Why It’s Easy Keeping Private Data Private

Thomas Ferrin, Ph.D.
Professor, Departments of Pharmaceutical Chemistry and Biopharmaceutical Sciences
Chair, Biological and Medical Informatics Graduate Program
Director, NIH Resource for Biocomputing, Visualization, and Informatics
Definition of Terms

Privacy -
• An individual’s desire to limit the disclosure of personal information.

Confidentially -
• A condition in which information is shared or released in a controlled manner. Organizations develop confidentially policies to codify their rules for controlling the release of personal information in an effort to protect someone’s (e.g. patient, customer, employee) privacy.

Security -
• Any of a number of measures that organizations implement to protect information and systems. It includes efforts not only to maintain the confidentially of information, but also to insure the integrity and availability of that information and the information systems used to access it.
Definition of Terms - continued

Protected Health Information (PHI) -

• An individual’s health information, or data collected from an individual, that is created or received by a health care provider, plan, or clearinghouse related to the past, present or future physical or mental health or condition of an individual, the provision of health care to the individual, or the past, present or future payment for the provision of health care to the individual; identifies or could reasonably identify the individual; and is transmitted or maintained in electronic or any other form or medium.
Patient privacy and confidentially issues are not new! (California Confidentiality of Medical Information Act enacted in early 1982)

However, Computerized Patient Records (CPR), also known as Electronic Medical Records (EMR), and other electronic access to information such as e-mail, the Internet, and the World Wide Web have raised the awareness level of everyone and brought these issues to the forefront of discussion and action.

Result: Health Insurance Portability and Accountability Act (HIPAA)
Recent Federal Legislation

Health Insurance Portability and Accountability Act (HIPAA)
• Authored by Senators Kassebaum (Kansas) and Kennedy (Massachusetts)
• Public Law 104-191, effective August 21, 1996
• Set uniform privacy and confidentiality standards
• Simplified many administrative procedures
• Required all health care providers, health plans, and clearinghouses to adopt these standards and procedures

Status:
• Privacy rule effective April 14, 2003
• Uniform Electronic Transaction Standards effective Oct. 16, 2003
• Security rules not yet finalized -- estimated early 2005
HIPAA Privacy Standards

Rights for Patients:
- Right to privacy of personal health and research records
  - “Protected Health Information” (PHI)
- Right to request access to their own health and research records
- Right to request amendment or addendum of information in their record
- Right to receive an accounting of disclosures of their information

Penalties for breaches of privacy of PHI:
- Up to $250,000 fine and imprisonment for 10 years
- At UCSF, possibly termination of employment
Some Uses of Patient Health Care Information

Patient care
Insurance billing
Epidemiological studies
Scientific research
Control of insurer risk
Maximize provider or insurer profits
Marketing advantage
Data Security Involves Tradeoffs

Confidentiality
• Ensuring that information doesn’t fall into the wrong hands

Integrity
• Ensuring that information is not damaged, corrupted or destroyed

Availability
• Ensuring that information can be made continuously available to the right people, at the right time

Each organization or application must find the proper balance among these competing principles
Security Practices

Organizational Practices:
- Security and confidentiality policies
- Security and confidentiality committees
- Information security officer
- Education and training programs
- Sanctions for violations
- Improved authorization forms
- Patient access to audit logs

Technical Practices and Procedures:
- Individual authentication of all users
- Per user access controls
- Audit trails (user accountability)
- Physical security and disaster recovery
- Protection of external electronic communication (encryption)
- Protection of remote access points (network firewalls)
- Protection from malicious software (malware)
- Authentication of data records
- Single sign-on
- Ongoing system vulnerability assessment
Authentication - Are you who you say you are?

Can be based on one or more of:

- Something you know (e.g. a password)
- Something you have (e.g. a “token” such as an ID card)
- Some physical attribute or “biometric” (e.g. fingerprint, retinal pattern, handwriting, voice identification, face recognition)

Potential problems:

- Someone else can know the same thing (e.g. shared or stolen password)
- Someone else can have the same thing (e.g. shared or stolen ID card)
- Someone else can masquerade as you (e.g. steal and then use the digital version of your fingerprint after the measurement is taken)
Tokens from Security Dynamics
Pretty Good Authentication Example

Use password and dynamic token:
  • Session #1 at 10:05am on 10/27/2004:
    - login: tef
    - password: mysecret
    - token: 18436572

  • Session #2 at 12:16pm on 10/27/2004:
    - login: tef
    - password: mysecret
    - token: 31415928
Biometric Identification

Fingerprints
Palmprints
Voiceprints
Iris or retina scans

Note that identification needs to be based on phenotypic features, not genotypic, since ~1% of the population has an identical twin.

Also note that once a biometric is stolen, you can’t just get a new one issued.
Roots of a Technological Breakthrough...

Awarded 2002 ACM Turing Award (the “Nobel Prize of Computing”) for this work. See http://www.acm.org/announcements/turing_2002.html
Public Key Cryptography (PKC)

A method for two parties to communicate privately without pre-arrangement and over a communications link that may be monitored by a hostile third party.

- Based on the premise that a message can be encrypted with one key and decrypted with another (there are several algorithms to do this)
- Requires that the parties agree on the algorithm to use for encrypting messages and two “keys”, one private and the other public
- Each user publishes his/her public key in some easily accessible place (e.g. on their web site) for the other party to see, while closely guarding the secrecy of their private key
- Useful approach both for transmitting secure documents and for providing digital signatures
Public Key Cryptography - Example

Clear text message

The quick brown fox jumped over the lazy dog.

Encrypted message

Te = f(Tc, K1)
Kadsf ormbvq wergsff sdfgk3ol xuasyasdf9 plqcukg szulws.

Encryption key
1843657299265

Decryption Key
3141592536387

Recovered message

Tc = f(Te, K2)
The quick brown fox jumped over the lazy dog.
Encrypted Documents

Clear text message

The quick brown fox jumped over the lazy dog.

Encrypted message

Te = f(Tc, K1)

Kadsf ormbvq wergssf sdfgk3ol xuasyasdf9 plqcukg szulws.

Tc = f(Te, K2)

Recovered message

The quick brown fox jumped over the lazy dog.

Recipient’s public key

1843657299265

Recipient’s private key

3141592536387
Digital Signatures

Clear text message
The quick brown fox jumped over the lazy dog.

Encrypted message
Te = f(Tc, K1)
Kadsf ormbvq wergsff sdfgk3ol xuasyasdf9 plqckg szulws.

Recovered message
Tc = f(Te, K2)
The quick brown fox jumped over the lazy dog.

Sender’s private key
1843657299265

Sender’s public key
3141592536387
Encrypted and Signed Documents

1. Original document
2. Encrypted with recipient's public key
3. Encrypted with sender's private key
4. Encrypted and signed document
5. Decrypted with sender's public key
6. Decrypted with recipient's private key
7. Verified clear text document
The RSA Cryptographic Challenge

Want to win a $200,000 prize?

- RSA Laboratories regularly sponsors a series of cryptographic “challenge” contests as a way to quantify the security offered by the U.S. government-endorsed data encryption standard (DES) and other secret-key ciphers with keys of various sizes. The information obtained from these contests is of value to researchers and developers alike as they estimate the strength of an algorithm or application against exhaustive key-search.

- The challenges are:
  - Find the factors of a very large number
  - Find the secret key used to encrypt a message using PKC
  - Find the secret key used to encrypt a message using DES (solved!)

- For more information go to
  - http://www.rsasecurity.com/rsalabs/node.asp?id=2091
Other Uses of PKC Technology

Secure (encrypted) transactions on the world wide web
  • The “s” in https://

Verifiable identities of computers on a network
  • Digital certificates

Secure telephone calls
  • Motorola’s Sectera system

Secure network transactions
  • Novell’s NetWare
So if cryptographic technology is so great, then what's the problem?
Answer: Lots of Stuff. Like passwords…

Traditional authentication relies on passwords
• And passwords are notoriously easy to steal or “guess”
  - Passwords need to be easy for a human to remember
    – But if it’s easy to remember, then it’s probably also easy to guess.
    – If it’s a word in a dictionary, then it’s susceptible to a “dictionary attack”
  - Typical passwords are not longer than 6 or 7 characters
    – Even a slow computer can try every possible combination of alphanumerical passwords in ~50 hours, and every possible keystroke password in ~500 hours.
  - If a password is not easy to remember, then the user probably will write it down
    – …like on a Post-it note that they stick on their monitor!
More on Passwords

People tend to re-use the same password for different applications
  • Here’s an approach that has been shown to be quite effective at collecting user names and passwords:
    - Put up a Web site with something interesting on it: porn, baseball scores, stock tips; whatever will attract the demographic you’re after
    - Don’t charge for it, but require that people register with a username and password
    - Collect these – chances are good that someone will use the same username and password that they already use on another system
    - Use this data to break into the target system

People also tend to share their passwords with others, especially when they need help getting work done
Like Malicious Software (Malware)...

Malware has two components:

- **Payload**
  - The part that does the damage

- **Propagation mechanism**
  - **Computer viruses**
    - Computer code that attaches itself to another computer program; once attached it co-opts the program to make copies of itself
  - **Worms**
    - A self-replicating program that does not hide in another program like a virus does
  - **Trojan Horses**
    - Computer code embedded in some “normal” application and designed to fool the user into thinking it is benign, but behind the scenes it is doing something bad.

*Malware is an exceedingly difficult problem to solve!*
Example Trojan Horse

Password authentication program

• Normally prompts a user to provide their username and password, then verifies that the password is correct and, if so, gives user access to the system

• Malware version: Do the same, except also send a copy of the username and password to the bad guys

As a user of this system, you have no idea that anything is amiss!
The Malware Explosion

Malware originally was most often propagated via floppy disks. In 1999 that changed, and e-mail became the propagation mechanism of choice. The ubiquity of networked computers and e-mail means that malware can infect 10 million or more computers in just a few hours (e.g. ILOVEYOU worm).

Antivirus software is only a partial answer, since new malware can always propagate faster than the antivirus signature updates.
Protection from Malware

Organizations must exercise and enforce discipline over user software:
  • Install virus-checking programs (and keep them updated) on all computers
  • Install network firewalls that limit exposure from outside attackers
  • Limit the ability of users to download or install their own software

This area is currently the least addressed and most technically challenging of the recommended Technical Practices and Procedures!

The software industry, especially Microsoft, must take more responsibility
  • Design and implement better security models to ensure proper operation of software (e.g. prohibit e-mail attachments viruses)
Security Practices: Where do we stand?

Technical Practices and Procedures:

- Individual authentication of users
- Per user access controls
- Audit trails (user accountability)
- Physical security and disaster recovery
- Protection of external electronic communication (encryption)
- Protection of remote access points (network firewalls)
- Protection from malware
- Authentication of data records
- Single sign-on
- Ongoing system vulnerability assessment

😊 = challenging, but not insurmountable problem
Conclusions

Until very recently, privacy and confidentiality protections were inadequate, largely because of inadequate organizational practices fostered by little or no incentive for change. The technology exists to address, at least partially, several of the technical issues.

The problems are urgent.

Federal HIPAA legislation and accompanying standards are underway, but more are needed.
Additional Information:


Questions?