F48 "De-NIST-ing": The Evidence and Science Behind the Term

Douglas R. White, MS, 100 Bureau Drive, MS 8970, Gaithersburg, Maryland 208998970; and Mary T. Laamanen, MS, NIST, 100 Bureau Drive, Gaithersburg, MD 20899*

After attending this presentation, attendees will understand the scientific process and handling of evidence which culminates in the generation of the National Institute of Standards and Technology (NIST) National Software Reference Library (NSRL) reference data set. This data set is commonly known via the term "de-NIST-ing" as applied to deduplication of case materials.

This presentation will impact the forensic science community by showing how the deduplication of case materials is a universal step in processing. A practitioner should have working knowledge of the scientific and evidentiary foundation upon which this step rests.

The NSRL collects software from various sources and incorporates file profiles computed from this software into a Reference Data Set (RDS) of information. The RDS is used to review files on a computer by matching file profiles in the RDS. This data set assists in automated deduplication involved in determining which files are important as evidence on computers or file systems that have been seized as part of an investigation. The data set can just as easily be used to target files of interest. Such uses include detection of unauthorized software installations (e.g., in corporate or web hosting environments or in intellectual property disputes) and discovery of exculpatory evidence by criminal defense teams.

The NSRL is comprised of: (1) a collection of original software, stored for evidentiary foundation; (2) a virtual collection of software, created via digital forensics tools; (3) a freely available metadata RDS for investigative use; and, (4) a standalone research environment enabling access to all NSRL data.

A rigorous, open, scientific process is followed to preserve original data and provide metadata describing various objects and states encountered on computer systems and storage. "De-NIST-ing" is a final step in that process, performed by practitioners. The "de-NIST-ing" term obscures the underlying process, and the multiple applications in which the RDS metadata may be used: deduplication, benign identification, malicious identification, etc.

This project is supported by the United States Department of Homeland Security, federal, state, and local law enforcement, and the NIST to promote efficient and effective use of computer technology in the investigation of crimes involving computers.

DeNISTing, Deduplication, Digital

*Presenting Author