B14 Tracking the Effects of Software on Systems: A Forensic Metadata Collection

John Tebbutt, 100 Bureau Drive, STOP 8970, Gaithersburg, MD 20899-8970; Mary T. Laamanen, MS, NIST, 100 Bureau Drive, Gaithersburg, MD 20899; and Alex J. Nelson, MS, 100 Bureau Drive, MS 8970, Gaithersburg, MD 20899-8970*

The goal of this presentation is to familiarize attendees with a data set available from the National Institute of Standards and Technology (NIST) and aimed at digital forensics investigators which tracks the lifecycle of targeted software and provides data on the files, Windows® Registry components, memory states, and network communications associated with the targeted software at specific points in its lifecycle, including after uninstallation.

NIST's National Software Reference Library (NSRL) is building a forensic data set which tracks the various effects of software on computer systems under controlled conditions, providing forensics investigators with a baseline record of artifacts produced at specific points in the software's deployment.[1]

Virtual machine (VM) environments built on common Windows® operating systems are used as installation environments for targeted software. At predetermined break points during the life cycle of the software, the VMs are paused and saved. Simultaneously, any network traffic originating from the VM is captured and is also saved at each break point in the life cycle. The software life cycle is defined as its installation, registration, online update, use and uninstallation. The sum of all data collected over the life cycle of the software is referred to as the software's "diskprint," while the data collected at each stage is referred to as a "slice." Once the software diskprint has been saved and recorded, the slices are processed and the associated files, Registry entries, RAM contents and network traffic are extracted for publication. Note that the baseline operating system installations were also processed in this manner: the data on clean operating system installations is also available.

The results are published in two formats: File metadata are folded wholesale into the NSRL RDS to supplement the file metadata garnered in the usual manner.[2] The file metadata, Registry information, RAM cells and network traffic are also published in an XML format that describes the diskprint in terms of its constituent slices.

NIST is working closely with the DFXML and the Cybox communities on the definition of a format for the publication of diskprints for easy ingestion into forensic tools.[3,4]

The intention behind this data set is to provide investigators with definitive evidence of the artifacts associated with software packages, thereby reducing the community's reliance on untested knowledge, inference and hearsay and introducing a new level of rigor into the discipline.

References:

1. The National Software Reference Library is online at http://www.nsrl.nist.gov
2. See, for example, "NIST's National Software Reference Library": http://www.nsrl.nist.gov/Documents/NSRL-CFS-April-2009.pdf
3. Digital Forensics XML in the Forensics wiki: http://www.forensicswiki.org/wiki/Category:Digital_Forensics_XML
4. Cyber Observable eXpression: A Structured Language for Cyber Observables: http://cybox.mitre.org

Diskprint, File System, Windows® Registry

** Presenting Author*