B15 The National Software Reference Library: Applications in Digital Forensics

Douglas R. White, MS, 4225 Angell Road, Taneytown, MD 21787-2601*

After attending this presentation, attendees will be familiar with the data generated by the National Software Reference Library (NSRL), which is freely available to digital forensics investigators and researchers.

This presentation will impact the forensic science community by calling attention to the various data sets produced by the National Institute of Standards and Technology (NIST) NSRL.

The NIST National Software Reference Library has collected computer software since 2000. Acquisitions have spanned Microsoft®, Apple®, Linux®, and other operating systems. Applications range from common business software suites to foreign, malicious, specialized executables. The NSRL retains copies of all applications in the collection.

Each application is described with metadata, detailing the manufacturer, publisher, system prequisites, etc. Every file in each application is described with metadata detailing the file name, path, dates, byte signatures, cryptographic hash (file fingerprint), etc. All of this metadata is available to the public. A subset of the metadata that is targeted for investigators is published quarterly and made available to the public.

All media in the collection are copied to network-based storage. All distinct files are stored in a corpus on network-based storage. “Archive” type files (Cabinet, Zip, tar files, etc.) are recursively extracted and content files are added to the corpus. As needs arise for heretofore uncollected metadata, NSRL can update and repeat processes to harvest additional information from the media and files. Access to media images and the file corpus may be obtained.

Software applications are installed in virtual machines to collect metadata on files, memory, and operating system structures (e.g., Windows® registry). Several points in the installation are chosen while the system is in a stable state, to enable identification of actions taken during the lifecycle of the applications.

Software acquisition includes mobile device applications, “clickwrapped” download-only applications, network-based multiplayer applications, and game console applications. The NSRL collaborates with other libraries and collections. Metadata taxonomies and standard software identifiers are used to enable information sharing between diverse data sets. A “Digital Forensics XML” (DFXML) schema has been drafted to facilitate discussion on interoperable data.

NSRL investigates methods of identification other than the use of cryptographic hashes. Block-level hashing introduces statistical probabilities into the process, and approximate matching involves algorithms that allow measurement of match characteristics.

File Identification, Digital Metadata, Registry

 ** Presenting Author*