D35 Hashing of File Blocks: When Exact Matches Are Not Useful

Douglas White, MS, National Institute of Standards and Technology, 100 Bureau Drive Stop 8970, Gaithersburg, MD 20899-8970*

After attending this presentation, attendees will understand some principles of eliminating benign information from investigations of computer systems, based on cryptographic hashes of files and partial files.

This presentation will impact the forensic community by introducing the rigor of cryptographic digital file identification at a granular level which supports statistical identification of objects.

Use of cryptographic hashes or “digital fingerprints” to automatically identify files is absolute when applied to a file as a whole; the file is unambiguously categorized. When dealing with morphing digital objects, such sorting leaves many files to be dealt with by manual review.

Block hashing is a method of applying the cryptographic algorithms to smaller-then-filesize portions of the suspect data. In this case study, the portions align with the blocks of the computer’s hard disk. The aggregation of the unambiguous block hash values allow statistical probabilities of identification of suspect files, taking the dynamic nature of digital objects into consideration. This is parallel to the use of latent fingerprints from a few of a suspect’s fingers rather than a complete tenprint set for identification.

Examples of practical applications of this technique, along with preliminary error rates will be presented.

Automated Investigation, File Hash, File Fingerprint

* *Presenting Author*