Protecting…

• What?
  – Security
    – Protecting information against accident, disaster, theft, alteration, sabotage, denial of service,
  – Privacy
    – Individual’s desire to limit disclosure of personal information
  – Confidentiality
    – Information sharing in a controlled manner

• Against what?
  – “Acts of God”
  – “Evil hackers”
  – Malicious insiders
  – Stupidity
  – Information Warfare

Privacy

• Right to be let alone; e.g.:
  – snooping on Dan Quayle by J. Rothfeder
  – “outing” of Arthur Ashe (HIV),
    Henry Hyde (adultery)
  – celebrity medical problems (Tammy Wynette, Nicole Simpson)

  – … applies mostly to known individuals

Privacy in obscurity

• Right to remain unknown

• Correlation among pervasive databases:
  – census
  – marketing
  – health

Confidentiality

• Use and sharing of information by multiple users at many institutions
• Should be controlled by coherent policy
• Enforced by appropriate technology

• E.g., who may use results of your life insurance physical exam, for what purposes?
NRC “For the Record” Motivation (1996)

- IT is critical to improve the quality and reduce the costs of health care.
- Privacy and security must be resolved if patients are to share sensitive health information with care providers.
- Protect patient privacy while ensuring that providers have legitimate access to information for purposes of care.

Health Information Held by Individual Organizations Can Be Protected

- **Technical practice**: A variety of practices provide effective protection in an operational environment and can be implemented with reasonable effort.
- **Policy and implementation**: Technical mechanisms must be accompanied by organizational mechanisms for developing access and release policies, training workers, and penalizing violations of policy.
- **Incentives**: Health care organizations need proper set of incentives to address privacy and security concerns.

Two Approaches to Protect Privacy

- **Pre-emptive controls**
  - Lock & key
  - “Need to know”

Two Approaches to Protect Privacy

- **Retroactive controls**
  - Community of trust
  - Checking up, not prevention
  - Sanctions

Threat Model

Must understand what you are protecting against:

- Nature: confidentiality, security
- Source: insider, outsider
- Means: tourist, cracker, …, NSA
- Information at risk
- Scale

Credible threats:
- accidental disclosures by insiders
- abuse of record access privileges by insiders
- insider access for profit or spite
- unauthorized physical intruder
- vengeful outsider who seeks to access, damage, disrupt

Authentication and Access

Eliminate undesirable (horrendous) current practices, e.g.,
- all doctors log in as “MD”
- nurses, receptionists use doctor’s account
- four-digit (or six-digit) “id+password”
- all data available to everyone
- no record of who creates, alters or destroys data
- poorly-controlled access from networks, remote sites
System and Software Discipline

• Standard workstations
  – hardware
  – approved software
• Control over networking
• Control over software installation/dissemination
  – viruses
  – network downloads
  – floppy drives
• Testing of security features

Physical Security

• Lock the computer room (wherever it may be!)
• Backups, recovery procedures
  – protect the backup data
  – test the recovery procedure
• Erase the disk when de-commissioning the computer

Observed Security Features

<table>
<thead>
<tr>
<th>Security Features</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical security</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Network security</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Software security</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Administrative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personnel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equipment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical access</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passwords</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Segmentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equipment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical access</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passwords</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Segmentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Observed Security Features

<table>
<thead>
<tr>
<th>Security Features</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical security</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Network security</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Software security</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Administrative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personnel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equipment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical access</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passwords</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Segmentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Segmentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Access control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Segmentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Observed Security Features

Recommended Organizational Practices for Immediate Implementation

- Security and confidentiality policies
- Security and confidentiality committees
- Information security officers
- Education and training programs
- Sanctions
- Improved authorization forms
- Patient access to audit logs

Policies and Governance

- Clearly stated policy:
  - Responsibility
  - Education
  - Data access
  - Guardianship
    - Associating people with their actions (identification, capabilities, temporary access, termination)
  - Enforcement
  - Testing
  - Transparency
- Governance
  - Policy-making body
  - Security officer
  - Buy-in
    - CIO
    - Human Resources
    - Entire community
  - Education

Enforcement

- Auditing
  - Periodic sampling of access logs
  - Users’ ability to check
- Human Resources (Personnel)
  - Emphasize importance
  - Explicit criterion of evaluation
  - Education and training
  - Reprimand, ..., termination for all levels of employees
  - “Public floggings”

Testing

sine qua non
- Monitoring and awareness
- Review of performance
- Auditing
- “Tiger teams”
- Published results

Recommended Security Practices for Future Implementation

- Strong authentication:
  - Single-session passwords,
  - Encrypted authentication sessions,
  - Token-based authentication
- Enterprise-wide authentication (single logon)
- Access validation—to ensure that retrieved information matches user’s access privileges
- Expanded audit trails
  - All internal accesses to information
  - Global audit trails to trace secondary distribution of data
- Electronic authentication of records
Systemic Concerns Regarding Privacy and Security

- Many concerns regarding patient privacy stem from sharing of information among organizations in healthcare industry.
- Existing data flows are largely unregulated and often occur without patient consent or knowledge.
- Possible development of a universal patient identifier could exacerbate such concerns.

Fair Information Practices (Federal Privacy Act, 1974)

- No secret databases that include personally identified information
  - Agencies must publish policies on all databases
- Right to see my information, with ability to correct
- Prevent data collected for one purpose from being used for another
- Agency responsible for reliability and security of data
- Right to sue

HIPAA

- Health Insurance Portability and Accountability Act of 1996 (Kennedy-Kassebaum Bill)
  - Is really mostly about health insurance portability
  - "Administrative Simplification" section led to regulations about health data exchange, privacy protection, etc.

History of Privacy Provisions

- Congress gave itself until Aug 21, 1999 to enact legislation -- it did not do so
- Backup was that Secretary of HHS was to promulgate rules by Feb 21, 2000 -- this was extended because of 70,000 comments
- Rule promulgated Dec. 2000
- Bush administration has put it on "hold", mainly because of cost complaints (among 30,000 new comments)
- Sec. Thompson agreed to issue the rule, Apr. 2001
- Went into effect Apr. 2003/2004 (dep. on size of org.)
- Congress may legislate later, based on experience
- Can be revised yearly; most have been minor.

Other “simplification” issues

- Standards for electronic health care transactions, including detailed data elements
  - unique health identifiers
  - Providers
  - Patients (did not happen, because of privacy concerns)
  - code sets
  - security standards
  - electronic signatures
  - transfer of information among health plans
- Target date: Feb 21, 1998

Sanctions

- Civil penalties for violations of standards: $100/person/violation, max $25,000/violation/year
- Knowing violations of health identifier or deliberate disclosure:
  - $50,000 + 1 year jail
  - $100,000 + 5 years jail if "under false pretenses"
  - $250,000 + 10 years jail if "with intent to sell, transfer or use ... for commercial advantage, personal gain, or malicious harm"
**Principles**
- Allow smooth flow of PHI for treatment, payment, related operations, public interest
- Prohibit flow of PHI for other purposes, without consent of subject
- Fair information practices
  - Allow subject to access PHI (later, excludes psych notes)
  - Allow subject to have records amended for errors or incompleteness
  - Allow subject to know who else uses PHI
- Require persons who hold PHI to safeguard it
  - Accountable for own use and disclosure
  - Legal recourse
- Minimal Necessary Use and Disclosure
  - Few limits on use for treatment, more for other functions

**Limitations**
- Applies only to "covered entities": health plans, providers, insurers, etc.
- But not to contractors, employers, 3rd-party administrators, marketing firms, etc.
- Covered entities must contract with business associates to pass along responsibilities.
- No private right of action.

**Consent vs. Regulation**
- Should we as a society make generally acceptable rules about privacy, or negotiate them privately?
- Current approach is based on
  - Policy: no authorization needed for treatment, payment, health care operations, public health, health oversight, law enforcement, coroners, State reporting, search warrants
  - Individual consent for other uses, incl. research
- Highly asymmetric power relations

**Health Care Operations**
- Quality assessment and improvement activities
- Review competence of professionals, organizations; conduct training; accreditation
- Insurance rating concerning existing (but not proposed) coverage
- Auditing
- Legal Proceedings
  - Added later:
    - Business planning and development
    - Management
    - General administration
    - Fundraising
    - "Internal Marketing"
  - But not research!

**NOT Health Care Operations**
- **These require explicit authorization**
- Marketing
- Sale, rent or barter of information
- Use in non health-related parts of org.
- Rate setting prior to subject’s enrollment
- Employment determination
- Research "to obtain generalizable knowledge"

**So what do researchers do?**
- Get consent from each patient
  - Virtually impossible in retrospect, costly, and introduces selection bias
- Use data from deceased patients
- Use de-identified data sets
  - "safe harbor" or statistician
- Use "limited data sets"
- Persuade IRB
  - "Minimal risk"
  - Impossible to do otherwise
  - Minimum necessary data
What makes data “identifiable”?

- Name, address, phone number, fax number, email address, URL, IP address, social security number, medical record n., health plan n., account n., certificate/license n., vehicle id, device id, biometric id, full-face photo; of patient, family, or employer
- Date of birth, death, treatment, admission, discharge.
- “any other unique identifying number, characteristic, or code”
- “actual knowledge that the information could be used … to identify”
- But, can retain 3-digit ZIP code and year in dates.

Limited Data Sets

- Data use agreement
  - Spells out uses
  - Promise not to disclose or re-identify
- Need to remove all info to make it de-identified except:
  - May retain town/city, state and ZIP code
  - May retain dates

Sweeney’s Cambridge

- 1997 Cambridge, MA voting list on 54,805 voters
  - Name, address, ZIP, birth date, gender, …
- Combinations that uniquely identify:
  - Birth date (mm/dd/yy) 12%
  - BD + gender 29%
  - BD + 5-digit ZIP 69%
  - BD + 9-digit ZIP 97%
- Unique individuals
  - Kid in a retirement community
  - Black woman resident in Provincetown

Problem of “other information”

- Governor Weld’s data found in Mass “de-identified dataset”
- Dates you visited a health care provider (over a lifetime) are probably unique
- Can be used to re-identify you if someone has both de-identified data and other data that link to identifiers

Danger of Re-identification

Protection via generalization
Computational Disclosure Control

- Make sure data cannot be traced back to a set of size < n
  - Generalization
  - Suppression of unique combinations
  - Account for leakage from what has been suppressed; e.g., back-calculating from aggregate statistics
- How to estimate “external information”?
- Every release becomes more external info.

Methods of Generalization/Suppression

- Underlying problem (find minimal generalization/suppression to achieve a level of anonymity) is NP-hard (Vinterbo)
- Mainly heuristic search over space of possible generalizations/suppressions
- Current research: quantify value of lost information and develop methods that minimize this