The National Software Reference Library

By

Douglas White

Presented At

The Digital Forensic Research Conference

DFRWS 2014 USA Denver, CO (Aug 3\textsuperscript{rd} - 6\textsuperscript{th})

DFRWS is dedicated to the sharing of knowledge and ideas about digital forensics research. Ever since it organized the first open workshop devoted to digital forensics in 2001, DFRWS continues to bring academics and practitioners together in an informal environment. As a non-profit, volunteer organization, DFRWS sponsors technical working groups, annual conferences and challenges to help drive the direction of research and development.

http://dfrws.org
National Software Reference Library

Douglas White
Disclaimer / Disclosure

Trade names and company products are mentioned in the text or identified. In no case does such identification imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the products are necessarily the best available for the purpose.

This research was funded by the National Institute of Standards and Technology Office of Law Enforcement Standards, the Department of Justice National Institute of Justice, the Federal Bureau of Investigation and the National Archives and Records Administration.
Physical software library

Database of file metadata

Physical purchases

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
National Software Reference Library

Physical software library

Database of file metadata

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

Reference data for investigative use

NSRL Reference Data Set
Version 2.25 06/1/2009

Data can be imported into many commercial digital forensics tools

Download-only distributions

Virtual software library

Physical purchases
National Software Reference Library

Physical software library

Virtual software library

Database of file metadata

Virtual machine installation

30,000,000+ unique software application files

Reference data for investigative use

Data can be imported into many commercial digital forensics tools

Reference data from installation

Registry data from installation

Metadata collection from installation, execution, patches, etc.

Metadata collection of registry key and value status after install, execution and removal

File corpus is available to researchers

RAM-based identifiers
Alternate cryptographic hashes
Common Platform Enumeration (CPE)
Security Content Automation Protocol (SCAP)
National Software Reference Library

14,320 Applications
2,281 Manufacturers
696 Operating Systems
30,193,306 Unique Files (corpus)
37,827,886 Unique Files (DB)
192,090,545 File Records
Many Registry entries

Shrinkwrapped & “Clickwrapped” products

Mobile device apps, iOS & Android
Scope: free, popular apps
$x = \text{int}(\log(\text{bytes})/\log(2))$, $y = \text{count of files}$

Majority are between 1024 and 8192 bytes
<table>
<thead>
<tr>
<th>File Extension</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>.GIF</td>
<td>4,653,105</td>
</tr>
<tr>
<td>.HTML</td>
<td>2,852,898</td>
</tr>
<tr>
<td>.CLASS</td>
<td>2,334,990</td>
</tr>
<tr>
<td>.HTM</td>
<td>1,768,049</td>
</tr>
<tr>
<td>.JPG</td>
<td>1,492,402</td>
</tr>
<tr>
<td>.PNG</td>
<td>1,070,950</td>
</tr>
<tr>
<td>.WMF</td>
<td>1,050,296</td>
</tr>
<tr>
<td>.C</td>
<td>1,026,247</td>
</tr>
<tr>
<td>.H</td>
<td>962,059</td>
</tr>
<tr>
<td>.DLL</td>
<td>907,594</td>
</tr>
<tr>
<td>.EPS</td>
<td>537,649</td>
</tr>
<tr>
<td>.XML</td>
<td>524,945</td>
</tr>
<tr>
<td>.WAV</td>
<td>402,867</td>
</tr>
<tr>
<td>.TXT</td>
<td>341,116</td>
</tr>
<tr>
<td>.EXE</td>
<td>331,244</td>
</tr>
</tbody>
</table>
National Software Reference Library

RDS – SHA1, MD5, CRC32
Additionally: SHA256, SHA512, SHA3 (upon FIPS)
Block hashes
Bulk_extractor
Approximate matching : SP 800-168
   http://csrc.nist.gov/publications/PubsSPs.html

Corpus of unique files
Collection of media (disk) images
Photographs of materials

Installation in virtual machines
Memory
Registry
Network
National Software Reference Library

Taxonomy of applications, OSes

Automated identification: SWIDTags (ISO 19770)

Interoperability with National Vulnerability Database

CYBOX, DFXML
Contacts

Barbara Guttman, Software and Systems Division
guttman@nist.gov

Douglas White, NSRL
dwhite@nist.gov -or- nsrl@nist.gov

www.nsrl.nist.gov

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