By THE COMMON DIGITAL EVIDENCE STORAGE FORMAT WORKING GROUP

Standardizing DIGITAL EVIDENCE STORAGE

Investigators have an increasing need to share digital evidence between different organizations and analysis tools. But today's investigators are hindered by a variety of independently developed and incompatible formats used to store digital evidence. Problems arise when dealing with different disk image formats, and the difficulties are exacerbated when dealing with diverse kinds of evidence, such as nework logs and the contents of mobile devices. Without standards that are both open and technically sound, the risk is that evidence may be lost, cases may be compromised, and innocent people may be improperly convicted—or guilty parties let free.

Forensic copies of storage media provide an illustrative example of weak standardization. The current de facto standard for storing information copied from a disk drive or memory stick under investigation is the so-called "raw" format: a sector-by-sector copy of the data on the device into a file. However, the raw format does not store metadata that can be vital to an investigation, such as the drive's serial number, the date and place that the drive was imaged, and a digital signature or cryptographic checksum to verify the data's integrity. Nor is the raw format error tolerant—if a portion of the evidence file becomes corrupt, we cannot isolate the damage and still use the intact remainder. The raw format cannot even distinguish between sectors that are blank and those that are inaccessible because of hardware error.

From a practical viewpoint, the biggest problem with raw files is their size. Raw files are not compressed. A raw file from a 200GB hard drive, for example, requires 200GB to store, even if the drive only had 100MB of actual files.

Proprietary formats that address some of

THE LACK OF A GENERALLY ACCEPTED FORMAT FOR STORING ALL FORMS OF DIGITAL EVIDENCE IS HAMPERING THE DEVELOPMENT OF DIGITAL FORENSICS AS A SCIENTIFIC DISCIPLINE, AND MAY RESULT IN COMPROMISED OR LOST EVIDENCE, AND SIGNIFICANT JUDICIAL CONSEQUENCES.

these issues are inherently limited by the desire of each vendor to create a format that is distinct from others. This has created a number of problems. Converting between proprietary formats may result in incorrect data, missing metadata, and lost time. A proprietary format can also create difficulties for individuals who do not have the access or ability to use software that reads such files. Some courts may not accept evidence stored in proprietary formats that are trade secrets and subject to change because they hinder validation and access to the evidence using other tools.

Even open file formats that are well documented can be data prisons if the format lacks sufficient expressiveness for the information it needs to embody, or if the standard is so complicated it cannot be implemented correctly.

Digital forensics practitioners are now working to define a standard format for storing and transmitting digital evidence and its associated metadata so that it can be processed efficiently by multiple tools and parties, and can ensure evidence integrity and effective case management.

One desirable feature of a common format is an audit trail that documents the chain of custody of the digital evidence. This requires that every action performed relating to an acquisition or alternation of digital evidence be recorded. But this goal is a long way from current practice. Today's best practice for assuring the integrity of digital evidence is for the examiner to keep a paper notebook in which he or she writes the MD5 hash of the acquired disk image. Therefore, improved methods are needed to support authentication and non-repudiation of digital evidence many years in the future, even when the person who collected the data is not available. Another desirable feature of a common storage format is the flexibility to accommodate many forms of digital evidence, including network traffic, memory dumps, and logical files that have been acquired as evidence.

A standard digital evidence storage format will be analogous to the evidence bags and tags used at physical crime scenes, where the evidence is placed in a sealed bag and related information is written outside the bag on a tag in a standard language and format, such as the acquisition location and time. The current state of digital evidence storage formats is similar to having no bag or bags with information written using private notations that are not widely understood.

A set of well-defined properties would enable persons without technical training to evaluate the reliability of the evidence. Decision makers in an organization or the courts would be able to define a minimum set of requirements for their context. A standard format would encourage the development and commercialization of better evidence management systems. Such a format would also permit better cooperation between both national and international agencies.

Today's digital forensics community is faced with a significant need and a growing urgency for coordinated technical effort in this area. The lack of a generally accepted format for storing all forms of digital evidence is hampering the development of digital forensics as a scientific discipline, and may result in compromised or lost evidence, and significant judicial consequences.

The Common Digital Evidence Storage Format (CDESF) working group is defining an open data format that can store both digital evidence and related metadata. For more details, visit www.dfrws.org/CDESF. C

THE COMMON DIGITAL EVIDENCE STORAGE FORMAT WORKING GROUP INCLUDES Frank Adelstein, Brian Carrier, Eoghan Casey, Simson L. Garfinkel, Chet Hosmer, Jesse Kornblum, Jim Lyle, Marcus Rogers, and Phil Turner.