D14 Identification of Known Files on Computer Systems

Douglas White; and Michael Ogata, National Institute of Standards and Technology, 100 Bureau Drive STOP 8970, Gaithersburg, MD 20899-8970*

After attending this presentation, attendees will learn about identification of known computer files and be able to implement automated processes to eliminate such files in their computer forensic practice.

This presentation will impact the forensic community and/or humanity by introducing one method of reducing the data in digital forensics cases.

The amount of data involved in digital forensics investigation can be greatly reduced by automated means by eliminating known files from computer systems. The method used to obtain the data reduction is based upon the National Institute of Standards and Technology (NIST) National Software Reference Library (NSRL) data set. The NSRL data set can be applied to several different operating systems and can be used with several off-the-shelf commercial software tools. In laboratory tests, data reduction up to 95% has been obtained, while in the field, rates of up to 80% have been obtained.

The National Institute of Standards and Technology (NIST) hosts a project that promotes efficient and effective use of computer technology in the investigation of crimes involving computers. The National Software Reference Library (NSRL) is designed to collect software from various sources and incorporate file profiles computed from this software into a Reference Data Set (RDS) of information. The RDS is a collection of digital signatures of known, traceable software applications.

Numerous organizations including law enforcement, government, and industry use the NSRL data set to reduce the amount of data involved in digital forensics cases. The NIST data is collected with the requirement of court admissibility. While a courtroom may not be the destination of the investigation, the possibility is not excluded due to this data set.

The RDS is a free resource. Instructions for obtaining the RDS will be given. Technical descriptions of the contents of the RDS will be briefly discussed. Methods to use the RDS to identify file “pedigrees” and application relationships will be shown.

Several commercial computer forensics software tools exist that leverage the information from the NSRL. Tips on use of these tools will be provided.

A collection of laboratory measurements of the application of the NSRL data set to known reference computer systems will be presented, to give the theoretical upper bound of data reduction capabilities. This will be followed by a presentation of similar real-world systems processed with the same methodology to show more realistic response.

Computer, File, Identification

 ** Presenting Author*