NATIONAL LAW ENFORCEMENT AND CORRECTIONS TECHNOLOGY CENTER

A program of the National Institute of Justice

From Fall 2003 *TechBeat*

TECH b•e•a•t

Dedicated to Reporting Developments in Technology for Law Enforcement, Corrections, and Forensic Sciences

Electronic Files: Criminal to Historical

Uncovering evidence of computer crime. Managing and preserving electronic records from the National Archives. Seemingly unrelated? Not really.

When a computer is seized from a crime scene, chances are it contains valuable evidence. Until recently, searching for such evidence by reviewing and analyzing the thousands of electronic files on a computer was tedious and time consuming. Now, investigators can employ the National Software Reference Library (NSRL) to automate their investigations.

According to Doug White, a computer scientist at the National Institute of Standards and Technology (NIST) and the lead scientist on the NSRL project, NSRL collects software from various sources and incorporates file profiles—including file name, byte size, and location—from the software into a reference data set (RDS). Law enforcement agencies, the Federal Government, and industry organizations can match seized files with profiles in the RDS. This allows them to determine more quickly and easily which files are evidence and which can be disregarded.

The NSRL project is supported by the National Institute of Justice, NIST's Information Technology Laboratory, the U.S. Department of Defense, the U.S. Department of the Treasury, and State and local law enforcement agencies. Its goal is to "promote the efficient and effective use of computer technology in the investigation of crimes involving computers," White says. The technology behind the NSRL involves digital signatures, or "hash sets." Says White, "The concept of having a unique string that can be identified with a particular data file is similar to unique fingerprints identifying a person. The contents of every file can be manipulated mathematically to give you a unique value or number." The value or number, he says, can uniquely identify the file.

When a computer is seized, investigators use computer forensic tools to create hash values of the files on the computer and compare those values with the reference hash set. White says this comparison allows automatic elimination of files that investigators do not need to investigate further and thereby saves a significant amount of time.

Using such hash sets can often eliminate 75 percent of the files on a computer. "An example we like to give is, if an investigator is looking for a bomb schematic or a facility map on a computer that is running Windows® 2000, Windows 2000 software has nearly 6,000 images as part of its operating system," White says. "By applying our hash set, the investigator won't even have to look at any of those files right off the bat."

White estimates that, since the project began 2 years ago, NIST has hashed just more than 4,000 software applications. Once the software is collected, it is shelved in a locked room in case the project team members need to recalculate hashes. "This actually results in court-admissible data, and that is the primary focus of the hash set, to be court admissible," he says.

Although other hash sets are available to law enforcement agencies, NIST was chosen for this project primarily because "we could keep the information traceable and repeatable," White says. "We collect enough information about every file in every piece of software on every disk or CD to uniquely identify it on our shelves." The information is stored in a massive database. The RDS is extracted from the database and put on CD-ROMs, which are published quarterly for a yearly subscription fee of $90. White says the RDS has a free redistribution policy, meaning that a subscriber agency can copy its CD and give it to other agencies for free. The NSRL website (www.nsrl.nist.gov) has sample data that agencies can download to evaluate and decide if they want to subscribe.

For agencies that for security or other reasons cannot send files to NIST to hash, White says that the goal is "to provide them with the cookbook to build a duplicate environment so that they could hash these files on their own." He says the project hopes to have an open-source version of the code used to generate the hashes available by late summer 2003.

The National Archives and NSRL

NSRL's applications extend beyond investigating computer crimes. According to White, the National Archives and Records Administration (NARA) is collaborating with NIST to research the use of NSRL to manage and preserve Federal electronic records collections, including those of former President George Bush. White is looking at using NSRL to identify duplicate and application files in the backlog of presidential library material and separate those from the presidential records.

According to Robert Chadduck, Research Director for NARA's Electronic Records Archive (ERA) Program, the project "builds on the great ideas and technology previously developed among NIST and NIJ but extends them to the new problem of managing and processing electronic records collections." In addition to the NSRL project, NARA is planning to build a state-of-the-art ERA to preserve the electronic records of the Federal Government for future generations. ERA is envisioned to be a comprehensive, systematic, and dynamic means of preserving electronic records, free from dependence on any specific hardware or software.

The NSRL project began in late summer 2002 and is ongoing. So far, NARA has provided NIST with a small set of electronic records from the computer systems of former President Bush for processing. Says White, "This is at a very early stage." Once NIST has finished processing the files, NARA will release the results.

For more information on the National Software Reference Library, visit the NSRL website, www.nsrl.nist.gov; call Doug White, 301–975–4761; or e-mail douglas.white@nist.gov. For more information on the National Archives' NSRL/ERA project, call ERA, 301–837–0740, or e-mail ERA.Program@ nara.gov.

This article was reprinted from the Fall 2003 edition of *TechBeat,* the award-winning quarterly newsmagazine of The National Law Enforcement and Corrections Technology Center system, a program of the National Institute of Justice under Cooperative Agreement #96–MU–MU–K011, awarded by the U.S. Department of Justice.

Analyses of test results do not represent product approval or endorsement by the National Institute of Justice, U.S. Department of Justice; the National Institute of Standards and Technology, U.S. Department of Commerce; or Aspen Systems Corporation. Points of view or opinions contained within this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice.

The National Institute of Justice is a component of the Office of Justice Programs, which also includes the Bureau of Justice Assistance, Bureau of Justice Statistics, Office of Juvenile Justice and Delinquency Prevention, and Office for Victims of Crime.