

Provider Perspective



Michael Bailey
University of Michigan

PREDICT Workshop
September 27, 2005
Newport Beach, CA





What do providers worry about?

- Privacy or Acceptable use policies
- Institutional Review Boards (IRB)
- Finding the governing or responsible body
- Legal Issues
- Practical Considerations



Privacy Policies

- Sensitive, Confidential or Personally Identifiable Information
- What information is and is not collected, how its used, and how long its kept
- Disclosure
- Safeguards



IRB

- The primary goal of the Institutional Review Board (IRB) is to assure that, in research involving human subjects, the rights and welfare of the subjects are adequately protected.
 - reviews all planned research involving human subjects
 - approves research that meets established criteria for protection of human subjects
 - monitors approved research to ascertain that human subjects are indeed protected.
- Network data publication is new idea to IRB. We were shuffled from Medical to Behavioral Sciences to Health and back again.
- Was required for University of Michigan and Merit (MSU, Wayne State, etc.). Facilitated Washington's process.



Who governs these activities?

- Research projects agencies
- Contracting
- Dean or Office of Research
- Office of Technology Transfer
- Provosts Office



- **Electronic Communications**
 - Wire And Electronic Communications Interception And Interception of Oral Communications
 - Electronic Communications Privacy Act (ECPA)
- **Providers (such as Michnet) may carry data for a variety of institutions such as hospitals, libraries, universities, and K-12 organizations**
 - Family Educational Right to Privacy Act (FERPA)
 - Michigan's Library Privacy Act
 - Federal Standards for Privacy of Individually Identifiable Health Information (implements the privacy requirements HIPAA)
- **Some providers (such as Michnet, University of Wisconsin, University of Washington) are public bodies**
 - Michigan's Freedom of Information Act (FOIA)



Practical Considerations

- Don't want to police or restrict network use
- Don't want to be subpoena'd
- Bad publicity
 - Getting sued
 - Student/Customer information leaking (e.g. San Diego State)
 - Exploits (e.g. recent Usenix Security papers)
- Apples are not oranges and not all oranges are created equal
 - Not everyone collects the same type of data
 - Not all data of the same type is collected in the same way
- Data volumes are huge
 - Merit /8 blackhole O(tens of GB) per day
 - Merit Netflow O(ones of GB) per day



Anonymization

- Identify what is sensitive, confidential, personally identifiable
 - In general not sensitive or confidential if you broadcast it already
 - routing information
 - Data sets you already publish through other venues (e.g. Internet2)
 - But even then you have special cases (e.g. blackhole data representing own'd boxes scanning)
 - Mostly worried about identifying WHO doing WHAT
 - IP Addresses generally are considered personally identifiable as they identify a WHO although they are not always (e.g. dial-up/dhcp, multi-user boxes)
 - Payloads are considered sensitive because they describe in detail WHAT a WHO is doing



Anonymization

- Mitigate the sensitive information
 - IP Addresses are bad, but organizations are ok so just kill/hash/substitute the last 8-bits (some have pushed for last 24-bits).
 - Payloads are very bad. Remove them entirely. MAY be able to create hashes of them. Blackhole MAY be released in their entirety. MAY be able to mitigate in other ways.