SAND 95-0576
Unlimited Distribution
Printed June 1995

Distribution
Category UC–705

Protecting Privacy and Information Integrity of Computerized Medical Information*

Kevin S. McCurley†
Algorithms and Discrete Mathematics Department
Sandia National Laboratories

Abstract

With the recent realization that health care spending is growing out of control, there has been a renewed interest in streamlining all aspects of the health care delivery system. One of the most promising approaches to streamlining health care is through the use of electronic information systems to manage health care information, and in particular the conversion from a paper-based patient record to an electronic patient record. The design and use of such a system raises serious concerns about the privacy and integrity of information about individuals. In this paper we survey some of the problems, and some of the technological tools and approaches that can be used to secure electronic patient records against abuse.

*A hypertext version can be found on the world wide web at `http://www.cs.sandia.gov/~mccurley/health.html`

†Author address: MS 1110, Sandia National Laboratories, Albuquerque, NM 87185. Email: `ksmccur@sandia.gov`

Contents

1 The Four Building Blocks of Security

There can be no doubt that the health care system as it exists today in the United States has some severe problems. Many of these problems are aggravated by the manner in which information is typically managed. Paradoxically, this mismanagement has contributed to protecting the integrity and privacy of medical records by making the information difficult to retrieve and use. As we enter the age of computerized medical information, this situation will change dramatically, and threats to individual privacy and health will become more serious unless systems are designed and managed intelligently. It is therefore important that during the design of such systems, we give appropriate level of attention to these security issues, and carry out some careful planning for the future.

There are four key components required for the protection of computerized medical information. They are:

Technology A wide range of technologies exist that can be used to protect medical information from improper use, dissemination, or modification. This will be the major topic of discussion in the paper. I will attempt to describe how technologies currently work, some current trends, and how I expect them to evolve in the future. Briefly, I expect a lot of money to be wasted on "closed systems" that are later discarded in favor of systems that adhere to open, non-proprietary standards.

Legislation When automobiles were first invented, they represented very little physical threat to humans because they were rare and the threats were very small. As automobiles became woven into the fabric of our society, so too did the threats, including pollution and threats to our personal safety. Legislation was drafted to regulate their use in order to protect society from autos. As our society becomes more dependent upon information, so too will we be vulnerable to threats from the abuse of information, and it is natural to expect some legislation and regulation of its use in order to protect society against this abuse. This was the basis for the enactment of the Privacy Act of 1974, which governs how government must protect information that it holds about individuals. It is time to revisit the rights of the individual regarding the use of information about them by other parties. An obvious place to start is with medical information.

Institutional Policies Technology alone cannot protect against abuse of information, since it is only a tool to enable proper handling of information. Legislation cannot and should not regulate every aspect of how information is treated. These gaps can be filled by crafting appropriate institutional policies to govern the use of technology and accomplish the broad guidelines established by legislation. On the other hand, if institutional policies alone are relied upon without legislation, then it will remain tempting for institutions to use very weak protections, particularly if it requires them to invest time or money (just as physical security against crime and invasion is not free, so too is security against invasions to our privacy). If institutional policies are used without appropriate technology, then the policies can be broken too easily. It should be mentioned that ongoing system administration of these policies will be required for effective use. Standards such as the ASTM 31.20 authentication standard should play an important role in forming such policies.

Education and training When appropriate technologies, legislation, and policies are put in place, users must still become informed of the proper use of technologies as well as their responsibilities. This will be an ongoing activity within each organization, particularly as new technology is added to enhance the capabilities of the system and new users enter the system.

If any one of these is neglected, then the others are likely to have too many demands placed on them to succeed. My own background is in the technology side of things, and much more information will be given later about this.

2 The Ultimate Architecture for Electronic Medical Records

2.1 Data has to follow the people using it

The discussion about health care reform in 1994 served to highlight the fact that our health care system is a highly diversified system of distinct institutions, including individual doctors, HMOs, PPOs, clinics, pharmacies, private and public hospitals, hospices, laboratories, etc. In order to provide effective health care, it is imperative that they be able to share information between them. Information systems that are not planned with this in mind are headed for ultimate obsolescence, because they will have failed to provide the principal benefit.

I take it on faith that we should be planning for a future where electronic medical records will be portable across systems. People move, people change doctors, and records need to follow them around. If a patient requires treatment while they are traveling, then remote access will be required, perhaps from the personal physician or from the remote physician. Caregivers will often be called upon to access information from their homes, or from computers that they carry with them on travel. If nothing else I expect to be alive thirty years from now. I don't really know what computers will look like in thirty years, but I sincerely hope that my medical history will still be available.

I hope this makes it clear that the importance of portability of data between systems cannot be overestimated. In particular, medical records should become just as usable as any other record in the new system, *independent of the system vendor.* This means that "open standards" are of paramount importance in the development of future patient record systems. Vendors selling systems into this market will compete in terms of the features that they provide, the service they provide, or any of a number of other factors. They should *not* however be encouraged to "captivate" their customers by creating records that can only be used within a "closed" system, because this will defeat the ultimate goal of portability of individual records.

2.2 The importance of universal networks

An often repeated axiom among the Internet cogniscenti is "the network *is* the computer". Personal computers have proved to be very useful tools to manipulate information, but the real power of information comes only when information is conveyed from one party to another. It is this realization that has created the intense interest in the Internet.

When the Internet was in it's infancy, standards were developed to allow computers to talk to each other in a common language, independent of their operating system, architecture, or vendor. More than anything else, it is this capability that has resulted in the tremendous impact of the Internet. This will be particularly true for medical information systems, since information will need to follow us around as we go from one place to another. Providing security for a static, physically isolated network is hard enough. Extending this to include roaming remote computers greatly complicates the situation, but must be accepted as a requirement for future systems.

The common language of network protocols that has been used to develop the Internet is commonly known as TCP/IP [6] (but should also include UDP). More than anything else, TCP/IP has demonstrated its scalability, adaptability, and suitability to a wide variety of applications. Riding on top of it is a set of protocols to support worldwide machine naming conventions (DNS), worldwide electronic mail transmission (SMTP), network management protocols (SNMP), file sharing (NFS), and an almost every imaginable service beyond these basic ones. TCP/IP has been identified most closel with UNIX machines, but there are now numerous implementations for Microsoft Windows, Apple Macintosh, and mainframes. It is notable that TCP/IP will be incorporated into Windows 95.

The type of system that I believe will ultimately be the dominant architecture for clinical information systems will consist of a distributed relational database system, spread around multiple institutions and/or locations. This database system will be accessible in a client/server fashion, using a variety of client workstations that are tailored to the particular user and/or facility. The client workstations may be fully capable desktop workstations, laptop computers, personal data assistants, or dumb terminals. The networking used to accomplish the client/server communication might be a combination of existing wiring or new wireless networks. We should expect to be able to buy server and client hardware and software from a variety of vendors, all of which can be expected to interoperate.

3 Technology for Securing Health Information

As our society becomes more and more dependent on information systems, so too will the need to protect this information. Many existing and emerging technologies are available to protect information, but the proliferation of technologies can make it difficult to sort out what will best serve the needs of an organization. From what I have observed, the health care community is not particularly sophisticated in evaluating information technologies. Moreover, the vendors tend to be offering "closed" systems that violate my basic assumptions about future architectures.

The protection of computerized medical information will require attention be paid to several areas, including access control, user authentication, remote system authentication, data authentication, data encryption, data replication (backups), audit trails, and timestamp information. In this rather long section I will attempt to describe how some of these technologies will play a role in the future.

3.1 Access Control and User Authentication

It is generally acknowledged that user authentication can be accomplished by using some combination of something known by the user (a password or PIN), something possessed by the user (a badge or token), or some unique feature of the user (e.g., a voiceprint). Very strong authentication methods are often built from a combination of these, such as when you present your ATM card at an ATM machine. In this case you generally present two things: the card itself (something you possess) and the PIN (something you know). The advantage of this is that if someone steals your card, then they still need to know your PIN in order to impersonate you.

Passwords The most common form of authentication used for multi-user computer systems is passwords. Passwords have so many problems associated with them that they can no longer be recommended as a reliable method of authentication for users. The biggest problems are:

- passwords are hard for most people to remember. This causes headaches for users and administrators alike. Furthermore, people write down their passwords in public places, or choose common passwords that are easy to guess.
- passwords are subject to eavesdropping. Many machines on the Internet have been attacked by this method. It is now almost universally accepted that passwords are a completely insecure means of user authentication on public networks.
- passwords are too easy to share between people, making it very difficult to enforce personal responsibility as part of a security policy.

Passwords can still be a strong component of user authentication, but they are best when combined with another method such as an intelligent token or biometric authentication technique. For information on intelligent managment as password usage policies, see [1].

Biometric user authentication In recent years the field of biometric identification has developed into an acceptable alternative for user authentication. Simply stated, a biometric identification technique is a measurement or observation of a feature or action of a human, for the purposes of uniquely identifying the human. Biometric identification systems have been built based on a variety of techniques, including fingerprints, handwritten signature dynamics, voice print analysis, the pattern of blood vessels in the retina, hand geometry measurements, and facial characteristics.

A biometric identification technique can fail in one of two ways: it can either reject the identification of a legitimate user, or else it can incorrectly admit an illegitimate user. Most biometric identification systems allow adjustments to the technology in order to control these error probabilities to some degree. Set the systems to be too sensitive, and many legitimate users will fail to be identified. Set the systems to be too lax, and too many illegitimate users will be accepted. There is usually a tradeoff between these settings, where reducing the false rejection rate will simultaneously raise the false acceptance rate. The time for verification may also be affected. Sandia National Laboratories has for some time engaged in testing of biometric identification devices, in support of their mission of securing nuclear facilities for the U.S. government. At the time of their last announced results, the hand geometry devices were found to be quite reliable and surprisingly accurate (these are now in use in the San Francisco airport). The voice recognition devices were found to be the least reliable, but they have the nice feature that the measurements can be taken remotely using only a telephone or microphone. In general, none were found to be accurate enough to be used alone in high security applications. This is a rapidly evolving technological area, and new methods can be expected to be developed in the future. Sandia expects to continue testing devices as they become available.

From a security standpoint, some forms of biometric identification suffer from one of the major problems with passwords: they can be vulnerable to replay attacks. If an adversary records your voice then they can later use it to impersonate you unless the system is designed to counter this (for example, with a challenge-reponse protocols 3.1).

Tokens and Badges for User Authentication Since most work environments require employees to wear some kind of identification badge, it is natural to expect that this badge might be useful for user authentication in a medical information system. When combined with either something the user knows or a biometric identification method, this can be an extremely effective method of user identification. Tokens can take a variety of forms, including cards of various sizes and other devices. Cards can also incorporate a variety of technologies, including pictures, magnetic strips containing secret or biometric information, or even a microprocessor and memory in a very small package. Because of the opportunities to use this processing power for other applications (particuarly data authentication), I will delay discussion of tokens until section 3.3.

One-time passwords In order to combat the problems of sharing and eavesdropping on passwords, one can employ a method known as "one-time" passwords. The idea here is that each time a user authenticates themselves to the system, they supply a different password. This new password generally needs to be computed for them by software or hardware that they carry with them. There are a variety of methods for doing so, including a variety of software and/or hardware tokens. They generally fall into two categories: a challenge/response system or a time-based algorithm. In the challenge/response system, a challenge is issued by the system, which is then used to compute the appropriate response. The response can be computed via a token, an automatic program, or a piece of user software. The only time-based algorithm that I am aware of is called SecurID, and is marketed by Security Dynamics. In such a system, the user carries a special token that displays the correct response of the moment. There are several such tokens sold, and one requires a PIN to be entered into the device before it displays the correct response. The time-based system carries the advantage that the user is not required to handle a challenge, but carries the disadvantage that it may be vulnerable to replay attacks for a short time period unless properly implemented. It is also a proprietary technology that limits the interoperability, choice of tokens, and administrative tools.

Network Authentication: Kerberos When a doctor walks up to a workstation to access the record of a patient, it is almost certainly going to be the case that the data will not be stored locally within the workstation, but will instead be across the network someplace. Because communication will need to take place between the client workstation and the server system where the data actually resides, the user will need to authenticate themselves to the server system, and not the client system. If a password is used, then it will be transmitted across a network in cleartext, and open to eavesdropping. [1] One system that is designed to address this problem is the Kerberos system developed at MIT. The major drawback of Kerberos is that it requires a common security policy among all systems using it, and a common "key server" for supplying encrypted tokens. This can cause problems in a heterogenous environment where machines are constantly entering and leaving the network, and can also be a performance bottleneck.

The Role of Firewalls A firewall is a term used for a "barrier" between a network of machines and users that operate under a common security policy and generally trust each other, and the outside world. In recent years, firewalls have become enormously popular on the Internet. In large part, this is due to the fact that most existing operating systems have essentially no security, and were designed under the assumption that machines and users would trust each other.

There are two basic reasons for using a firewall at present: to save money in concentrating your security on a small number of components, and to simplify the architecture of a system by restricting access only to machines that trust each other. Firewalls are often regarded as some as an irritation because they are often regarded as an impediment to accessing resources. This is not a fundamental flaw of firewalls, but rather is the result of failing to keep up with demands to improve the firewall.

There is a fairly large group of determined and capable individuals around the world who take pleasure in breaking into systems. Other than the sense of insecurity that it has instilled in society, the amount of actual damage that has been caused is relatively slight. It highlights the fact that essentially any system can be compromised if an adversary is determined enough. It is a tried and true method to improve security within DOD projects to have a "black hat" organization that attempts to break into systems rather than have them found by your real adversaries. By bringing the vulnerabilities of systems to the forefront, the Internet hackers have essentially provided this service, and an impetus to improve existing systems. It is probably a stretch to say that we should thank them, but I believe that it is better to raise these issues early rather than later when our society will be almost 100% dependent on information systems.

[1] This is a very real problem. Software to do this on an ethernet is readily available for personal computers, and is actually shipped as a diagnostic tool on many UNIX workstations.

3.2 Data authentication

We are all familiar with the notion of data authentication in the form of the handwritten signature applied to the paper record. It is useful to ponder for a moment what properties a handwritten signature has that has resulted in it being accepted for data authentication for paper-based documents:

- once a handwritten signature is affixed to a document, it is supposedly difficult to transfer that signature to another document.
- once a handwritten signature is affixed to the document, the document is supposedly difficult to change.
- a handwritten signature can supposedly only be created by that individual.
- many people are able to compare two handwritten signatures and decide if they were created by the same person.

A method to protect electronic information should share these properties to the maximum degree possible. One approach that has been proposed is to "digitize" a picture of a handwritten signature and append this picture to electronic documents. Unfortunately, this is essentially useless because this kind of "signature" does not depend on the document, so that it does nothing to authenticate electronic information. Systems of this type have been offered for sale in the past, but they offer essentially no security whatsoever, and serve only to make people feel comfortable with something that "looks good".

Digital Signatures Note that in practice, a user could "sign" a digital document by encrypting it with their secret key and a traditional symmetric encryption method. The disadvantage to such a system is that the verification of the signature would require knowledge of the secret key, and would therefore render other documents vulnerable to forgery by the party verifying the signature. In order to overcome this, a technique called *digital signatures* has been developed that uses a newer form of cryptography called "public key" or "asymmetric" cryptography, where there are two keys that are mathematically related to each other. Generation of digital signatures requires knowledge of the secret key, but verification of signatures requires only knowledge of the public key. A digital signature for a document is then a "string of bits" that is derived from two pieces of information: the document itself, and the secret key of the signing party. The public key is distributed freely to anyone who wishes to verify signatures. The other key is a secret key known only by the party that is authorized to generate the signatures associated with that public key. For more information on public key cryptography, see [5]. To summarize, digital signatures have the following properties:

- The signature can only be created by someone who knows the secret information.
- Once a signature is created, anyone can validate that the signature is associated with the individual and the document, without needing to know the secret information used to create the digital signature.
- Since the signature depends on the document, it cannot be used to validate a different document.
- It is impossible to produce a document that differs from the original document by even a single bit but for which the signature is still valid.

In this way, the digital signature replicates the desirable features of a handwritten signature, and offers even stronger forms of authentication if the proper procedures are followed in the handling of secret information. For example, handwritten signatures can be verified only by experts practicing what can only be described as an inexact science. By contrast, the mathematical procedure for verifying digital signatures can be verified by any number of independent agents, and there is no room for disagreement among these agents. The numbers are either correct or they are not.

There are now at least two common methods of generating digital signatures. One is a proprietary technology of RSA Data Security, Inc., and is called the RSA digital signature, after it's inventors Rivest, Shamir, and Adleman. This technology is being deployed in a large number of applications, and has been licensed by nearly every big telecommunications and computer company in the U.S, including AT&T, Apple, IBM, Microsoft, Novell, and Sun Microsystems. The only serious competitor to RSA signatures is that of the Digital Signature Algorithm (DSA), which was proposed by NIST as a Federal Information Processing Standard called the Digital Signature Standard (DSS) [7]. There is some controversy surrounding the licensing of DSA, since the holders of the RSA and other patents claim that DSA is covered by their patents. NIST has declared that "The Department of Commerce is not aware of any patents that would be infringed by this standard" [7]. In the next few years this point will become moot anyway, as many if not all of the relevant patents will expire by the year 2000 anyway.

RSA and DSA signatures use similar kinds of calculations, and these can be performed by a number hardware and/or software solutions. They rely upon the notion of a "Key Certification Authority" (CA) that is responsible for issuing and/or certifying keys. The primary role of a key certification authority is to provide assurance that a user's public key is accurate. The CA need not be online to answer questions about the legitimacy of keys. Instead, when a user is issued (or chooses and registers) their public/private key pair, the CA simply issues a digital signature of these keys to certify that these should be recognizable as having been issued to the particular user. These credentials generally have expiration dates and may convey other information such as a role for which the user is certified. When a digital signature is generated by a user, the credentials for the keys used to create the digital signature may be included with the digital signature (but this need not be the case). A user may obtain multiple credentials for their public key, and there is no need for the user to obtain multiple public/private key pairs for multiple applications. In particular, there should be no need that the public/private key pair of a user cannot be used for signing health documents as well as their email or their bank card transactions. For a user that accesses a system held by a given hospital, they may get credentials for their public key from the hospital itself. In order to access the system in the role of a nurse, they may want to present credentials for the same keys that were issued by an accreditation agency for nurses.

In addition to basic key certification through digital signatures under the certification authority's own public key, a key certification authority can provide:

- construction and issuance of private and public keys for users.
- recertification of keys whose certificates have expired.
- issuing key revocation lists for keys that have been compromised.

There is *no reason* to have secret user keys stored on the certification authority machine, and there are good reasons for them *not* to be stored there.

For a "raw" digital signature without certification credentials, we should expect the signature to only about 100-300 bytes, depending on the level of security chosen, but not depending on the size of the message. When a certificate chain is added to the signature, the credentials will grow to a size that is proportional to the length of the chain. It is therefore advantageous to set up key certification hierarchies that are not too deep.

Biometric User Authentication vs. Data Authentication Data authentication and user authentication are closely linked in the sense that it is often an individual (e.g., the author) that certifies the authenticity. In the case of a digital signature system, there needs to be a way to associate an individual with their public key. The usual way of doing this is through the key certification authority, who verifies the identity of the individual through some independent means. The user can only sign information by remembering the secret key and using it in the signing process, but they no longer need the certification authority to be available for the signing or verifying process.

An alternative procedure might be to use biometric authentication techniques to identify the user at the time of signing. Such a system can be built using an on-line signing service to accept the biometric measurements, and if the identification is positive, to use the secret key on behalf of the user to generate the cryptographic binding of the document to the identify of the user. This has the advantage that the user need not be responsible for keeping secret keys, but it has the disadvantage that the signing authority needs to be online.

A more serious potential problem is that the "template" against which the biometric measurement or observation is compared is sensitive information, and cannot generally be made public because it can open the system to impersonation attacks. This is essentially the same problem as that of passwords, that the information used to verify identity is sensitive, and needs to be protected. This implies also that unless public key cryptography is used, the signing authority needs to be online whenever the authenticity of a signature is validated. This severely limits the transferability of authentication information from one system to another, and can potentially lead to a bottleneck and single point of failure in large systems. This will be a big problem with a distributed network consisting of machines within different administrative/security domains. I therefore believe that biometric authentication is better suited to user authentication than data authentication.

3.3 Tokens dumb and smart

As was mentioned in the section on user authentication, one common form of authentication is to rely on something the user possesses, such as a key or token. Tokens can be either smart or dumb, depending on whether they contain processing power. Examples of dumb tokens are standard magnetic strip cards, picture badges, or keys. While they can be quite useful for user identification, dumb tokens are fairly useless for data authentication functions.

Tokens with processing power and storage capabilities can take various form factors. These include:

smart cards In this paper I use the term "smart card" to describe an ISO-standard smart card with the exact same size and shape as a standard credit card, but with an embedded microprocessor and I/O channel. France seems to be ahead of the rest of the world in developing smart card technology, and they are starting to be used for several applications. ISO standard smart cards can be inserted thousands of times into a standard reader that connects to a serial port of a personal computer. They are extremely convenient, since they can easily be carried in a wallet, and can be combined with a magentic strip or printed picture on the card to create a very secure identification device. The major disadvantage at present is that there are relatively few reader devices in the US at present, but the reader devices are extremely cheap, costing less than $100 each.

PCMCIA cards Unlike smart cards, PCMCIA cards are much thicker, and have much more flexibility in the kind of technology that can be incorporated into the cards. A type II card (the most common) is approximately the same dimensions as a credit card, with the exception that they are about 1/8" thick. Almost all laptop computers are now being sold with at least one PCMCIA slot in it, so the technology is becoming fairly common. One disadvantage of PCMCIA cards is that the technology is not intended for as many insertions into a reader as an ISO standard smart card, since it relies upon pins rather than smooth contacts. Another disadvantage is the fact that PCMCIA cards are inconvenient for carrying in wallets.

SmartDisk[2] The SmartDisk looks exactly like a standard 3.5" floppy diskette, but instead of containing a spinning magnetic medium, it uses a magnetic read/write head to communicate with a host computer. Inside the 3.5" floppy disk is space for quite a bit of sophisticated electronics, including batteries, megabytes of memory, and powerful microprocessors. Fischer now markets a card that can be used to control the booting of workstations, but they are also developing a cryptographic module. The major advantage of the SmartDisk is that virtually every desktop computer in existence has a reader already in place that can read the SmartDisk. The SmartDisk is somewhat larger than a smart card however, and must therefore be considered somewhat more inconvenient.

3.4 Key Management

The only reasonable ways to protect the integrity and privacy of information is to rely upon the use of secret information in the form of private keys for signing and/or encryption. The management and handling of these pieces of secret information is generally referred to as "key management". This includes the activities of selection, exchange, storage, certification, expiration, revocation, changing, and transmission of keys. Most of the work in managing information security systems lies in the key management.

Key Management Issues for Public Key Cryptography The concept of public key cryptography was most appealing because it greatly simplifies some of the problems involved in distribution of secret keys. When applied to encryption, it allows a person sending a message to send a message that can only be read by the receiver, without having a need for the sender and receiver to agree on any secret key. The reason for this is that in public key cryptography, the key used for encryption is different from the key used for decryption. In practice, the methods that have been developed for realizing public key encryption are comparitively slow, and public key cryptography is generally used for encrypting "session keys" that are then used for a faster traditional single-key encryption method such as the Data Encryption Standard (DES).

In addition to the convenience of key management for encryption provided by public key cryptography, it also provides a means to implement digital signatures. The separation of public and private keys is exactly what is required to allow users to sign their data (with their secret key), allow others to verify their signatures with the public key, but not have to disclose their secret key in the process.

The secret key then provides the link between the public key and the individual, and will remain a valid link only if the user properly maintains the secrecy of the private key. If for some reason a user's secret key for a digital signature scheme is compromised, then the public key may need to be revoked. If it is known when the private key was compromised, then there is no need to invalidate all of the documents that were signed prior to this date.

The Real Importance of Key Escrow In April of 1993, the United States Government announced a program called "clipper" for key escrow, designed to preserve the government's ability to eavesdrop on telephone and data transmissions over domestic networks. The proposed mechanism required that individuals split their secret encryption keys into two pieces, and hand them over to escrow agents of the government's choosing. In the event of a court order, the escrow agents would release their shares of the secret key in order to allow eavesdropping on encrypted communications. This is coincidentally related to a well known principle in handling of launch codes for nuclear weapons, where two parties are required to collaborate in order to produce the secret information.

The government's clipper program sparked an enormous controversy about the rights of individuals vs. the rights of government, and this controversy races on. Unfortunately, this controversy served to make "key escrow" a dirty word among those concerned with privacy rights, and obscured the fact that the principle of key escrow is a crucial element to the protection of information.

Probably the best way to destroy information is to encrypt it with a secret key and then throw away or lose the key. If we value information, then we should be concerned as much with the threat from failure of access as much as we are concerned with the threat to open access. When files are encrypted, it is assumed that we will someday want to retrieve the information. If the decryption key is held in only one place, then it is extremely vulnerable to catastrophic loss. People die, buildings burn down, and people simply lose information. Luckily, it is fairly simple to set up a split knowledge key escrow system, in which you break your key into two pieces so that knowledge of only one piece makes it no easier to decrypt the information. These keys can then be given to trusted third parties as a hedge against the situation where the original keys are accidentally lost. Several companies will soon announce services in this area, with the intent to provide a service for organizations that are concerned about losing data. I expect that this will become a very common service in the future as we come to use cryptography more routinely in business.

3.5 Audit Trails

If we are to come to rely on computerized record keeping systems for storing critical care information, then we must balance the need to protect this information against prying eyes against the need to provide access in emergency situations. As soon as someone's health is harmed as a result of security procedures keeping information out of the hands of health care providers, then the system will have failed in its primary function. For this reason we must expect that ultimately there will be procedures to allow access to clinical information in case of emergency. The definition of such an emergency access incident is a matter of local policy, but we need to provide for this capability.

The mechanism for accomplishing emergency access is also dependent on the local resources and policies. One approach that I find appealing it to allow authorization of the emergency access by a different legitimate user. In such a situation, if a user has lost their key, forgotten their PIN, or cannot access the system for some other reason, a third party that is themselves a legitimate user may step in an authorize access to the system for the user. The ability to authorize access by others should be severely restricted by policies in order to prevent abuse. A factor that can be used to enforce this is to insist that both users share responsibility for actions taken under such a situation, and to label all entries made under such an access, to be verified later when the user is reinstated in the system.

In order to provide a continuing sanction against improper use of information in a medical information system, we need to combine legal sanctions with access control. These legal sanctions can only be enforced if there is an adequate trail of evidence that documents the misuse of information. An important part of this trail of evidence can be found in properly constructed access logs. Whenever a part of the record for an individual is accessed, a log entry should be made that records exactly what information was accessed, who accessed it, where they accessed it from, and for what purpose. Such logs may themselves become a large database, but they should be considered as important as the primary medical information database. Access to these logs needs to be protected in a manner similar to the primary record, because access patterns to records may in themselves convey information about patients (e.g., if they are seeing an HIV specialist).

The notion of a digital signature can prove extremely useful to guarantee the accuracy of audit trail logs. For example, if it is claimed that a record was accessed by a particular user, it makes sense to demand that a digital signature be created by that user in order to provide access, and this digital signature should then be countersigned by the authority that provided access to the records. In this way, the audit logs cannot later be tampered with by either party in order to make it appear that access either did or did not occur.

3.6 Digital Notary Timestamp Services

Ideally, audit trail logs should show the *order* in which access events occur. In a distributed system it is not obvious how to guarantee the exact sequence of events that occur across the network, but it is relatively easy to cryptographically chain a given access event on a single server to the previous event on that server. This can be accomplished by cryptographically hashing the previous event, combining this with the current event, and hashing the result as the input to the next event.

The chaining of events across multiple servers can be accomplished by having a single timestamp server that everyone contacts to get their events sequenced. It does not take much imagination to see that this technique will soon become unwieldy when it is used to timestamp for every small auditable event. Timestamping across distributed networks is probably best accomplished by the use of a proprietary technique discovered by Stuart Haber and Scott Stornetta at Bellcore [4], [2]. They designed a method in which messages are hashed together across different machines in a tree-like structure, producing common hash values that closely link the sequence of events between different legs of the tree. Bellcore created a new company called Surety Technologies to develop this technoloy. Information is available about their services on the Internet at `http://www.surety.com`.

An important aspect of digital signatures is their temporal characteristics. Certificates for public keys generally have an expiration date, at which point the user needs to have a new certificate issued. There are several good security-related reasons to do this, as well as policy reasons. One of these is the fact that the ability of a user to generate a certain kind of signatures is often tied to their professional certification. For example, a physician has to be certified to practice in a given state, and their ability to sign prescriptions or reports is generally tied to their continued certification. It is natural to tie the certification of their signing keys to their certification in the role that they play in the system.

In order for digital signatures to provide any form of nonrepudiation, it is a good idea to have them incorporate timestamp information. The reason for this is that if a user chooses to repudiate their signatures (say on the basis of arguing that their secret key was compromised), then they should not have complete freedom to repudiate individual signatures. In addition, if it is discovered at some time that a user's secret key was compromised, then it may be necessary to issue a revocation of signatures generated after that date.

4 Evaluation and Design of Appropriate Security

4.1 What are the risks?

In a computerized medical information system, we should be clear about what the risks are for the parties involved before we decide on an effective strategy to implement security. After weighing these risks, we can then assess the seriousness of these risks and the potential for these risks to be experienced. Only then can we design an appropriate security mechanism.

Risks that can be recognized immediately include:

- adverse health affects due to improper records, including possible loss of life. If a record contains incorrect information, then the patient may receive incorrect treatment and suffer ill health effects. This is a problem that exists already in paper-based systems, and it seems unlikely that it will be exacerbated by going to an electronic system.

- liability of providers. If incorrect records result in adverse health effects or other personal injury, then the party responsible for collecting, storing, and displaying this data may be held liable. When the medium of storage and delivery of medical information changes to electronic methods, there will be very little case law to guide liability decisions. This implies that computerized medical information will require much stronger methods of authentication and auditing than is currently used in practice.

- personal embarrassment. Our medical information is often some of the most sensitive information that exists about us. It can be an indicator of past or present behaviour that we might not like to be known widely. This is particularly true of celebrities, for whom there is a great deal of interest in medical details.

- loss of opportunity. As medical technology accelerates, we have observed that much more detailed and specific data collection is possible.

At the same time, much more sophisticated (and some dubious) inferences can be drawn from this data. For example, a person's medical information may be used to make decisions about suitability for insurability, employment or promotion, political office, or even as a basis for investment decisions.

4.2 What are the Threats?

In order to design a security system, it is important to know what the potential threats are so that appropriate counter-measures can be taken. It is impossible to predict every potential threat that may exist to a system, but it is also far too expensive and impractical to plan a security system that will protect against every conceivable threat. It is an axiom of computer security that the only completely secure computer is the one that has never been turned on.

Just because it may be *possible* to subvert a system does not mean that it is a credible threat. An example is provided by the Data Encryption Standard (DES), which is an encryption method proposed by NIST (then the National Bureau of Standards). This has become the standard method of protecting unclassified government information as well as the standard method of protecting banking information worldwide. Rumors have swirled around it's design from the very beginning because of suspicion that the National Security Agency crippled the design back in the mid 70's. Lately, a cult belief has developed that DES is completely insecure, but the fact of the matter is that as far as anyone knows, single key DES can only be broken by a determined and fairly rich adversary [8]. As far as anyone knows, triple key double encryption DES can only be broken if it is used improperly.

The kinds of threats that exist for systems vary, depending on the type of system that is deployed. I am most interested in focusing on systems that exchange information with other systems outside of their administrative domain, that are accessible from multiple physical sites, and from multiple access points. In my opinion this is the direction that computerized medical record systems will move in the future, and this is the most interesting system to try and secure. Such systems are much more usable, but have many more avenues of attack against them. For even the simplest closed systems, I believe that the biggest security threats for medical record systems will come from:

- health care participants wishing to rewrite history in order to avoid and/or introduce liability.

- individuals, including nosy coworkers and reporters seeking information about celebrities.
- system administrators and employees seeking to profit from the monetary value of information.
- incompetent system administrators who fail to administer security correctly.
- murderous intent (e.g., to prescribe an improper medication).
- kids (and pseudo-adults) looking for kicks.

In addition to these malicious threats to the integrity of systems, we should not overlook the threats that come from software, hardware, or network failure, or the threats that come from simple human error. Security systems can help in mitigating these threats, if for no other reason than they impose a discipline of thinking about the protection of data.

4.3 What will security cost?

It's needs to be stated up front that security will cost something in dollars, human capital, and ease of use. At the same time, it needs to be said that you cannot devote the same level of resources to the protection of medical records that we devote to protecting nuclear weapons.

For a long time the computer security market has been starved, primarily because the only market has been the government captive market. This market is largely driven by regulation rather than technical needs, and for such uses the assumption was that you would only be networked to "trusted machines". Products generally need to be certified to be secure for a given level according to DOD regulations, and by the time they were certified the products were sometimes obsolete for other technical reasons. Market forces are changing rapidly now that more and more computers are being networked together. We can expect in the future that security will become a more natural element in the design of systems, and that security will not be an add-on that you buy, but rather an integral feature of the system.

For now however, it is usually the case that security products are purchased separately from complete systems. This makes it extremely difficult to have any real security, since real security requires that the entire *system* have protections built into it, rather than stuck on the outside like a bandaid. Buying a "separate" security system for an existing information system that was never designed with security in mind makes about as much sense as buying a yacht and later turning it into a motorhome by adding wheels.

The major costs of security should be incurred during the time of system design and development. Additional costs will come from delivery time consulting for configuration, and continuing operating costs. Initial outlays for software and/or hardware are not likely to be a big cost to the customer. A good rule of thumb for the cost of administration and setup might be 10-15% of the overall system budget, and a somewhat smaller percentage for initial capital expenditures. If you are buying an entire system from a single vendor, then the 10% figure is probably still a useful guideline for how much more a system will cost that incorporates appropriate security features. When buying such a system, you may want to use outside consultants to judge the security features, just as you might use consultants to judge the software/hardware features. Security is complicated, but if you buy a system that is designed properly with security in mind, it should not be overwhelmingly expensive.

5 What Barriers Remain?

Some Legal Issues In theory, digital signatures should carry greater legal weight than handwritten signatures. Unfortunately, many "quill pen" laws are still on the books that only recognize the validity of paper records, and the legal status of digital signatures has not yet been tested adequately in the court system. There is evidence to suggest that if appropriate technology is combined with appropriate policies and procedures, the digital signature will eventually gain even stronger legal status than the existing record for handwritten signatures. In addition, the GAO has issued an opinion that digital signatures using the DSS are acceptable substitutes for handwritten signatures in most dealings with the US government [3]. For further information regarding the legal status of digital signatures, the interested reader is referred to [5, Chapter 3].

There has been very little precedent in the area of liability for software systems, particularly when it comes to security features. Is a vendor better off to make no claims (or effort) toward providing security services in a system, in order to avoid liability claims, or is it better for them to make a good faith effort toward security, only to be compromised later by a clever hacker. In either case a party may be injured, exposing the software developer to liability. My hope is that legal liability does not get turned on its head, resulting in a greater threat to our well being through a lack of widespead tools for information security.

National Policy on Information Security Technologies Computer and communication security technologies have traditionally been regarded as something required only by governments, and treated as munitions for the purposes of export control. It is certainly true that diplomatic and military superiority depends to a great extent upon the ability to protect information. At the same time, we have come to realize that the security of a society depends in large part on its ability to maintain a high standard of living, and business is realizing more and more that this means protecting their information. Unfortunately, the defense establishment is loathe to allow widespread deployment of security technologies, for fear that they will fall into the hands of our military adversaries. The resulting tension serves to inhibit the development of technology for protecting medical information. This is particularly true in the case of cryptography, where there is very little expertise that the private sector can draw upon to secure their data. Moreover, there is currently no effective support structure for research into information security methods to meet the needs of the private sector.

Who is the customer? The ultimate customer for privacy and integrity of medical information is the patient, who is not directly involved in choosing a system to maintain their records. In the pressure to cut costs in health care, there is a very great risk that security will be left out of the design process because the constituency for security is not directly involved. As I mentioned previously, the primary costs should be incurred at the time of system design and development. Unfortunately, market pressures dictate that products need to be rushed to market, and all development that does not help in this goal will be under pressure for cuts. This is how we end up with security getting stuck on the outside after the system has been designed, and leads to higher overall costs, less convenience, and less security.

6 Summary

Many of us look forward with anticipation to the potential benefits to be derived from large-scale distributed medical information systems. The next decade will see a great deal of experimentation, particularly in the organization of information within databases, the design of user interfaces, and the introduction of new technologies such as wireless communication. I believe that we must resign ourselves to the fact that a great deal of the money that is invested will later be viewed as less than successful. It is important during this time to keep clear view of the fact that the ultimate goal is to have universally accessible and transferable records.

The sudden realization that the Internet can be used to make money has been greatly tempered by the simultaneous realization that the Internet is sadly lacking in adequate security methodology. Luckily, money has a way of solving many problems in life, and I expect that the financial rewards possible on the Internet will fuel a concentration on developing better and cheaper security technologies. This is particularly true for the development of digital signatures and a key certification hierarchy infrastructure, since digital signatures are the key to enabling electronic commerce. We have already seen a rapid acceleration in the process of standards development within ANSI X9 and ANSI X509.

The transition to an electronic record will however require some change in habits of the health care community, since things that have seemed routine will no longer be needed, and things that we took for granted under the previous system will no longer be true. It is important that the transition to an electronic record should at the very least preserve the same degree of assurance for the integrity of medical records that the paper record held. Luckily this should be possible, but it will not be free. Vigilance will be required to maintain the momentum toward securing our medical information.

References

[1] Standards on password usage. Technical Report FIPS PUB 112, National Institute of Standards and Technology, May 1985.

[2] Dave Bayer, Stuart Haber, and W. Scott Stornetta. Improving the efficiency and reliability of digital time-stamping. In R.M. Capocelli, A. DeSantis, and U. Vaccaro, editors, *Sequences II: Methods in Communication, Security, and Computer Science*, pages 329–334, New York, 1993. Springer-Verlag.

[3] Comptroller General of the United States. *Matter of National Institute of Standards and Technology — Use of Electronic Data Interchange Technology to Create Valid Obligations*, December 13 1991. File B-245714.

[4] Stuart Haber and W. Scott Stornetta. How to time-stamp a digital document. *Journal of Cryptology*, 3:99–111, 1991.

[5] Office of Technology Assessment. *Information Security and Privacy in Network Environments*, September 1994.

[6] W. Richard Stevens. *TCP/IP Illustrated*, volume 1, The Protocols. Addison-Wesley, 1994.

[7] U.S. Department of Commerce/National Institute of Standards and Technology. *Federal Information Processing Standard FIPS PUB 186, Digital Signature Standard (DSS)*, May 1994.

[8] Michael J. Wiener. Efficient DES key search. Presented at the rump session of Crypto '93, August 1993.

Internal Distribution

George Davidson	1415	MS 0318	(1)
Ed Barsis	1400	MS 0321	(1)
Bert Westwood	1000	MS 0360	(1)
Samuel Varnado	9400	MS 0431	(1)
Sharon Fletcher	9411	MS 0777	(1)
Judy Moore	9415	MS 0777	(1)
Brad Wood	9415	MS 0777	(1)
Richard Craft	9415	MS 0777	(1)
Roxana Jansma	9415	MS 0777	(1)
Mike Sjulin	9417	MS 0777	(1)
Albert Anaya	9417	MS 0777	(1)
Belinda Garcia	9417	MS 0777	(1)
Doug Brown	13221	MS 0811	(1)
Randy Simons	13221	MS 0811	(1)
Joselyne Gallegos	9426	MS 0974	(1)
Art Hale	1424	MS 1109	(1)
Dick Allen	1422	MS 1110	(1)
Peter Gemmell	1423	MS 1110	(1)
David Kravitz	1423	MS 1110	(1)
Yair Frankel	1423	MS 1110	(1)
Ernie Brickell	1423	MS 1110	(1)
Sudip Dosanjh	1402	MS 1111	(1)
Dona Crawford	8900	MS 9003	(1)
Pete Dean	8910	MS 9011	(1)
Ray Cline	8920	MS 9011	(1)
Central Technical Files	8523-2	MS 9018	(1)
Technical Library (5)	13414	MS 0899	(5)
Technical Publications	12615	MS 0619	(1)
Document Processing for DOE/OSTI	7613-2	MS 0100	(2)