA project. These efforts include the development of an efficient credential transfer system, a prototype for a centralized anonymous medical record repository, and a system for filling electronic prescriptions (including claim processing) that protects the privacy of patients and doctors.

Credential transfer systems are a controversial departure from the centralized control adopted by today's medical information systems. Yet, an interest in credential transfer systems is justified by the HIPAA privacy standard of minimum disclosure (164.501(b)). Our investigation tries to answer the question of whether such systems could be made efficient enough for wide deployment. The flexibility inherent in credential-based systems would allow for the development of an information infrastructure capable of explicitly supporting privacy protection goals at a technical level. Also importantly, credential systems can be described at a high level of abstraction suited to policy-making discourse, thus facilitating the involvement of common citizens in the ongoing health privacy debate.

The anonymous medical repository project explores privacy issues that have not been addressed by HIPAA regulation, namely which security practices should attend the storage of identifiable health data. For instance, digital signature standards have not been adopted by HIPAA. Yet, the HIPAA requirement of signatures in consent forms (164.506(c)(6)) implies the need of such adoption, because support of fully electronic formats is an essential part of the HIPAA mandate. Our anonymous repository prototype implements several privacy and security mechanisms: use of pseudonyms, support for patient's control of health information, audit records, and policies for role-based access control. Several of these mechanisms have been explicitly recommended by security experts in previous reviews of existing practices [10].

Another area in which HIPAA is short on specifics is the management of medicine prescription data. While pharmacies and pharmacy benefit managers are not covered entities under the jurisdiction of the regulation, they may be covered as business associates, in case they receive identifiable health information from health care providers that can be considered an extension of the provider's health care service operations (section 164.504(e)). Moreover, there is ample legal precedent for coverage of medicine prescription data under the same umbrella as other identifiable health information. For instance, in many American states there is such legislation [15], and some courts have agreed with an expansive interpretation [16] even in the absence of laws to that effect. Our design of a privacy-preserving electronic prescription system uses group signatures to provide anonymity to both patients and doctors while maintaining the rigorous levels of accountability and non-repudiatability which the law stipulates for these systems.

In the following subsections we describe each of these topics in further detail.

4.1 A Simple Credential-Transfer System

The HIPAA Privacy Regulation defines several rules for use and disclosure of health information (section 165.501). Section 164.502(a) of the regulation requires a covered entity to disclose protected health information only to the individual who is the subject of the information and to HHS for enforcement of the privacy regulation. In most circumstances, a covered entity can choose not to disclose information. HHS expects covered entities to rely on their professional ethics and exercise their own best judgment in deciding when they will permit the use and disclosure of protected health information. However, health plans and providers routinely hire other companies and consultants to perform a wide variety of functions for them. Health plans and providers, for example, may work with outside attorneys, bill collectors, computer specialists, or accreditation organizations. All of these *business associates* need to access some patient information, but they are not directly subject to the privacy regulation. To allow information to be shared with business associates and to protect the information as it is disclosed to them, the rule establishes specific conditions on when and how covered entities may share information with business associates. These rules are included in sections 160.103, 164.502(e) and 164.504(e) of the regula-

tion [14]. A covered entity is permitted to disclose protected health information to a business associate or to allow the business associate to create or receive protected health information on its behalf, if the covered entity obtains satisfactory assurance that the business associate will appropriately safeguard the information. Generally, this safeguard will take the form of a written contract which, among other things, requires the business associate to not use or disclose the information other than as permitted or required by the contract or as required by law, and to implement appropriate safeguards to prevent inappropriate uses and disclosures.

Instead of relying on the good faith of the parties involved, it is preferable to make the process completely automatic so that it is technically impossible to deviate from the regulation. In our privacy framework, users would select multiple pseudonyms that allow them to interact anonymously with multiple organizations. The idea of using pseudonyms instead of anonymous transactions has several advantages. For instance, anonymous re-mailers allow a user to send a message which has no identifying information when it reaches the intended recipient. However, the header information that routes reply messages has been altered in the re-mailing process, so a recipient may not reply to an anonymous message as he might to an ordinary e-mail message. A pseudonymous system, instead, receives a user's message, strips it of any identifying headers, then attaches a pseudonym, or alias, by which the user has chosen to be permanently identified. The message is then sent to the desired destination. These pseudonyms are unlinkable, i.e., given two or more pseudonyms, it is impossible to determine whether they correspond to the same user. However, in real-life environments often the user must prove to one organization a statement about his relationship with another. This statement is called a *credential* that can be linked to pseudonyms. When information about an individual is to be sent from one organization to another, the first organization issues a certificate that is a "credential on a pseudonym" to the individual, showing that the particular credential applies to his pseudonym used with that organization; then the individual transforms this certificate into a "credential on a second pseudonym" used with the second organization.

The basic properties of a credential transfer are *unforgeability* and *unlinkability*. Users should not be able to forge credentials (unforgeability) and organizations should not be able to trace an user by linking two or more credentials (unlinkability). Stated informally, unforgeability protects the organizations from malicious users, and unlinkability protects the users from malicious organizations. The process of transferring credentials should not involve any external entity (or third party) but rather only the user and the target organization.

Our future work in credential transfer systems will seek to find solutions that are optimized for the requirements of health care operations. For instance, disclosures of medical information between organizations that know the patient under different pseudonyms may take place without the patient's on-line participation: The only requirement is the patient's consent. Our chosen approach to the development of credential transfer systems uses the cryptographic primitive named

group signature. We have developed solutions involving group signatures computed through optimized use of a semi-trusted party. We are also investigating the possibility of credential transfer systems with easily shared parameters (such as discrete log-based systems). Please refer to appendix A for a more detailed description of credential transfer systems based on group signatures.

4.2 A Centralized, Anonymous Repository for Medical Records

Today medical records exist in a variety of formats, such as primary practitioner notes, hospital admittance records, radiological images, to name a few. While much information is still in paper format, there is a trend to move to electronic formats. For instance, use of personal digital assistants in clinics have facilitated the entry of practitioner notes directly into the clinic's electronic repository.

From a purely medical perspective, there is a strong motivation to gather all the acquired data about the patient, as the resulting record will provide a more complete view of the patient's health status, and omissions of critical data and observations would be less likely to happen.

Centralizing and sharing identifiable medical information brings its own problems. If all of a patient's medical records can be obtained from one place, the outcome of a breach of privacy can be more damaging. There are also higher incentives for intruders to try and compromise a single protection mechanism.

Here we describe a prototype for a centralized medical records repository that incorporates strong patient privacy safeguards. While we do not advocate the creation of national medical databases, we believe that such services will be available at least on a voluntary basis, and in fact similar services already exist [17].

System Entities: A "person" refers to anyone who is using the medical records repository. Persons have unfettered access to their own records (in that case, we refer to them as *record owners*). In order to access another person's records however, one must have proper permissions. Such person is referred to as *authorized user.* Unlike record owners, authorized users typically possess a limited set of privileges over the records they can access. Authorized users can be doctors, nurses and other medical personnel, lab technicians, pharmacists, as well as non-medical users, such as privacy officers and government employees.

The *Health Identification Provider (HIP)* is a trusted central authority that verifies persons' proof of identity to ensure they have only one medical record in the system. The HIP also establishes the set of permissions granted to each person. These permissions will depend on the roles played by these persons in the system, as determined by credentials and certificates presented with identification. The *Centralized Medical Database (CMDB)* is a repository that contains the medical records of all the people in the system, including the auditing information.

The CMDB supports a single access point to a person's entire medical record, that is, one has to interact only with the CMDB in order to review and amend the person's medical record. The person's anonymity is preserved in the interactions

that take place in the system, through the use of pseudonyms: The owner's real name is not needed, instead a randomized encryption of the HIP-issued identifier is used. For example, laboratory technicians have no real need to know the names of patients whose samples they are analyzing.

System Description: Technology plays an important role in preventing inappropriate access to patient information. Strong user authentication (to ensure access control) and audit records of accesses (to ensure accountability) are powerful abuse deterrents. The patients can specify who may read/update their medical record, and for what duration. So, in compliance with the HIPAA regulations (Privacy Rule, section 164.524 [18]), patients have control over their medical records. All reads of and appends to a person's medical record are audited. Audit trail records contain details about the information access, including the identity of the requester, the date and time of the request, and the source of the request.

Persons have read access to their own entire medical record and they can make clarifications to their medical record by appending to it. The system is able to distinguish the parts of a medical record that were updated by the owner from parts updated by someone else. A crucial aspect of data access control is that access to a person's medical record may occasionally need to be done without the person's consent. (Those accesses are still audited.) Such situations include: trusted emergency medical workers accessing a person's medical record, trusted privacy officers revealing a person's identity, and trusted government employees (e.g. FBI agents) accessing a person's medical record and revealing their identity. The audit information also reveals if parts of a medical record were accessed without the owner's consent, distinguishing these from explicitly permitted accesses.

The repository accommodates access during medical emergencies. After registration, each user should create an emergency-access token to be used by emergency medical technicians whenever the patient is unable to issue an access token (e.g. when the patient is unconscious). The emergency-access token is not granted to any specific person and only certain authorized users can use it without patient's consent.

Structure of Records in the CMDB: Each person's record consists of a series of updates. Even a read-only access to the record generates an update due to the addition of audit information. Updates consist of one or more entries, each of which belongs to a single category. The category can be medical related (biographical, prescription, allergy, lab result, etc.) or can be related to security and audit control. The latter, **special**, categories are: *RevealIdentity* - allows certain authorized users (privacy officers or FBI agents) to retrieve this person's identity from the HIP; *RevealHID* - allows certain authorized users to obtain the identity of the originator of the update; *ReadAudit* - for auditing purposes; an entry of this category is generated whenever someone (including the owner) views this person's medical record, and it contains a description of the request.

To insure integrity, only append operations are allowed: The system doesn't allow deletion from a medical record. This design decision was made in compliance with the practices used in the health care industry [19].

The system enforces role-based mandatory access control by associating read /append permissions with each category. Permissions define how a person is allowed to access another person's medical record. An authorized user's effective permissions are the result of a combination of the permissions granted by the record owner and permissions intrinsic to the user's role in the system. Role-based permissions have an expiration date and may assume three values: 'allow', 'consent required', 'deny'. Permissions granted by the owner have either value 'allow' or 'deny' for each category. Thus, for instance, a general practitioner may have role-based 'allow' permission to basic health categories (such as allergies), while only 'consent required' permission to psychiatric entries. So when the patient issues an access token to his practitioner, the doctor will automatically have access to allergy entries, but will need explicit patient consent to review the patient's psychiatric history. Persons always have full access to their own medical records, because the database recognizes the owner's identity and ignores the role-based permissions for such accesses. The owner may then issue herself an access token that has 'allow' permissions for all categories and exercise her right of review.

Although the CMDB is a trusted server in the current implementation, it is possible to lower the level of trust in the CMDB by using encryption: If the medical records are represented as XML documents, their structured format would allow the use of technologies like XML Signature and XML Encryption in order to sign and encrypt specific entries of the medical record.

Interactions in the System: Users first must register with the HIP and obtain a credential on a pseudonym of their choice. This credential is then used to register with the CMDB under another pseudonym. Later, whenever authorized users wish to access medical records, they need access tokens from the record owners. Again, this access will be qualified by the user's intrinsic permissions (granted by the HIP during registration) and the permissions explicitly listed in the access token.

For implementation details, see appendix B.

4.3 Private Electronic Prescriptions

Keeping drug prescription information private is part of an overall strategy to protect medical information, as prescriptions contain revealing information about (at least some aspects of) the medical history of the patient.

The project on private e-prescriptions approaches the issuing of drug prescriptions as a business process composed of several distinct subtasks or workflows. Issuing a prescription may entail adding entries to a patient's medical records; performing queries in expert systems for possible drug interactions or medical conditions which, if present, may counter-indicate the use of the medication; creating evidence records of authorized use of the medicine, to comply with laws and regulations; issuing of claim forms for billing purposes.

In specific situations, further events may use this information. For instance, the medical prescription, in combination with other parts of a patient's medical record may be disclosed at future points in time, together with the patient's identity, for legal investigative purposes (such as in the context of a malpractice lawsuit) or to comply with legal requirements, such as the right of review enjoyed by patients, or for other purposes of law enforcement.

A prescription system will eventually interact with all the following parties: Patient, Doctor, Pharmacist, Insurer, Privacy Officer, Enforcement agent, Judge, Certification Board and Certification Authorities. By *Doctor* we mean the person issuing the prescription, who in practice might be a licensed practitioner or paramedic. By *Pharmacist* we denote a server within a Prescription Benefit Management system, organizations which process prescription claims. (Such organizations process over 99% of all prescription claims in the United States.) *Privacy officers* are the persons and systems within a medical organization with the responsibility of overseeing compliance with privacy regulations and policies. A computer server maintaining a database that translates pseudonyms into patient (or doctor) names is an example of a computer server executing functions of a privacy officer. *Enforcement agents* must be able to link prescriptions per-patient and per-doctor to perform statistical analysis for fraud-prevention purposes. A *Judge* may revoke the privacy of a party in a transaction as part of a legal proceeding. A *Certification Board* grants powers to entities to issue prescriptions. *Certification Authorities* issue digital certificates affirming such capacities, roles and responsibilities, possibly in the form of a pseudonymous certificate.

The goal of protecting patient confidentiality must be balanced against potential for fraud in a truly anonymous system. Thus the privacy of the patient must be *revocable* under provisions of the law. For similar reasons, it is desirable that patient participation in transactions should result in *non-repudiatable* evidence of patient engagement. More restrictively, transactions by the same patient should be *linkable* by the PBM. Otherwise current fraud-prevention investigative practices using statistical treatment would be rendered useless. One could argue that linkable anonymity is no anonymity at all. Our counter-argument is that patients should have the right to request a change in pseudonyms if they have reasons to believe their privacy is under risk of being compromised, and privacy officers may place reasonable restrictions on how often such pseudonym changes may take place. Another solution would be issuing different pseudonyms every time enrollment is renewed and a new smart-card is issued. That would limit histories to shorter periods of time, reducing risks of privacy breach while still allowing investigative profiling to take place.

Confidentiality of the doctor's identity must be similarly revocable, and her participation non-repudiatable. However in our view there is no good reason for doctors' transactions to be linkable by the pharmacist/PBM. Instead fraud on the part of doctors could be investigated by the insurer. In other words, doctors' transactions could (and probably should) be *unlinkable* for the pharmacist, but linkable from the perspective of the insurer.

A detailed description of an anonymous electronic prescription system is presented in [28].

5 Conclusion

The relevance and urgency of security and privacy problems faced by medical information systems have recently led researchers to develop approaches tailored to these systems. In this paper we introduced MIPA, a project dedicated to the understanding of both theoretical and practical issues involved in improving the security and privacy of electronic medical data. To this date, the MIPA project has sponsored the design of credential transfer systems to support requirements of minimum disclosure, the development of a prototype of a centralized anonymous repository of medical data, and the design of a system for anonymous electronic prescriptions.

References

1. S. D. Warren and L. D. Brandeis. *The right to privacy.* Harvard Law Rev. 4, pages 193-220, 1890.
2. A. F. Westin. *Privacy and Freedom.* Atheneum, New York, 1967.
3. United Nations, Universal Declaration of Human Rights. `http://www.unhchr.ch/`, December 1948.
4. D. F. Linowes and R. C. Spencer. How employers handle employees' personal information. `http://www.kentlaw.edu/ilw/erepj/v1n1/lino-main.htm`, 1997.
5. Hospice Patients Alliance. When getting prescriptions filled, beware of medication substitutes! `http://www.hospicepatients.org/substituteRx.html`.
6. S. Lehrman. Keeping your genes private. *GeneLetter.*
7. N. Keene, W. Hobbie, and K. Ruccione. Childhood cancer survivors. `http://www.patientcenters.com/survivors/news/jobs.html`, OncoNurse.com.
8. C. Jabs. The myth of privacy: Technology is putting your medical history on public view-and you in jeopardy. *FamilyPC*, 2001.
9. Health Level Seven. `http://www.hl7.org`
10. For the record: Protecting Electronic Health Information. Computer Science and Telecommunications Board, National Research Council 264 pages. Washington, DC: National Academy Press 1997.
11. R. J. Anderson. A security policy model for clinical information systems. In *Proceedings of the 1996 IEEE Symposium on Security and Privacy*, 1996.
12. D. Clark and D. Wilson *A comparison of commercial and military security practices.* In Proceedings of the IEEE Symposium in Security and Privacy, IEEE Press, 1987.
13. T. Albert. Doctors ask AMA to assure some privacy for their prescription pads. `http://www.ama-assn.org/sci-pubs/amnews/pick_00/prl11225.htm`, American Medical News. 2000.
14. Health Insurance Portability and Accountability Act. `http://www.hhs.gov/ocr/hipaa/`
15. Ohio State Board of Pharmacy. Confidentiality of patient records. `http://www.state.oh.us/pharmacy/rules/4729-05-29.html`. 1999.

16. T. Albert. Records privacy extended to pharmacies. http://www.ama-assn.org/sci-pubs/amnews/pick_01/prsb0402.htm, American Medical News. 2001.
17. WebMD Health. *My Health Record,* http://my.webmd.com/my_health_record.
18. Office for Civil Rights. Standards for privacy of individually identifiable health information. http://www.hhs.gov/ocr/hipaa/finalmaster.html. 2001.
19. Hipaadvisory.com Final standards for individually identifiable health information; §164.526, Amendment of protected health information. http://www.hipaadvisory.com/regs/finalprivacy/526.htm
20. D. Chaum and E. van Heyst. Group signatures. In *Advances in Cryptology – EUROCRYPT '91*, vol. 547 of *LNCS*, pp. 257–265, Springer-Verlag, 1991.
21. D. Chaum, *Security Without Identification: Transactions Systems to Make Big Brother Obsolete*, CACM Vol. 28, No. 10, October 1985.
22. D. Chaum and J. Evertse. A secure and privacy-protecting protocol for transmitting personal information between organizations. In *Advances in Cryptology*– CRYPTO'86, pp. 118-167. Springer-Verlag, 1986.
23. I. Damgård. Payment systems and credential mechanisms with provable security against abuse by individuals. In *Advances in Cryptology – CRYPTO '88*, pp. 328–335, Springer-Verlag, 1988.
24. L. Chen. Access with pseudonyms. In Cryptography: Policy and Algorithms, pp. 232-243. Springer-Verlag, 1995.
25. A. Lysyanskaya, R. Rivest, A. Sahai, and S. Wolf. Pseudonym Systems. In Selected Areas in Cryptography. Springer-Verlag 1999.
26. Jan Camenisch and Anna Lysyanskaya. Efficient Non-transferable Anonymous Multi-show Credential System with Optional Anonymity Revocation. In Eurocrypt'01. Springer Verlag, 2001.
27. G. Ateniese, M. Joye, J. Camenisch, and G. Tsudik. A Practical and Provably Secure Coalition-resistant Group Signature Scheme. In In Advances in Cryptology - CRYPTO 2000. Volume 1880 of LNCS, pages 255-270, Springer Verlag, August 2000.
28. G. Ateniese and B. de Medeiros. Anonymous E-Prescription. In ACM Workshop on Privacy in the Electronic Society (WPES '02), Washington D.C., USA, November 2002.

A Credential Transfer System

Assume that a user U selects a master secret x which is not revealed to others. The user U is known to the organizations O_1, O_2, ... , O_n by the pseudonyms $f_1(x)$, $f_2(x)$, ..., $f_n(x)$, respectively. Each function $f_i(\cdot)$ is a one-way function that does not reveal any information about the argument. An example may be $f_i(x) = g_i^x$ taken modulo a prime p such that $p = 2q + 1$ for another prime q. The base g_i is a generator of the set of quadratic residues in $\mathbb{Z}_p^*$. We will employ group signatures as building blocks. Group signature schemes are a relatively recent cryptographic concept introduced by Chaum and van Heyst [20] in 1991. In contrast to ordinary signatures they provide anonymity to the signer, i.e., a verifier can only tell that a member of some group signed. However, in exceptional cases such as a legal dispute, any group signature can be "opened" by

a designated group manager to reveal unambiguously the identity of the signature's originator. At the same time, no one — including the group manager — can misattribute a valid group signature.

We propose to use the following technique: Suppose the user U has a credential from the organization O_i. The user U can join the group of people with such a credential so to receive a group certificate containing $f_i(x)$ which states that $f_i(x)$ is a member of the group of people with that particular credential. In order to transfer the credential to the organization O_j, the user U may generate a group signature and proves that the secret x inside the signature is the same in $f_j(x)$.

Consider the following example. Employers often check the health condition of applicants before hiring them to determine whether they are suitable for a particular job or whether they could be a threat to other employees. Section 164.512(b) of the federal privacy regulation explicitly states that, in certain circumstances, a health care provider may disclose to an employer protected health information about an individual who is a member of that employer's workforce. However, this information is often abused, leading to health discrimination or prejudice. The law, in most cases, prohibits employers from denying privileges, such as health coverage, based on a worker's pre-existing medical conditions. Nevertheless, it is hard to prove that health information was abused. Health discrimination is a sad reality of our society. However, information can be made technically impossible to abuse. Suppose O_1 and O_2 represent the doctor and the employer, respectively. The user U wants to prove to the employer O_2 that he is healthy. The user U then gets the credential from the doctor O_1 and transfer it to O_2. The transfer is completely anonymous and unlinkable even if O_1 and O_2 collude, thus the information inside the credential cannot be abused since it cannot be linked to any real identity.

In our system we will assume that the long-term secret x uniquely identifies a particular individual who is not willing to share x. In order to generate a pseudonym with the organization O_i, the user U computes $f_i(x)$ and transfers a credential from the certification authority to O_i, proving that the secret in the credential is the x in $f_i(x)$.

`Pseudonym Generation (registration):`

- $U \longleftarrow CA$: credential $c(x)$;
- $U \longrightarrow O_i$: $f_i(x)$, proof of knowledge of x;
- $U \longleftrightarrow O_i$: transfer of $c(x)$ (proof that the same x is in $c(x)$ and $f_i(x)$);

In the system we propose, issuing a credential is equivalent to joining a group. Once a group signature is selected, the organization O_i has to form several groups, one for each credential. Therefore the user U, known as $f_i(x)$ to the organization, obtains a credential by joining the group G. The organization has to make sure that the secret inside the group certificate for user U is the same in the pseudonym $f_i(x)$. The user U has to prove this without revealing any information about x to the organization.

`Issuing a Credential:`

- $U \longrightarrow O_i$: request to join group G;
- $U \longleftrightarrow O_i$: release of group certificate, where x is secret;

Transferring a credential is as easy as generating a group signature under the group certificate. The target organization O_j generates a challenge that will be signed by the user. (U and O_j run a three-pass identification protocol.) The organization O_j receives also the group public-key signed by the organization O_i. Finally, the user U has to prove that the secret inside the group certificate is the same secret in the pseudonym with O_j.

`Transferring a credential:`

- O_j : selects a challenge c;
- $U \longleftrightarrow O_j$: execute the group signature procedure on c;
- $U \longleftrightarrow O_j$: U proves that the secret inside the group certificate is the same in $f_j(x)$;

When computing the group signature (or the response to the identification protocol), the user encrypts the group certificate under a trusted third party's public-key. The third party will then be able to open the signature and reveal the group certificate or, if cooperating with the certificate authority, the real identity of the user in case of disputes or emergencies.

The credential transfer problem was introduced by Chaum [21] in 1985. Then, Chaum and Evertse [22] developed a model for pseudonym systems incorporating the notion of credential transfer and presented an actual scheme based on the RSA problem. However, the scheme relies on a trusted center that performs the transfer of a user's credential from one organization to another. Subsequently, Damgård [23] developed a model and a scheme of pseudonym systems not requiring trusted centers. However, the scheme is not very efficient since it is based on zero-knowledge proof constructions. Later, Chen [24] proposed a practical scheme based on the discrete logarithm problem but a component of her system, the certification authority, must be totally trusted as it can transfer credentials between users. As noticed in [25], a common weakness of all these schemes is that they do not prevent a user from sharing his credentials with others. For instance, a patient can share his health insurance with all his friends. This problem is solved in [25], but the scheme proposed is more suitable for single-use or *one-time* credential models, where a credential can be transferred between organizations only once[1]. Recently, a multiple-use credential protocol has been proposed in [26] based on the techniques developed in [27]. The protocol is secure under standard assumptions and it is efficient when the group of participants is static. However, in case of large and highly dynamic groups, the efficiency degrades notably and linearly with respect to the number of participants.

[1] The authors of [25] present a construction for multiple-use credentials that, however, do not conform completely to the specifications of their model.

B Centralized Anonymous Repository for Medical Records

We assume that both the Health ID Provider (HIP) and the Central Medical Database server (CMDB) have public/private key pairs. Let these be HIP_PK /HIP_SK and CMDB_PK/CMDB_SK, respectively. All protocols in this appendix assume the existence of secure channels between parties.

Protocols: Following is a short descriptions of the interactions that take place in the system.

A person registers with the HIP using a pseudonym (let this be HIPID_PK) and gets a certificate signed by HIP which proves that the person has registered with the system (protocol A). The person uses this certificate and another pseudonym (let this be DBID_PK) to register with the CMDB (protocol B). In all future interactions with the system the person's real identity is not used. Instead one of these pseudonyms is used to uniquely identify the user. These identifiers will suffice in many health care transactions.

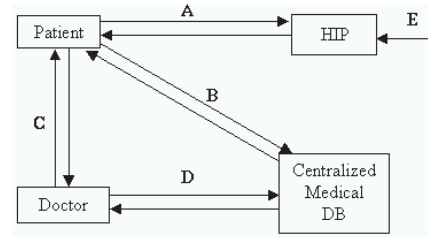

Fig. 1. protocols

We use the following notation in this section: $S_{SK}(m)$ for signing a message (with private key SK), and $E_{PK}(m)$ for encrypting a message (with public key PK).

Protocol A: A person (P) contacts the HIP (N), providing a proof of identification, e.g., birth certificate or driver's license. The person may also use protocol A to prove competence for certain health care roles. For instance, P may demonstrate the ability to write prescriptions by showing a license to practice medicine and a doctor's degree certificate. The protocol ends with N issuing a certificate to P that authenticates the person with the system and establishes P's role-based permissions.

1. P → N: *HIPID_PK*, `Name`, `Proof of Identity`
2. N → P: r (a random challenge)
3. P → N: $S_{HIPID\_SK}(r)$
4. N → P: Validate response to challenge, check proof of identity, add person to the database, then send $Cert_{HIPID}(P) = S_{HIP\_SK}(HIPID\_PK$, `Permissions`$)$

Protocol B: A person (P) contacts the CMDB (C), providing a credential from HIP and a new pseudonym $DBID_{PK}$. C verifies the credential and creates an account in the database for P under the given pseudonym and with the proper permissions.

1. P → C: *DBID_PK*, $Cert_{HIPID}(P)$
2. C → P: Validate $Cert_{HIPID}(P)$, send $r1$ and $r2$ (random challenges)
3. P → C: $S_{HIPID\_SK}(r1)$, $S_{DBID\_SK}(r2)$

4. C → P: Validate response to challenges, add person to the database, and return OK message

Protocol C: in the general case, protocol C is executed between a Person (P) and a health care provider (D), with P issuing an access token that can be used by D to access/append P's medical record.

The access token contains a list of health record categories and one of the values 'allow' or 'deny' associated with each. When accessing P's record, D will have privileges for those categories listed with permission 'allow' in the access token, provided that D's own set of role-based permissions does not list the value 'deny' for the same category. D will also have privileges for the categories for which D's role-based permissions list the value 'allow,' regardless of what value is listed for this category in the access token. In general, a role-based permission may be associated with the values 'allow,' 'deny,' and 'consent required.' Only those categories for which D's role-based permissions list the value 'consent required' are affected by the access token permissions, and for those, D's effective permissions over Q's record are those granted by the token, as shown in the table below.

P's role-based permissions	Access token permissions	Effective permissions
deny	*	deny
allow	*	allow
consent req'd	allow	allow
consent req'd	deny	deny

where $*$ = don't care

The access token is encrypted so that the provider cannot see the pseudonym (DBID_PK) of the person.

1. D → P: $Cert_{HIPID}(D)$
2. P → D: r (a random challenge)
3. D → P: $S_{HIPID\_SK(D)}(r)$
4. P → D: Validate response to challenge; if OK, create $\text{Token}_{\text{Access}}$ and send to D

The $\text{Token}_{\text{Access}}$ has the following format:

- $\text{Cert}_{\text{Access}} = \text{S}_{\text{DBID\_SK(P)}}(\text{DBID\_PK(P)}, \text{HIPID\_PK(D)}, \text{Permissions})$
- $\text{Token}_{\text{Access}} = \text{E}_{\text{CMDB\_PK}}(\text{DBID\_PK(P)}, \text{Cert}_{\text{Access}})$

where DBID_PK(P) and HIPID_PK(D) are, respectively, P's and D's public keys with respect to the database.

There are two special cases in protocol C. The first is when a Person (P) generates a self-access token. In this case, D = P, the token lists all categories with value 'allow', and it does not expire. This token is used by P to access her own record exercising the role of 'record owner.'

The second special case is when P creates an emergency access token. This token is not issued to a specific person (HIPID_PK(D) = null), and it does not

expire. It lists all categories with 'deny' permission. This is sensible because medical emergency technicians have all categories listed with permission 'allow' in their list of role-based permissions. On the other hand, if this token is stolen, it can generally not be used to compromise P's privacy; even doctors have only 'consent required' associated with the 'RevealIdentity' category and hence they are not able to identify the owner of the lost token.

After successfully completing protocol B, P should create an emergency access token as described above. If P is unable to execute protocol C (e.g. P is unconscious), an emergency medical technician (or a privacy officer) obtains the emergency access token created by the patient and uses it as if acquired through a regular execution of protocol C. The system should provide one or more ways for retrieving emergency access tokens. For instance, it could be carried by the patient inside a convenient storage media.

Protocol D: A registered user (P) contacts the CMDB (C), requesting execution of an operation (read/append) on a third person's (Q's) medical record. The request will succeed if the combined role-based permissions (P's) and the permissions granted in the provided access token allow it.

1. P → C: $HIPID\_PK(P)$, $Token_{Access}$, `Permissions`, `Operation`
2. C → P: r (a random challenge)
3. P → C: $S_{HIPID\_SK(P)}(r)$
4. C → validate response to challenge; if OK, execute the Operation

There are two special cases of protocol D. When a Person modifies her own medical record (P = Q), all permissions are granted regardless of which permissions P may have; and the audit information is created with 'null' for the value of the ID field of the person who made the `operation`, so as not to reveal the medical record owner's pseudonym. In the case that P is a medical technician requiring emergency access to Q's medical record, P holds a token that was issued to an unspecified (null) recipient. Only the DBID_PK(P) is used from the emergency access token. P will only be able to access those categories for which she has the value 'allow' in her list of role-based permissions.

Protocol D - Operation types

- *Read Operation* - C combines the permissions that P has, the ones that Q granted to P in the Access Token, and the permissions that P requested, and sends back the appropriate part of Q's medical record (a ReadAudit entry is also generated).
- *Append Operation* - P sends C a request to append Q's record with the entries of an array. If P has appropriate permissions, Q's medical record is updated, and an OK message is returned.
- *Validate Operation* - C returns to P what permissions she has on Q's record, combining P's role-based permissions and the permissions in the access token.

Protocol E: A person (P) contacts the HIP (N) to reveal a third person's (Q's) identity (her name) given Q's pseudonym with the HIP (HIPID_PK(Q)). No authentication is needed in this protocol because possession of someone's

HIPID_PK is sufficient proof that you were in a position to know their name, or can be trusted with their name. In fact, P may only obtain HIPID_PK(Q) from Q if Q has given P an access token with the permission 'RevealIdentity' (in which case Q has trusted P with her identity) or if P has role-based permission value 'allow' for the 'RevealIdentity' category. This latter case is only true for some *trusted* entities in the system, such as a judge or privacy officer.

1. P $\rightarrow$ N: $HIPID\_PK(Q)$
2. N $\rightarrow$ P: The name of the person whose pseudonym with the HIP is $HIPID\_PK(Q)$

Note that medical records never contain the record owner's pseudonyms (public keys), and that a patient will never use their HIPID_PK, except to create/update an account with the CMDB. However, the HIPID_PK of a doctor or authorized user is included in entries in medical records to register who made the changes. This does not compromise patient identities because the HIPID_PK never appears in their own medical records: Whenever someone modifies her own record, the identity of the initiator of changes is explicitly set to null - see Protocol D.