Email Analysis

Conan C. Albrecht, PhD

Email Analysis

- Today's electronic communication tracks often holds a gold mine of information
- Places to search
  - Local email client(s)
  - Webmail
  - Cell phone memory
  - PDA memory
  - USB drives
  - IM chat logs

What Is Email?

- Simple text files sent between servers on port 25
  - Example with telnet
  - Example of IMAP email file
- Parts of the file
  - Headers
  - Body
  - Attachments (encoded to text)

Getting Email Data

- This can often be retrieved from the corporate mail server
  - MS Outlook and Exchange
  - Lotus Notes
  - Thunderbird (or others) and IMAP Email
- Email from other sources
  - ISP-based Email (usually POP)
  - WebMail

MS Outlook and Exchange Server

- Server-side access
  - Modern Exchange servers store email in a SQL Server database
  - Access is primarily through the administrator interface
  - 3rd party tools exist as well
- Client-side access
  - .pst file contains all email
  - Read with Outlook or 3rd party tools

Lotus Notes

- Server-side access
  - Administrator programs can access email files
- Client-side access
  - .nsf file can be read by many applications
  - Many .nsf -> .pst (Notes to Outlook) converters exist

IMAP Email Access (Thunderbird)

- Server-side access
  - Maildir format: Most popular modern format – each email is kept is a separate text file in the user's home directory
  - Mbox format: Older format – all email files in a folder are in a single text file
  - Reading both formats requires a simple text editor, and searching is easy with regular expressions
- Client-side access
  - Varies depends upon the client used
  - Thunderbird uses mbox format internally

ISP-Based Email (POP)

- Server-side access
  - When the user checks email, all files are transferred to the client
  - No copies are kept on the server, so little exists here
  - This is one of the worst cases because it requires seizing the client computer
- Client-side access
  - Varies depends upon the client used
  - Thunderbird uses mbox format internally

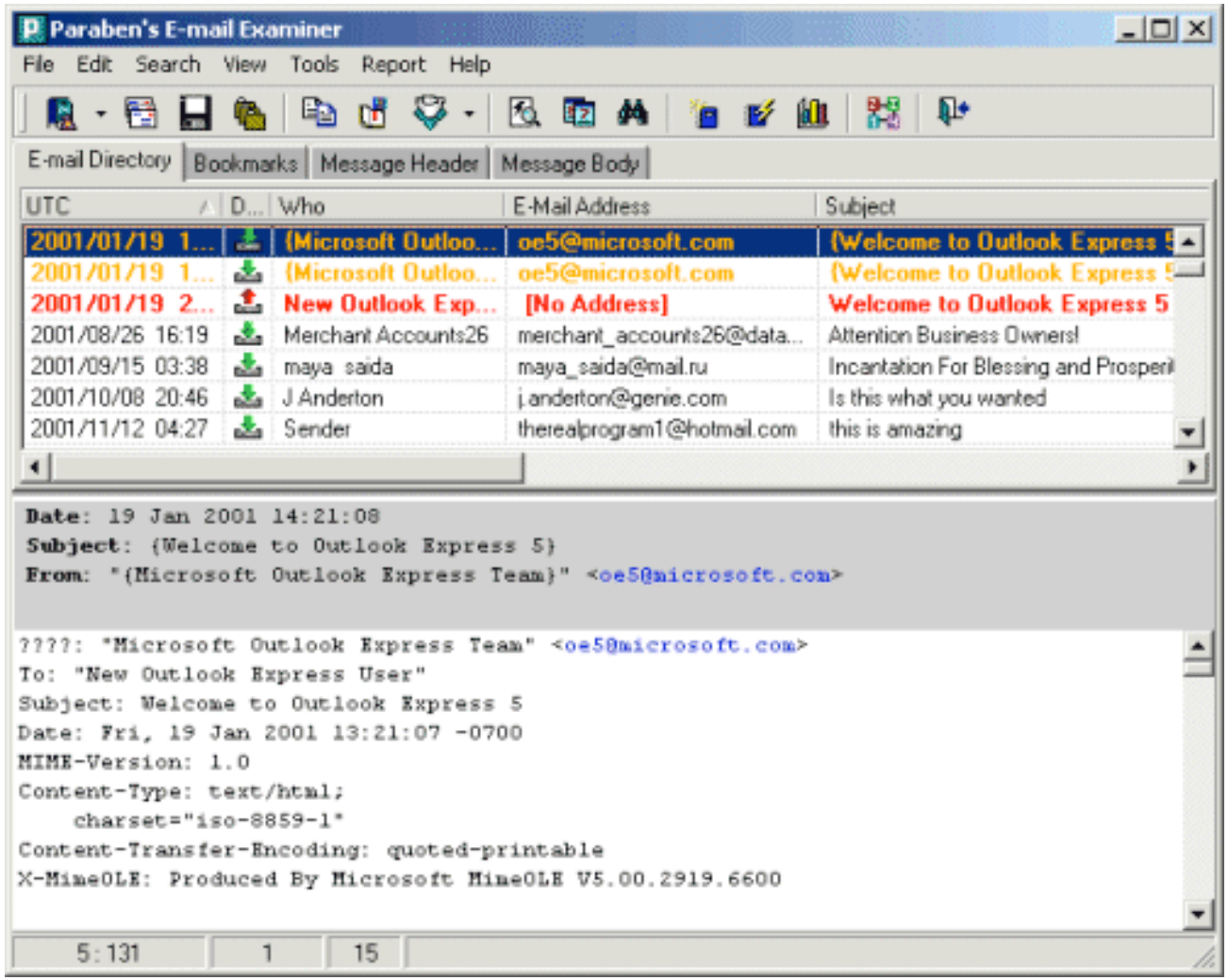
Webmail (Hotmail, Yahoo, Gmail)

- Server-side access
  - All email is kept on the server
  - In 2007, the Sixth U.S. Circuit Court of Appeals said government needs warrants to get emails that are less than 180 days old
  - After 180 days, administrative subpoena or different court order required
- Client-side access
  - Some emails may be in the browser cache
  - Password may be autosaved, but be sure it is legal for you to get in

http://blog.wired.com/27bstroke6/2007/06/appeals_court_s.html

Full Text Analysis of Email

Method 1: Use a Client Like Paraben's Email Examiner



Paraben's Email Examiner

- Reads and searches email from almost anys source (15+ types)
  - Microsoft Exchange
  - Lotus Notes
  - Novell Groupwise
- Reads deleted messages

Method 2: Use Full-Text Searching

- Simple search tools like Windows Search can search files within directories
- More powerful full-text tools exist
  - dtSearch
  - Apache Lucene
  - Sphinx open source engine
  - SQL Sever predicate logic searching
  - MySQL full text searching functions

Method 3: Load Into a Data Analysis Application

- Since email files are simple text, use a script to load them into a data analysis application like Picalo, ACL, or IDEA
- Search using standard data analysis functions like regular expressions