Report to U.S. Election Assistance Commission

July 13, 2004

The National Institute of Standards and Technology (NIST) has run the National Software Reference Library (NSRL) project since 2000. The NSRL provides the means to match files that are found on a computer with known files from software applications, thus making it possible to verify that two files are either the same or different.

The NSRL has a collection of more than 5,000 software products in a secured room. The information to uniquely identify each file from those products is stored in a database. A portion of the database information is published quarterly, and is available for free.

Law enforcement agencies and researchers use the published database information to automatically eliminate known files from an examination. This process identifies unknown or suspect files that are not expected on the computer. The concept can be turned around and used to verify that specific files do exist on the computer.

NIST was asked by the National Institute of Justice (NIJ) to host the NSRL because NIST is unbiased, not related to a law enforcement agency nor a forensic software vendor; NIST has a history of producing high quality data; the process NIST uses to produce the data is open and transparent. NIST can provide traceable, repeatable data sets.

The NSRL collection is a combination of products that have been purchased, and products that have been donated. All of the software in the NSRL is in the form that the researchers would find on a computer. The NSRL software is not available for use, it remains in escrow and is handled as potential evidence for admission to court.

The critical technology in the NSRL is a “hash” or “fingerprint” of a file. A hash uniquely identifies a file, and is similar to a human fingerprint, in that a person cannot be recreated from a fingerprint – likewise, a file cannot be recreated from a hash. The hashing algorithm primarily used in the NSRL is the Secure Hash Algorithm (SHA-1) specified in Federal Information Processing Standard (FIPS) 180-1.

In addition to use in the law enforcement community, the NSRL process has been applied to National Archives and Records Administration (NARA) data. Using the NSRL, we have separated the application files from a collection of non-classified Presidential materials on 94 computer systems, leaving Presidential data content. The process also identifies duplicate files.

The NSRL concept can potentially assist in addressing voting systems needs in several areas. Officials could determine that the software used during elections is the expected software. The tested, certified version is definitively identifiable. Verification that the software remains the same during distribution, installation, setup, or use is possible, supporting a “chain of custody.”

The technology used in the NSRL can stand up to the call for transparency in the voting process.

The NSRL methodology is in the public domain, available for inspection, as are the underlying algorithms. Jurisdictions can share knowledge with each other, as the hash values do not allow reproduction of proprietary software.

How can this technology specifically strengthen the voting process? An official can verify that operating system file contents have not been modified on a computer, that the device is configured as intended. An official can verify that application file contents have not been modified, that the voting software is the expected, certified, undisturbed version. An official or automated monitor can verify that known static sections of files have not been modified, providing assurance during the time of use.

Following Chairman Soaries' remarks about Electronic Voting Security Strategy, in particular to "invite every voting software vendor to submit their certified software to the NSRL", NIST has been contacted by two vendors, and are getting the legal issues for access to the software ironed out. At the Technical Guidelines Development Committee (TGDC) organizational meeting on July 9, 2004, a resolution was passed: "RESOLVED: that the TGDC recommends to the Election Assistance Commission the expediting of registration of currently certified voting system software into the National Software Reference Library at NIST."

As software is provided to NIST and the NSRL, it will be handled as potential evidence in a court case. The information that the NSRL gathers for file identification will be made available to the jurisdictions, testing agencies, vendors, and to law enforcement. This information will support the actions mentioned previously, as well as provide a security check outside the scope of the EAC and HAVA – if an illicit copy of voting software is identified during an unrelated investigation, that leak can be reported to appropriate authorities.

The NSRL staff has begun to identify research issues in this field. Issues include:

1. There can be differences in the hashes made from software distribution media vs. hashes made from the certified installation.
2. If there is any setup after the hashes are made, how are valid changes tracked and managed?
3. Can it be possible and/or practical to have on-location, time-of-certification hashing by a trusted agent?
4. Given system specifications, can the hashing technology provide the requested verification within time, space and security constraints?

Thank you for the opportunity to testify. I look forward to assisting the Commission in the future. I would be happy to answer any questions the Commission may have.

Respectfully,

Douglas White
NIST/ITL