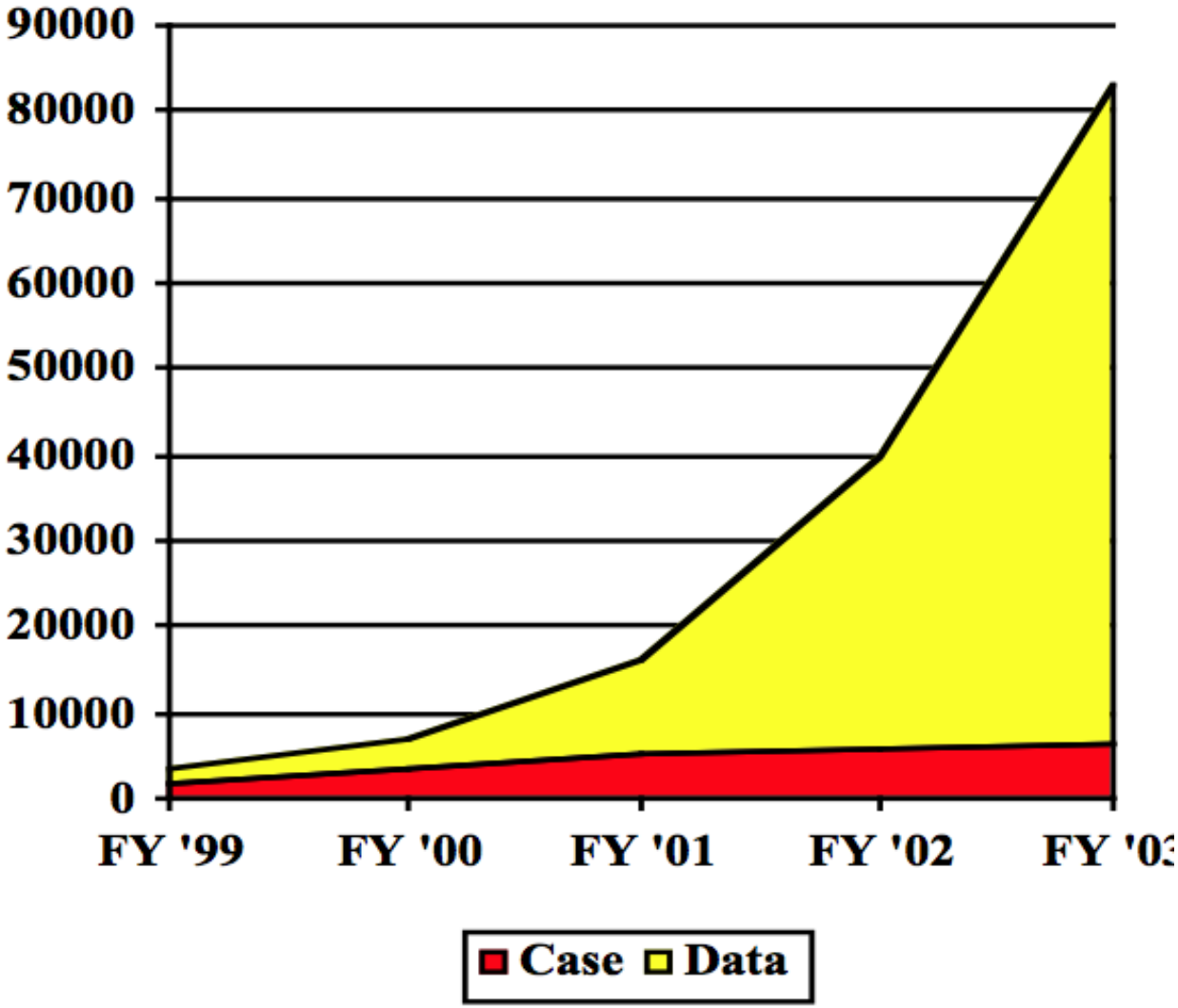
John Tebbutt

November 30, 2012

National Software Reference Library

Goal: to promote efficient and effective use of computer technology in the investigation of crimes involving computers.

NSRL collects software from various sources and incorporates file profiles computed from this software into a Reference Data Set (RDS) of information. The RDS can be used by law enforcement, government, and industry organizations to review files on a computer by matching file profiles in the RDS.



FBI's Cyber Caseload and Dataset Size Growth

Source: FBI CART, Oct 2003

History

NSRL evolved from FBI's Known File Filter

2001 : 233,281 hashes (400,000 files)
460 products
primarily Windows software

2012 : 28,530,178 hashes (91,961,336 files)
12,785 products
Various OS, languages
Contains malware

Not Merely A Library

The NSRL is conceptually four objects:

- **A physical collection of software**
- **A database of meta-information**
- **A subset of the database,**
  the Reference Data Set
- **A research environment**

All metadata is traceable back to original physical media

Library Contents

1,526 Manufacturers

most represented:
Adobe, Apple, Dell, HP, Intuit,
Microsoft, Oracle, Sun, Symantec

552 Operating Systems

most represented:
Windows (95,98,NT,2000,XP,Vista),
Linux, Mac OSX, Macintosh, Solaris, DOS

12,785 Products

most represented types:
Operating systems, games, office suite, database,
antivirus, financial, graphic/photo editor

Metadata

NSRL collects metadata that describes every file on all media in the physical collection.

Media tracking
Manufacturer information
Operating system requirements
Product description
File metadata includes:

Directory path, File name, Bytes, Digital signature (hash), etc.

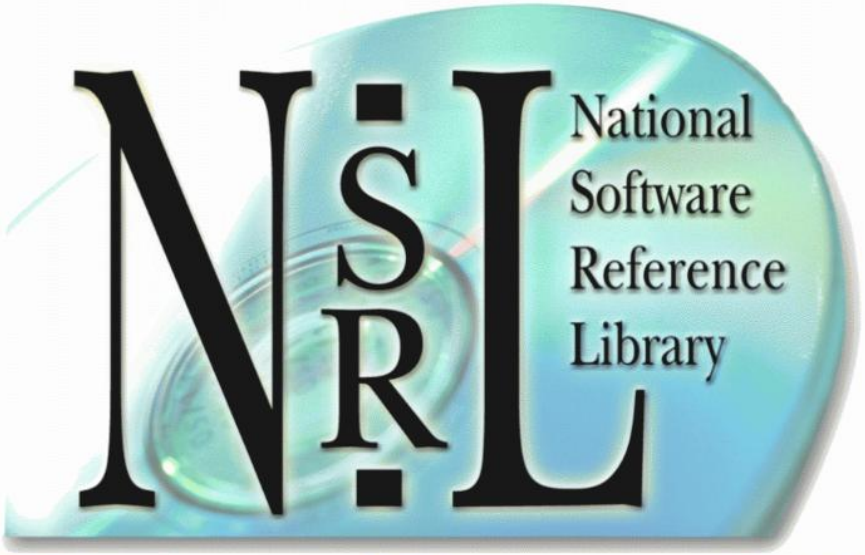
Hash Examples

```
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
```

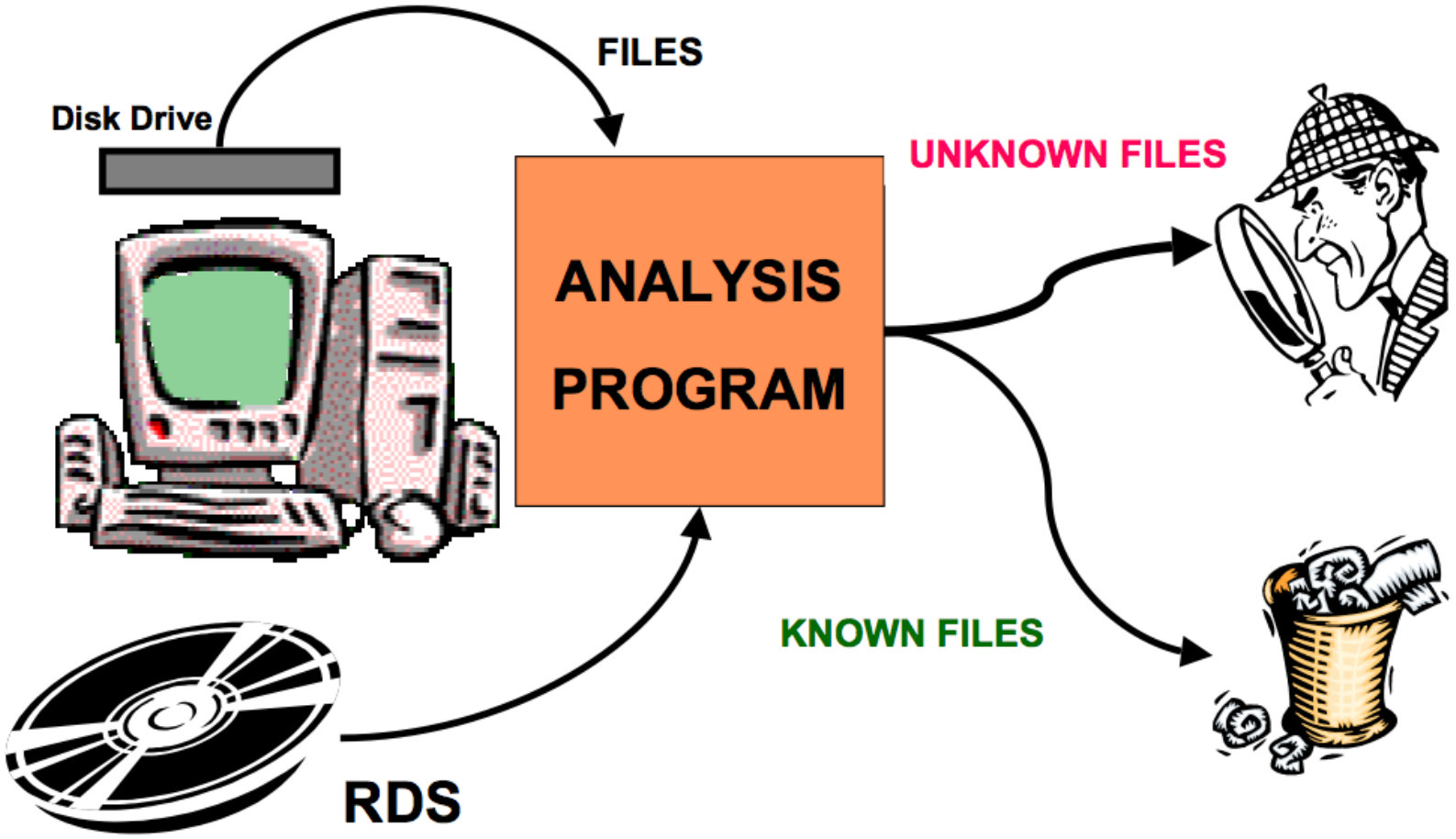
NIST Special Database #28



Reference Data Set
Version 2.38 10-01-2012

NIST

How the RDS is Used



Use in Commercial Products

Typically the RDS data is imported into one of many forensics and e-Discovery software tools.

NIST works with vendors to provide RDS import functionality.

NIST works with investigators to address evolving threats and trends.

Metadata Intent

The project sponsors were initially concerned with identification of known application files, to allow known files to be ignored, focusing investigation on user-generated data.

NIST does not assign “malicious” nor “notable” values to applications.

NSRL Impact

Essential to FBI CART, copied for every field office.
Imported into EnCase, FTK, Ilook, Hashkeeper, Maresware, etc.
Used by FDA in FL Botox case.
International use - UK NHTCU, EU JRC, etc.
Used by private organizations to eradicate P2P use.
Used by ISPs to track app sharing on servers.
Used by sysadmins to confirm valid OS file state.
Referenced by Simpson Garfinkel in 2002 efforts with reclaimed disks.
Referenced in 2001 seizure of bogus MS media in CA.

Media Image Collection

NSRL is using dcfldd to create a library of media images.

The media images can be processed by many algorithms automatically.

While the images are not available publicly, the metadata generated by algorithms is available, and researchers may access the images in our laboratory.

Foreign Languages

Metadata from English software assists foreign language investigations (and vice versa).

Applications available in various languages yield hash sets with a 50% identification rate.

Analyzing text and HTML tags can increase the identification rate to 99%.

Digital-only Software

Software is increasingly sold without physical media, via download mechanisms.

NSRL documents the acquisition of these products, and they enter the processing at the same point as the images of physical media.

Diskprints

Media images can be used to install any operating system in a virtual machine.

Any appropriate application can be installed on a VM OS.

NSRL is installing software and capturing the “footprints” of the installation files and system states, to augment the media file hashes.

Mobile Devices

The NSRL process can be applied to mobile devices, e.g. iPhones.

iOS apps can be processed by NSRL and can be identified on iOS devices and computers where sync occurs.

Technology Transfer

The code implementing the NSRL environment is freely available.

NSRL is occasionally asked to process data which cannot legally be accepted.

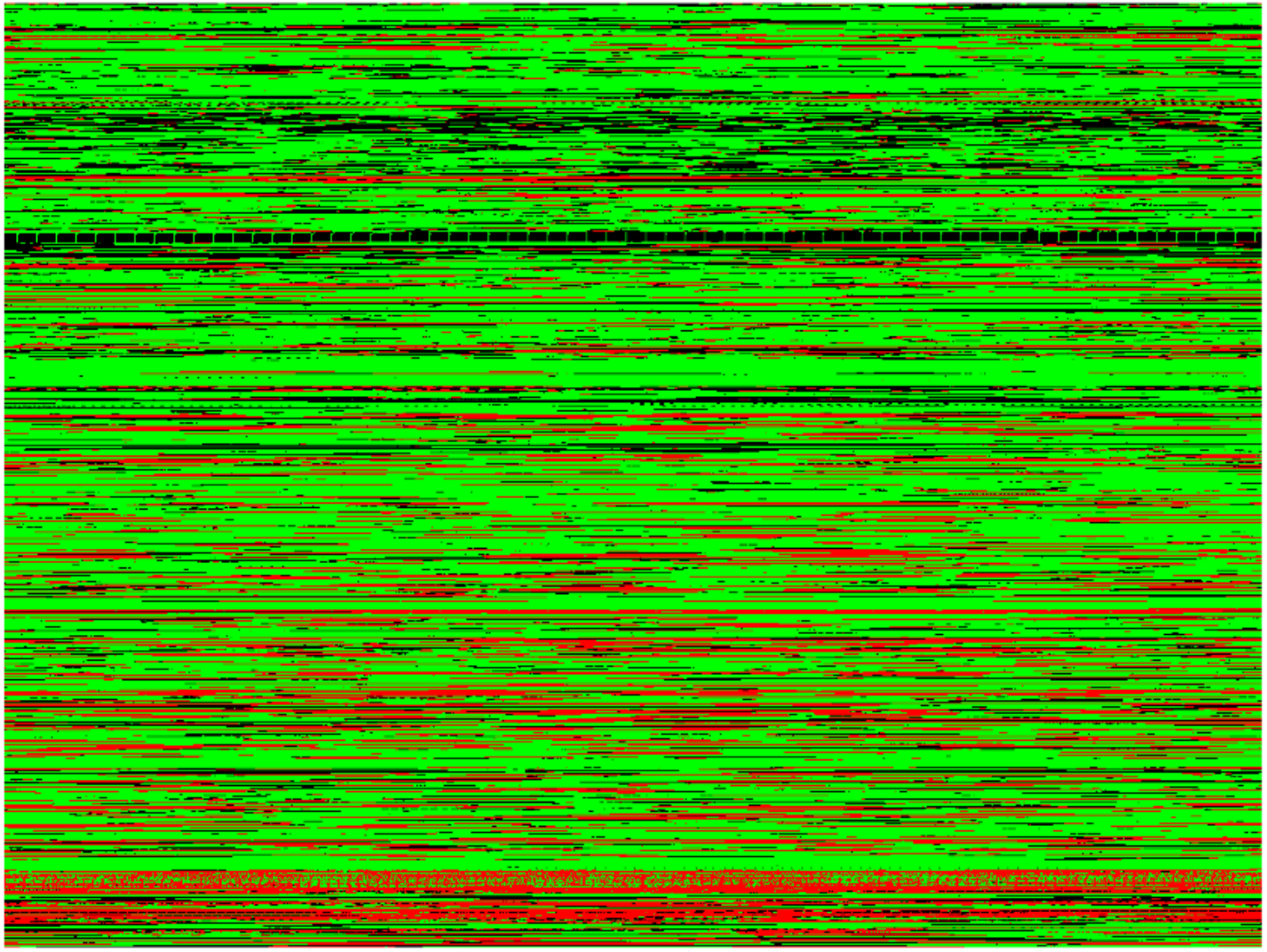
“NSRL in a Box” enables data holders to produce RDS-format hashsets.

External Research

- Block hashes
  - Smaller granularity, statistic interest
- Similarity digests
  - Capable of measurement
- Bloom filters
  - Distribution of large sets
- RAM resident identification
  - Use block hash and known static data
- Data relationships
  - Inter-system correlation of objects, metadata

Known - Unknown - Zero

2nd 512 MB in W2K NTFS VM



Similarity Metrics

fourscore.doc
fd375c1f4fe60fb4fbcef5b3f1bb035042e34fdd
0e0a6f71bc90534877f6018b50b94c2e97cab8f7
fivescore.doc

fourscore.doc
96:f+pIKe/OQxx1av5BVh:QSKcRuEuGtXo2s2Rhf6Pe2Qx+fV
96:BmpIKe/OQxx1av5BVK:vcRuEuOw5us2GNGrBSPe2Qx+fV
fivescore.doc

National Vulnerability Database & Common Platform Enumeration

The NVD is the U.S. government repository of standards based vulnerability management data. This data enables automation of vulnerability management, security measurement, and compliance. The NVD contains content (and pointers to tools) for performing configuration checking of systems implementing the FDCC.

http://nvd.nist.gov/

The official CPE Dictionary is hosted by NIST as part of the NVD. CPE is a structured naming scheme for information technology systems, platforms, and packages.

http://cpe.mitre.org/

Applicability Beyond Law Enforcement

Data preservation/curation

Electronic voting

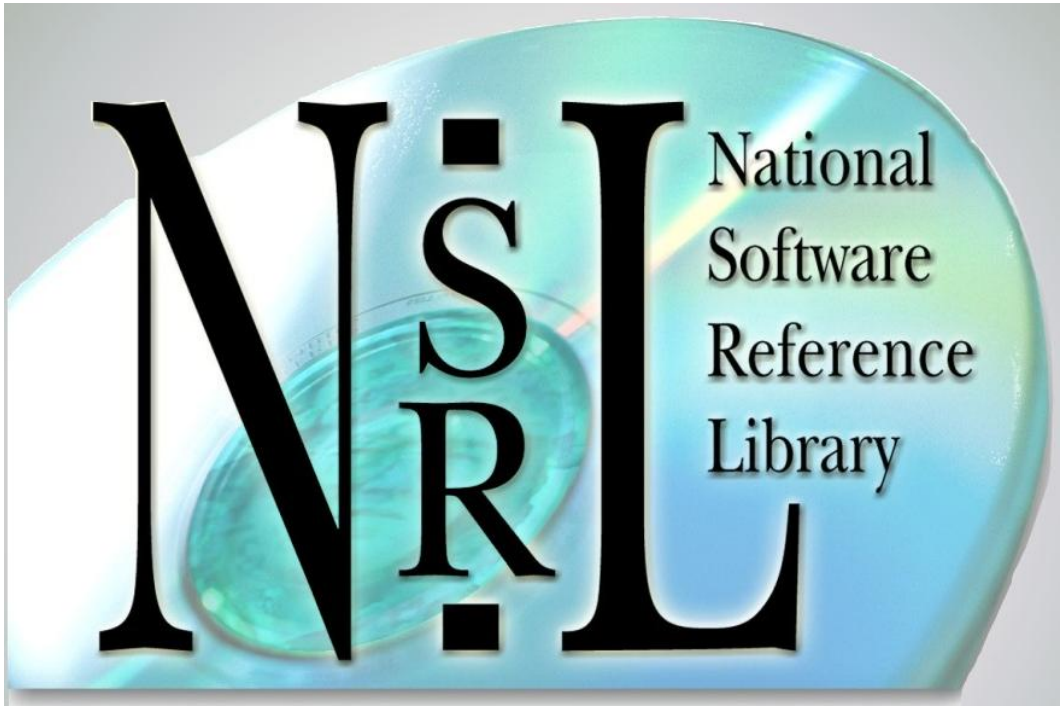
Critical infrastructure/SCADA

Contacts

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

Barbara Guttman
barbara.guttman@nist.gov

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



www.nsrl.nist.gov

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