National Software Reference Library

Douglas White

nsrl@nist.gov  www.nsrl.nist.gov
NSRL Core

Physical purchases → Physical software library → Database of file metadata → Reference data for investigative use

Data can be imported into many commercial digital forensics tools

File name, size, path, dates, SHA-1, MD5, etc. are recorded
All published data is traceable to original media. Collects metadata about files which can be used to uniquely identify files and their provenance. Metadata is used during investigations to automatically

- Eliminate known files
- Target files of interest

Supported by the U.S. Department of Homeland Security, federal, state, and local law enforcement.
NSRL Core + Storage

- Physical software library
- Database of file metadata
- Reference data for investigative use

Physical purchases → Virtual software library

- File name, size, path, dates, SHA-1, MD5, etc. are recorded
- Data can be imported into many commercial digital forensics tools
NSRL Core + Storage

All media are copied to network storage using forensic methods. Repeatable processes can be performed. Media degradation can be managed. Easy to incorporate new algorithms. Easy to extend metadata collection and measurement.
NSRL Expansion

Physical software library → Database of file metadata

Physical purchases

Download-only distributions

Virtual software library

Data can be imported into many commercial digital forensics tools

File name, size, path, dates, SHA-1, MD5, etc. are recorded

Reference data for investigative use

NIST Special Database #18

Reference Data Set
Version 3.29 06/1/2009

NSRL

FORENSICS @ NIST | December 3-4, 2014 • #NISTForensics
NSRL Expansion

Physical media have become less popular.
Downloaded software has a provenance.
Expansion into acquiring downloads enables
• Greater coverage of popular software
• Ability to add mobile apps
• Route for collaboration with other collections
NSRL Expansion + Corpus

Physical software library

Download-only distributions

Virtual software library

Millions of unique software application files

Database of file metadata

File name, size, path, dates, SHA-1, MD5, etc. are recorded

Data can be imported into many commercial digital forensics tools

Reference data for investigative use

File corpus is available to researchers

Non-file-based identifiers
Application of alternate algorithms
Common Platform Enumeration (CPE)
Software Identification tagging (SWID tag)
Security Content Automation Protocol (SCAP)

NSRL

Virtual Software Release Library
Reference Data Set
Version 3.28
06/1/2009

NSRL
National Institute of Standards and Technology
U.S. Department of Commerce

FORENSICS@NIST | December 3-4, 2014 • #NISTForensics
NSRL Expansion + Corpus

Research environment apart from core. Corpus built from all distinct files from media. Algorithms and processes can be applied to media images and to subsets of files.

File corpus is available to researchers
Non-file-based identifiers
Application of alternate algorithms
Common Platform Enumeration (CPE)
Software Identification tagging (SWID tag)
Security Content Automation Protocol (SCAP)
NSRL Expansion + Corpus

Alternate algorithm examples:

- SHA-256, SHA-512, SHA-3
- ssdeep, sdhash
- fiwalk, bulk_extractor
- Memory carving
- Manifest processing
- Block or sector processing
NSRL Next Generation

Physical software library → Database of file metadata

Download-only distributions → Virtual software library

Physical purchases → Virtual machine installation

Millions of unique software application files

Reference data for investigative use

Data can be imported into many commercial digital forensics tools

File corpus is available to researchers
Non-file-based identifiers
Application of alternate algorithms
Common Platform Enumeration (CPE)
Software Identification tagging (SWID tag)
Security Content Automation Protocol (SCAP)

Reference data from installation
Metadata collection from installation, execution, patches, etc.
Registry data from installation
Metadata collection of registry key and value status after install, execution and removal
NSRL Next Generation

NSRL uses virtual machine (VM) technology to investigate the forensics of the software life cycle.
NSRL Next Generation Diskprinting

Mary Laamanen
Alex Nelson
NSRL Expanded Core

Physical software library

Database of file metadata

File name, size, path, dates, SHA-1, MD5, etc. are recorded

Reference data for investigative use

Data can be imported into many commercial digital forensics tools

Physical purchases

Download-only distributions

Virtual software library
NSRL Diskprints
Motivation

Gather data on the specific effects of individual software packages on a system over the software's lifetime.

Provide digital forensic investigators with new reference data.

Extend the NSRL research environment for use by forensic researchers to develop new tools and techniques.
Systems and Software

All software is part of the NSRL library
- Provides Traceability

Operating System
- Focus on versions of Microsoft Windows

Software applications
- Chosen based on recommendations
Virtual Machine Advantages

VM state can be captured at any time
- VM may be paused / suspended

VM is preserved as a set of files
- Hard drive, RAM contents, etc

Can be copied off for external processing

Saved for future reference
Artifacts

Baseline

Installation

Running

Uninstallation

Rebooting
Captured Data

Filesystem (file hashes, MAC times, etc)
- Executables
- Libraries
- etc.

Configuration information
- Windows Registry

Memory mapping information
- System RAM

Network communication
- pcap files
Snapshot Metadata

Snapshot Id – Unique Id

Application Lifecycle State – Record the application lifecycle stage

Snapshot Notes – Record all user actions taken when generating the snapshot including unexpected behavior
Processing Workflow

What do we do with all this data?
NSRL Next Generation – Diskprinting

Mary T. Laamanen\textsuperscript{1}, Alex J. Nelson\textsuperscript{2,3}

\textsuperscript{1} NIST
\textsuperscript{2} Prometheus Computing
\textsuperscript{3} University of California, Santa Cruz