Computer Forensics:
Tool Testing
&
National Software Reference Library

Jim Lyle & Doug White
Information Technology Laboratory
20 August 2003

United States Department of Commerce
National Institute of Standards and Technology
Outline

- Overview of computer forensics at NIST
- Description of CFTT and NSRL projects
- Questions and answers
Computer Forensics Partners

- NIST (OLES)
- DoJ (NIJ, FBI)
- DoD (DCCC)
- TREASURY (USCS, USSS)
- National State/Local Agencies
- Homeland Security
The National Institute of Justice (NIJ) is a major funding source:
- CFTT to date: $3.5 M
- NSRL to date: $2 M

The Program Manager for Forensic Sciences, Susan Ballou, of the Office of Law Enforcement (OLES) at NIST, directs NIJ funding to the appropriate expertise whether within NIST or beyond.
A Shocking Revelation . . .

Computers can be involved in crime . . .

- As a victim
- As a weapon
- As a witness
- As a record
- As contraband
Outline of an Investigation

- Get proper authorization
- Seize evidence (Hard drives, floppies …)
- Create duplicates for analysis
- Analyze the duplicates
  - Exclude known benign files
  - Examine obvious files
  - Search for hidden evidence
- Report results
Investigators Need …

Computer forensic investigators need tools that …

- Work as they should and
- Produce results admissible in court
- Reference data to reduce analysis workload
Goals of CF at NIST

- Establish methodology for testing computer forensic tools (CFTT)
- Provide international standard reference data that tool makers and investigators can use in investigations (NSRL)
Why NIST/ITL is involved

• Mission: Assist federal, state & local agencies
• NIST is a neutral organization – not law enforcement or vendor
• NIST provides an open, rigorous process
Computer Forensics in ITL

Located in Software Diagnostics and Conformance Testing (SDCT) Division
- Includes development of specifications and conformance tests for use by agencies and industry
- Work is funded by Federal agencies and NIST internal funds
- Homeland Security support of agencies investigating terrorist activities
Computer Forensics Tool Testing (CFTT)
A Problem for Investigators

Do forensic tools work as they should?

- Software tools must be …
  - Tested: accurate, reliable & repeatable
  - Peer reviewed
  - Generally accepted

- … by whom?

- Results of a forensic analysis must be admissible in court
CFTT Presentation Overview

- Project Tasks
- Current activities
- Challenges
- Testing Hard Drive Imaging Tools
- Benefits of CFTT
Project Tasks

- Identify forensics functions e.g.,
  - disk imaging,
  - hard drive write protect,
  - deleted file recovery
- Develop specification for each category
- Peer review of specification
- Test methodology for each function
- Report results
Current Activities

- Hard drive imaging tools
- Software hard drive write protect
- Hardware hard drive write protect
- Deleted file recovery
Challenges

- No standards or specifications for tools
- Forensic vocabulary incomplete
- Arcane knowledge domain (e.g. DOS)
- Reliably faulty hardware
Hard Drive Imaging

- SCSI vs IDE
- Drive access
- Clone vs image
- Excess sectors on dst
- I/O errors
- Corrupt image file
Testing Hard Disk Drive Imaging Tools

Need to verify…

• Source disk not changed
• Copied information is accurate
• Behavior if source is smaller than destination
• Behavior if source is larger than destination
Testing Hard Disk Drive Imaging Tools

Testing support Tools

• Detect change
• Compare Source to Destination
• Track relocated information

<table>
<thead>
<tr>
<th>ASCII String</th>
<th>25 bytes</th>
<th>Fill Bytes 487 Bytes</th>
</tr>
</thead>
</table>

8/20/2003
Testing Hard Disk Drive Imaging Tools

Setup Source
Wipe
Load OS
Hash

8/20/2003
Testing Hard Disk Drive Imaging Tools

Select Source
Wipe Destination
Run Tool
Compare Src : Dst
Hash Source

8/20/2003
Impact

- Release 18 (Feb 2001) - A US government organization was doing some testing and uncovered an issue under a specific set of circumstances.
- Linux doesn’t use the last sector if odd
- Several vendors have made product or documentation changes
Benefits of CFTT

Benefits of a forensic tool testing program

- Users can make informed choices
- Neutral test program (not law enforcement)
- Reduce challenges to admissibility of digital evidence
- Tool creators make better tools
Lab Facilities
CFTT/NSRL Team
NSRL Project
Outline

- NSRL Description
- RDS Description
- RDS Use
- Project News
- Your Needs
What is the NSRL?

- National Software Reference Library (NSRL)
  - Physical library of software, 2400 products
  - SQL Server database of known file signatures
  - Reference Data Set (RDS): 16,200,000 file signatures

- Goals
  - Automate the process of identifying known files on computers used in crimes
  - Allow investigators to concentrate on files that could contain evidence (unknown and suspect files)
Addressing Law Enforcement Needs

- LE needed an unbiased organization
- LE needed traceability for the NSRL contents
- No repositories of original software available for reproducing data
- NSRL needs to work with many CF tools
Scope of the NSRL

- NIST has collected software for 2 years
- Software is recorded as the original source for known files and stored as a part of the NSRL
- Versions of OS, DBMS, photo editors, word processors, network browsers, compilers...
- Data formats, data dictionary and project status information is available on the website for RDS users and industry reference
What is the RDS?

NIST Special Database #28

National Software Reference Library

Reference Data Set
Version 2.1 06/02/2003

NIST
What is the RDS?

- Reference set of file profiles
  - Each profile includes file name, file size, 3 file signatures (SHA1, MD5, CRC32), application name, operating system, etc.
  - Extracted from files on original software CDs, diskettes, and network downloads
  - A single application may have thousands of separate file profiles
What is in the RDS?

- “Known” files – not “known good”
- Off-the-shelf, shrinkwrapped programs, documented downloads
- Includes hacker tools, port scanners, network security tools, encryption
- Permutated index available at www.nsrl.nist.gov/index
RDS Use

- Commercial tools import the RDS as a single hash set
- You may need to process the RDS data before importing it
- Perl scripts and unix shell scripts available on www.nsrl.nist.gov
- 4,300 separate hashsets on website
Hashes

- Compute a unique identifier for each file based on contents
- Primary hash value used in the NSRL RDS is the Secure Hash Algorithm (SHA-1) specified in Federal Information Processing Standard (FIPS) 180-1, a 160-bit hashing algorithm
- SHA-1 values can be cross-referenced by other products that depend on different hash values
Hashes

- Other standard hash values computed for each file include Message Digest 4 (MD4), Message Digest 5 (MD5), and a 32-bit Cyclical Redundancy Checksum (CRC32), which are useful in many CF tools and to users outside LE.
- Separate, parallel, and independent process is used to validate the results of the primary RDS implementation.
- Once verified and validated, the RDS is written to a master CD, duplicated, and distributed through NIST’s Standard Reference Data Office as Special Database #28 (www.nist.gov/srd/nistsd28.htm).
# Hash Examples

<table>
<thead>
<tr>
<th>Filename</th>
<th>Bytes</th>
<th>SHA-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>NT4\ALPHA\notepad.exe</td>
<td>68368</td>
<td>F1F284D5D757039DEC1C44A05AC148B9D204E467</td>
</tr>
<tr>
<td>NT4\I386\notepad.exe</td>
<td>45328</td>
<td>3C4E15A29014358C61548A981A4AC8573167BE37</td>
</tr>
<tr>
<td>NT4\MIPS\notepad.exe</td>
<td>66832</td>
<td>33309956E4DBBA665E86962308FE5E1378998E69</td>
</tr>
<tr>
<td>NT4\PPC\notepad.exe</td>
<td>68880</td>
<td>47BB7AF0E4DD565ED75DEB492D8C17B1BFD3FB23</td>
</tr>
<tr>
<td>WINNT31.WKS\I386\notepad.exe</td>
<td>57252</td>
<td>2E0849CF327709FC46B705EEAB5E57380F5B1F67</td>
</tr>
<tr>
<td>WINNT31.SRV\I386\notepad.exe</td>
<td>57252</td>
<td>2E0849CF327709FC46B705EEAB5E57380F5B1F67</td>
</tr>
</tbody>
</table>
Hashing Installed Files

- Currently testing methods for hashing installed files
- Installation of known packages in NSRL onto virtual machines
- Virtual machine state can be preserved on CD on NSRL shelf for repeatability
- Comparison installation on physical machine
Installed Hash Findings

- Installed MS W2K Pro on virtual machine and physical machine; approx. 4,500 files
- RDS identified 79% of files on VM, 60% of files on PM
- Hashed installed files on VM and PM
- VM hashes identified another 5% on PM
- PM hashes identified another 3% on VM
Installed Hash Findings

- Hashed 2 W2K Pro PCs “in the wild”
- RDS, VM and PM hashsets identified 17% of the 4,500 W2K files on the “wild” PCs
- The “wild” PC hashsets identified 80% of the files on each other

- Installed hashes are necessary
- Patch/hotfix/update hashes are most critical
Data Verification

- Multiple and independent techniques from different perspectives
  - We use test files with known signatures
  - Parallel database system: Match results with other system
  - Human verification
  - Database rules and constraints
  - Periodic database queries: Predefined procedures to search for and report anomalies in the database
  - User feedback: Error reports and RDS updates
Project News

- Hashing code (Mar. ’03) available
- Late Sept. – LAMP environment, cookbook
- Peer-to-Peer hashes
- Block size hashes – evidence chain, deleted files
- Multiple language research
- File format NOT changing ever again
- Conversion tools available
- Dec. (?) – database on public internet
- Interesting hashes – 200 steg tools, etc.
Your Needs

- Opinions on “known” vs. “known good”
- What can we do to make your work
  - Faster?
  - Simpler?
  - Easier to explain in court?
- What software do you recommend to be hashed?
- ???
Contacts

Jim Lyle
www.cftt.nist.gov
cftt@nist.gov

Doug White
www.nsrl.nist.gov
nsrl@nist.gov

Mark Skall
Chief, Software Diagnostics & Conformance Testing Div.
www.itl.nist.gov/div897
skall@nist.gov

Sue Ballou, Office of Law Enforcement Standards
Steering Committee Rep. For State/Local Law Enforcement
susan.ballou@nist.gov

8/20/2003