B19 Forensic Data Extraction in Computer Child Pornography Investigations

Sigurd Murphy, BA, Defense Computer Forensic Laboratory, Linthicum, MD; Donald P. Flynn, JD, Defense Cyber Crime Center, DC3, Linthicum, MD; Thomas J. McAleer, Defense Computer Forensic Laboratory, Linthicum, MD; Andrew Medico, BS, Defense Cyber Crime Institute, Linthicum, MD; Daniel Raygoza, Defense Computer Forensic Laboratory, Linthicum, MD ; and Michael J. Salyards, PhD*, U.S. Army Crminal Investigations Laboratory, 4930 North 31st Street, Forest Park, GA 30297*

After attending this presentation, attendees will understand the unique technical, legal, and investigative challenges in child pornography investigations, learn about the key forensic steps in conducting an automated examination, and understand the importance of a user-friendly display.

This presentation will impact the forensic science community by explaining how forensic data extraction is a powerful tool that could revolutionize the way digital evidence is examined primarily in child pornography investigations and possibly in other types of offenses.

This presentation will describe a new and powerful model for examining digital evidence in child pornography examinations. Child pornography introduces three challenges. First, anecdotal reports from local, state, and federal laboratories show that the sheer numbers of cases and volume of media associated with child pornography overwhelms most digital evidence laboratories. Second, computer forensic examiners are often asked to make determinations outside of their expertise like medical determinations (especially about the age of children in digital images) and legal determinations about whether the content of a picture of document contains pornographic material. Finally, examiners are burdened with lab requests that ask for "all information about all files." These types of examinations can take 3-9 months to complete, and often contain detail that is neither understood nor used by customers in the investigative and legal communities.

Forensic Data Extraction (FDE) was designed to meet these challenges and consists of four key elements. First, a robust extraction tool uses commercially available forensic software tool to search cases for images, videos, Peer-to-Peer file sharing logs, email messages, internet chat logs, and web browser history. The tool searches both allocated files and unallocated space, and automatically looks inside compressed files. Operating system and application files are eliminated using the NIST NSRL. This step is performed in a highly efficient machine-driven manner that was designed to be run in a massively parallel operation.

Second, all of the recovered files are stored in a database with their associated metadata (original path, size, last modified time, etc). MD5 hashes are computed for each recovered file so that files can be matched against lists of known child pornography files. Images and videos not found in the known files list are ranked by an algorithm that judges human (flesh) content. In addition, thumbnails of videos are generated after skipping past opening title/credit screens so the investigator can easily see a preview of the content of the video.

Third, a highly robust and user-friendly front end allows for easy viewing, sorting, and classification of the files by the investigative, medical, and legal communities. Known child pornography files and files that are known to be transferred via peer-to-peer (P2P) software are grouped together and highlighted. The remaining images are sorted by human content rating so that the investigator can view more likely files first. This front is delivered inside of a mini virtual machine to facilitate support for the largest possible set of customer machine configurations.

Finally, a follow-up mechanism allows the customer to quickly and securely request that a follow-on examination be conducted in a manner that they control. This technique marries files selected by customers with the back-end database to allow for timely follow-up examinations on files of interest.

This model results in dramatic increases in productivity and timeliness while simultaneously allowing the customer to maintain an increased amount of control over their casework. Unexpected benefits include increased guilty pleas, identification of additional victims and acceptance of the customer front end by the judicial community. Details will be presented about how FDE works, statistics showing the effects on productivity and some noteworthy case examples.

Digital Evidence, Computer Forensics, Child Pornography

 ** Presenting Author*