

Data Transmission

National Syndromic Surveillance Workshop
October 20-22, 2003
New York Academy of Medicine

Bill Lober, MD

University of Washington

Northwest Center for Public Health Practice

lober@u.washington.edu

DRAFT VERSION Oct 13



Objectives

- Introduce data extraction concepts, survey techniques
- Offer some buzzword compliance
 - Data transfer
 - Networks
 - Encryption
- Provide context for CDC standards
- NOT expect you to do this work...
- NOT geek out...



Plan

- Review basic steps
- Automated/manual transmission techniques
- Provide framework to understand proposed solutions & security implications of:
 - ftp, email, sftp, scp, SSL posts, secure tunneling, VPNs, etc...
 - restricting access, firewalls, networks, etc...
 - authentication, certificates, encryption, etc...
- Reliability/monitoring



Review Steps:

- Obtain agreements
- Select data elements
- Implement data transmission
- Process and store
- Retrieve and surveil...!



Obtain agreements

- What framework?
- What expectations?



Select data elements

- What data?
- What standards?
- Practical: what can you get?



Implement Data Transmission

- 2 sides, 2 partners
- Models for data extraction
- Data transfer methods



Models for data extraction:

- Manual, automated, and semi-automated transfers
- Occasional, Daily, and Near-real time queries (batch)
- Real-time interfaces (HL7)



Manual

- Sending
 - Run report & email file
 - Run report & ftp file
- Receiving
 - Cut and paste/import data
 - Pass data through multiple apps
- etc...



Automated

- Data generated daily, hourly, real-time by scheduled process
- Transmitted when generated, or on fixed schedule
- Received without intervention
- Processed/stored when received, or on fixed schedule



Push vs Pull



- Push – sending data to a person, process, or program when it is available
eg., hospital transmits report when run
- Pull – that person, or process periodically requests data (polling)
eg., a program periodically looks in a file directory to see if anything there to process



Semi-automated

- Ability to generate as-needed data
- Some lack of automation on either sender or receiver side
- Useful for manual process to be able to interact with automated process

eg., manual upload of missing data,
manual database queries in addition to
automated analysis



Tool for Automation

- cURL (Command Line URL)
 - <http://curl.haxx.se>
 - Open source
- Lets programs pretend to be users
- Lets servers pretend to listen to users



Open Source



- Many different licenses, terms
 - eg., non commercial only, etc.
 - Some are transitive
- Source code available
 - Do we care?
- Platforms: Windows, Linux, etc.
- Open Source \neq Linux, better, worse...



Background: Data, networks, and encryption

- Networks
- Data communication
- Network security
- Encryption



Networks 1



- Transport Control Protocol
 - Reliable connections
 - No lost packets
- User Datagram Protocol
 - Fast, lossy
 - Audio, video
- UTP – UDP Transport Protocol
 - New... Hybrid



Networks 2



- Internet Protocol
 - IP address - eg., 128.95.120.1
 - Routing information
- TCP/IP
 - Web site access
 - File transfer



Networks 3



- Plumbing
 - Connections
 - TCP/IP or “UDP/IP”
 - Ports (services)
 - http (80), https (443), ftp (21), ssh (22), telnet (23), Windows file sharing (137-139), smtp (25), IMAP (143), POP3 (110)...



Data Communication

- File transfers
 - Reliable transfer of entire files
- Streams
 - Pipe of characters, open in real time
- Messages
 - Open stream or series of short transfers



Network Security

- Listeners vs. Talkers
 - Open ports
- Security strategies
 - Minimize Listeners
 - Identify the participants
 - Directed conversations
 - Encrypt the conversation



Minimize Listeners

- Don't listen...
 - Windows file sharing
 - RPC/DCOM exploit – MS Blaster
 - Buffer overruns
 - Mail, Web servers,
 - Exploratory probes
 - Telnet, many others...
 - Denial of service

(Patch services: Windows update, etc.)



Identify the participants

- Username/pwd
- Public Key Infrastructure (PKI)
- SecureID



Directed conversations

- Firewalls
 - Host based
 - Independent
- Restrictions
 - IP address
 - Ports
 - Protocols



Encryption

- Conversations
 - PKI
- Content
 - PGP



Data transfer methods

- Insecure
- Secure Batch
- Secure Real Time Interfaces
- Reliability



Insecure

- Email
- ftp
 - server
 - client
 - password protected - bad



Secure

- Sftp, ftps, etc.
 - Encrypted authentication
 - Encrypted conversations
- Scp
- SSL
 - http posts



Secure Conversations

- Stunnel
- SSH tunneling
- VPNs
 - Client-server
 - Client-subnet
 - Subnet-subnet



MS Model for VPNs, as examples

- PPTP, L2TP
- Workstation support
- Server support
 - Remote Access Tools
 - IPSEC



Reliability:

- Fault tolerance
- Notification



Fault tolerance

- Inherent confirmation in protocols
- Retransmission
- Restoring interrupted connections



Notification

- Pager/email
- Examples



Testing

- Software testing plan
- Test point relation to notification



Example 1

- Concepts used in Seattle system
 - Hub & spokes – minimal listeners
 - Periodic reporting and HL7 interfaces
 - Authentication, IP restriction
 - https posts
 - sftp, scp, payload encryption
 - Notification/monitoring



Example 2

- CDC standards (PHIN)
 - ebXML/ebMS/SOAP/https
 - Bidirectional messaging
 - Authenticated Services
 - ebXML messages (HL7 2.3, 3.0, etc. content)
 - Encrypted data, strong authentication
 - Translate and manipulate codes, local mapping
 - Meet or exceed HIPAA security standards



Summary

- Introduced data extraction concepts, survey techniques
- Offerd some buzzword compliance
 - Data transfer
 - Networks
 - Encryption
- Provided context for CDC standards

