Outline

- Overview of Computer Forensics at NIST
- Computer Forensic Tool Testing (CFTT)
- National Software Reference Library (NSRL)
- Questions and Answers
These organizations fund and provide guidance to CFTT & NSRL
Vendors submit software and provide feedback in peer review
NIST is the technical component
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 an 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 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
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

National Software Reference Library
Outline

- NSRL Description
- RDS Description
- RDS Use
- Research Areas
- Project News & Accomplishments
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?

Reference Data Set
Version 2.1 06/02/2003

NIST Special Database #28
National Software Reference Library
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
- Permuted index available at www.nsrl.nist.gov
RDS Use

- Commercial tools import the RDS as a single hash set
- Users may process the RDS data before importing it
- Software tools available on www.nsrl.nist.gov
- 4,300 separate hashsets on website
How to Use the RDS

- Eliminate as many known files as possible from the examination process using automated means.
- Discover files that do not contain expected contents (.exe file containing a bomb schematic, facility map).
- Look for files that should be installed, but are missing (incomplete deletion of pirated software).
- Look for files that could be suspect (hash matches, but file name does not).
- Provide rigorously verified data for forensic investigations.
RDS Field Use

ANALYSIS PROGRAM

UNKNOWN FILES

KNOWN FILES

40-95%

Disk Drive

RDS
RDS Field Use Example

You are looking for facility maps on a computer which is running Windows NT 4.0 Workstation.

Windows NT 4.0 operating system software contains 6753 images which are known gifs, icons, jpeg files.

By using the RDS and an analysis program the investigator would not have to look at these files to complete his investigation.
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
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
- RDS, VM and PM hashsets identified 17% of 4,500 files on two “wild” PCs

- Installed hashes are necessary
- Patch/hotfix/update hashes are most critical
Use on Actual Machines

Clean OS
- 4622 files, 92% known – w98
- 7720 files, 89% known – w2k
- 5412 files, 93% known – wme

Actual NIST PCs
- 39631 files, 80% known – w2k
- 18262 files, 65% known – w98
- 75834 files, 45% known – w2k, mgmt
NSRL News

- Hashing code (Mar. ‘03) available
- Late Fall ‘03 – LAMP environment, cookbook
- Peer-to-Peer hashes
- Block size hashes – evidence chain, deleted files
- Multiple language research
- Conversion tools available
- Winter ‘03 – database on public internet
- Interesting hashes – 200 steg tools, etc.
NSRL Accomplishments

- RDS CD Version 1.2 distributed 6/6/2002
  - 124 subscriptions (Vendors, corporations, universities, and law enforcement agencies)
  - Free redistribution, NIST traceable
- Incorporated into vendor products
- Used by FBI, DCCC, Secret Service, Customs Service (Homeland Security)
Contacts

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

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

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

9/5/2003