

Computer Forensics:

Tool Testing

&

National Software Reference Library

Jim Lyle

Information Technology Laboratory

9 September 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

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 advertized and
- Produce results admissible in court

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)

In the Beginning

- Houston, we have a problem
- Money, who's got the money?
- Setting up the program

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 meet vendor specs?

- Software tools must be ...
 - Tested: accurate, reliable & repeatable
 - Peer reviewed
 - Generally accepted
- ... by whom?
- Results of a forensic analysis must be admissible in court

Presentation Overview

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

Project Tasks

- Identify tool categories e.g.,
 - disk imaging,
 - hard drive write protect,
 - deleted file recovery
- Develop standards for each category
- Peer review of standards
- Test methodology for each category
- Report results

Current Activities

Evaluating test methodology for ...

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

Challenges

- No standards 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

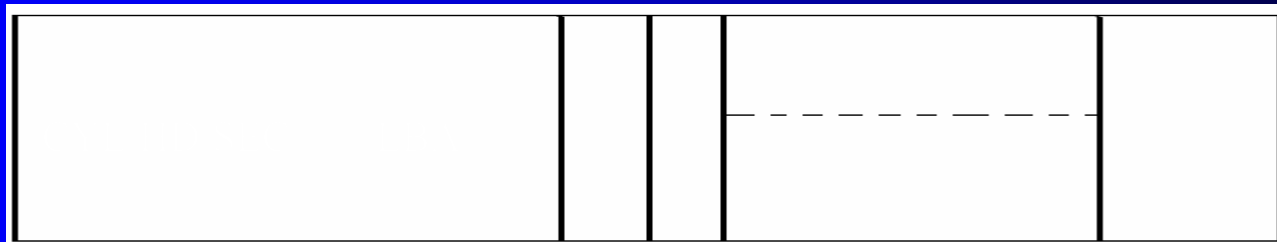
Need to verify...

- 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 SHA-1
- Compare Source to Destination
- Track relocated information



ASCII String 25 bytes

Fill Bytes 487 Bytes

Testing Hard Disk Drive Imaging Tools

Setup Source

Wipe

Load OS

Hash



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

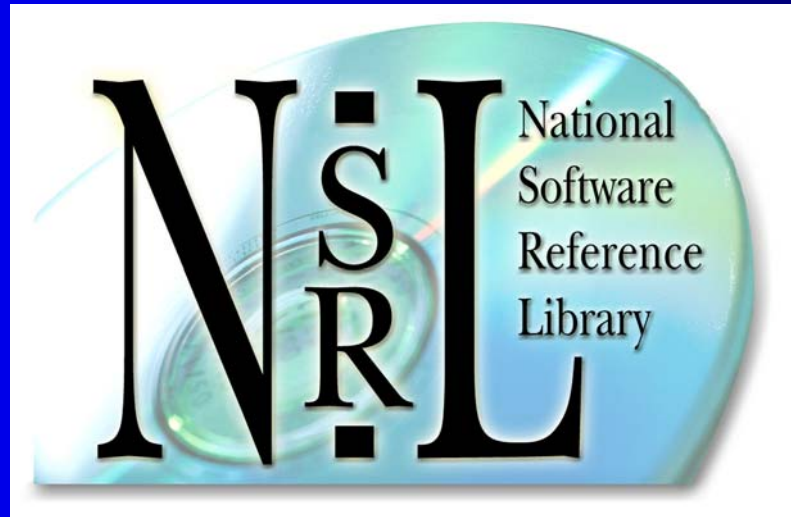
Benefits of a forensic tool testing program

- 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



NSRL Project



Outline

- NSRL Description
- RDS Description
- RDS Field Use
- Hash System Overview

What is the NSRL?

- National Software Reference Library (NSRL)
 - Physical library of software, 1900 products
 - SQL Server database of known file signatures
 - Reference Data Set (RDS)
 - Extract of database on CD: 6,500,000 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

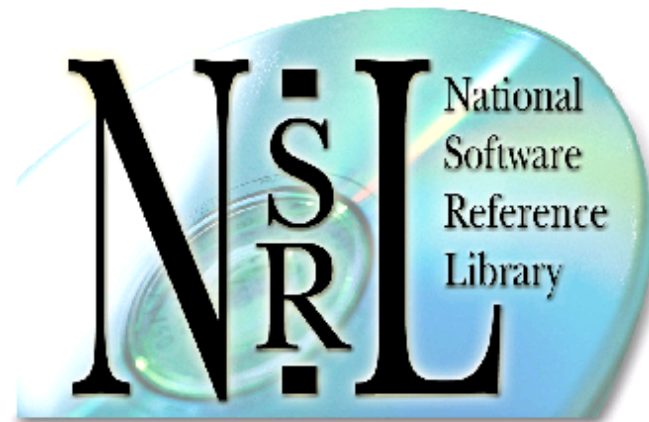
Scope of the NSRL



- 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



Reference Data Set

Version 1.2 06/06/2002

NIST



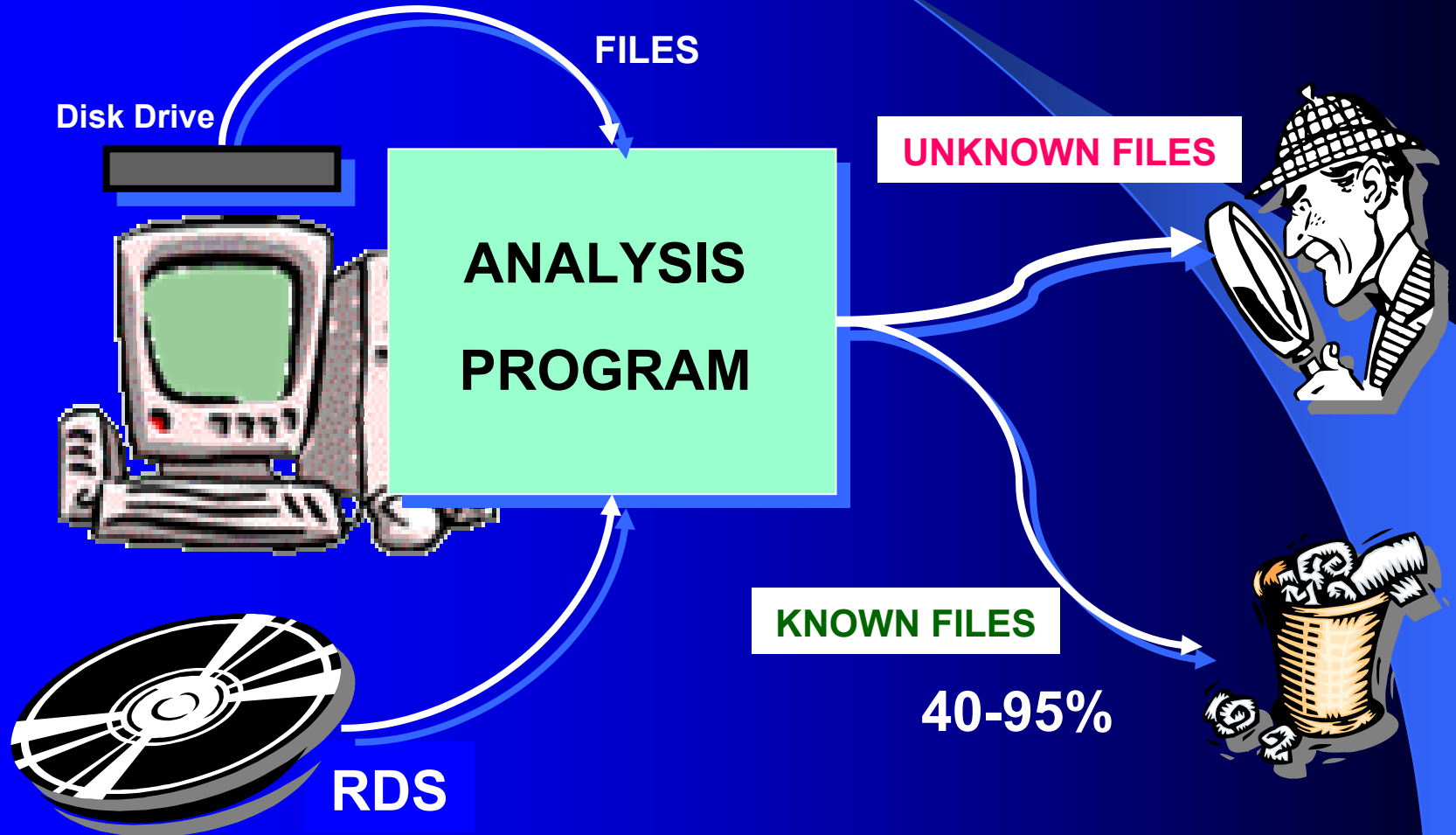
What is the RDS?

- Reference set of file profiles
 - Each profile includes file name, file size, 4 file signatures (SHA1, MD4, MD5, CRC32), application name, operating system, etc.
 - Extracted from files on original software CDs, diskettes, and network downloads
 - A single application may have 10,000 separate file profiles

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



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

e.g.,



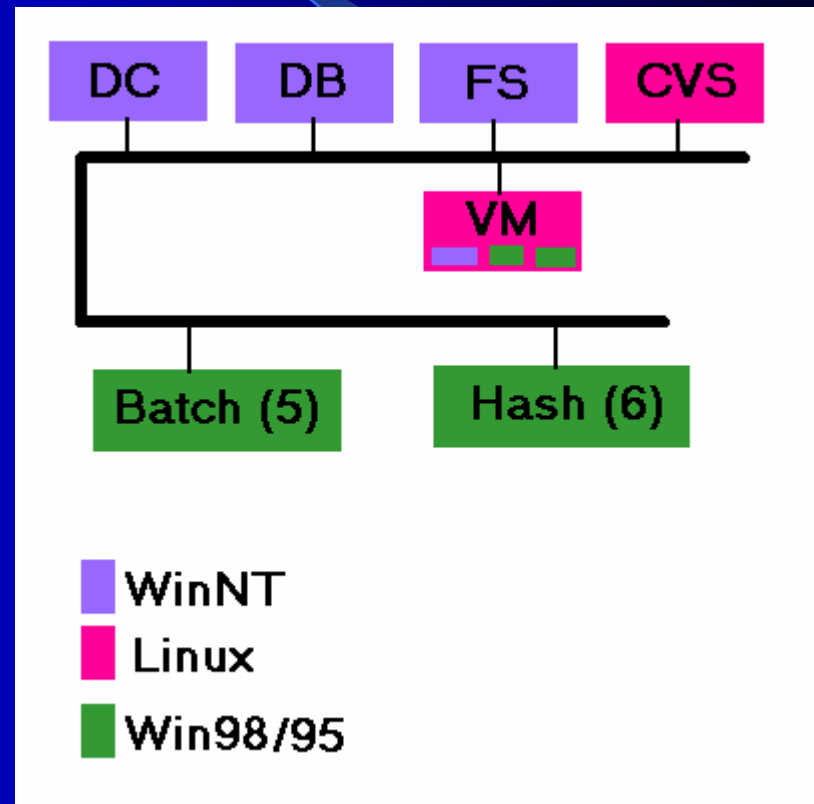
By using the RDS and an analysis program the investigator would not have to look at these files to complete his investigation.

Hash System Overview

- Environment
- Input Process
- Hashing
- Verification
- Future

Environment

- Isolated LAN
- Domain controller
- Database server
- File server
- CVS repository
- Virtual Machines
- Batching stations
- Hashing constellation



General Process

Acquire

Batch

Hash

Verify

Accept/Reject

Input Process


- Package is acquired
- Web interface used to enter information about manufacturer, product, OS and assign an ID
- Media are batched
- Approximately 15 minutes per package

NSRL Package Information - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites History Print Mail News RSS

Address http://192.168.58.4/login/script.asp Go Links



Package Information

NSRL Builder

*Application Name:	<input type="text" value="Your Eyes Only"/>	
*Version:	<input type="text" value="win95 1.0"/>	
Bar Code:	<input type="text" value="037648120487"/>	
*Language:	<input type="text" value="English"/>	Please Specify: <input type="text"/>
Manufacturer:	<input type="text" value="Symantec"/>	
*Application Type:	<input type="text" value="Utility"/>	Please Specify: <input type="text"/>
Packaged Within:	<input type="text"/>	
Comments:	<input type="text"/>	
*Location:	<input type="text" value="G2"/>	

Contact Us:

National Institute of Standards and Technology
 ATTN: NSRL Project
 100 Bureau Drive, Stop 8970
 Gaithersburg, MD 20899-8970 USA

Done Internet

Hashing Process

- July 2002 – accepting software, performing installations
- Currently hashing 5 p.m. until done
- 6 200MHz PCs in hashing constellation
- Averaging 10.5 hashes/sec.
- DVD with 250,000 files needed 30 hours

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

Filename	Bytes	SHA-1
NT4\ALPHA\notepad.exe	68368	F1F284D5D757039DEC1C44A05AC148B9D204E467
NT4\I386\notepad.exe	45328	3C4E15A29014358C61548A981A4AC8573167BE37
NT4\MIPS\notepad.exe	66832	33309956E4DBBA665E86962308FE5E1378998E69
NT4\PPC\notepad.exe	68880	47BB7AF0E4DD565ED75DEB492D8C17B1BFD3FB23
WINNT31.WKS\I386\notepad.exe	57252	2E0849CF327709FC46B705EEAB5E57380F5B1F67
WINNT31.SRV\I386\notepad.exe	57252	2E0849CF327709FC46B705EEAB5E57380F5B1F67

Installation Hashes

- 300-800 files are “missed” by current RDS
- Compare automated hashes with real-world installed hashes
- Compare installed file sets across machines and OS'es
- Installed on virtual machines which can be saved in the NSRL on media

Use on Actual Machines

Clean OS

- 4622 files -360, 92% known – w98
- 7720 files -864, 89% known – w2k
- 5412 files -370, 93% known – wme

Actual NIST PCs

- 39631 files -7902, 80% known – w2k
- 18262 files -6395, 65% known – w98
- 75834 files -41638, 45% known – w2k,mgmt

OS/Apps	Files installed on HD	HD Files not in Hashkeeper	HD Files not in NSRL	Files on distribution CD(s)
Virgin Win 98	4,266	142 (3%)	297 (7%)	18,662
Virgin NT4 WS	1,659	1,211 (72%)	239 (14%)	17,904
Virgin Win 2Kpro	5,963	783 (13%)	839 (14%)	16,539
Virgin Win ME	5,169	2,973 (57%)	383 (7%)	11,512
Win 98+Office 2K	23,464	313 (1%)	596 (2%)	43,327
Win ME+Office 2K	24,112	3,119 (13%)	526 (2%)	32,758
NIST PC #1 W2K	18,048	13,137 (72%)	11,839 (65%)	N/A
NIST PC #2 W2K	59,135	46,277 (79%)	47,124 (80%)	N/A
NIST PC #3 WNT	14,186	7,543 (53%)	6,618 (46%)	N/A
NIST PC #4 W98	16,397	8,360 (51%)	7,404 (45%)	N/A
NIST PC #5 W98	34,220	8,366 (25%)	8,667 (25%)	N/A

Lower percentage is better

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

Future Tasks

- Byte signature file type verification
- Self-extracting EXE files
- Redundant hashing in constellation
- Scheduled rebatching
- Additional algorithms - AES

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)

CFTT/NSRL Team



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