C6 Defining, Measuring, and Mitigating Errors for Digital Forensic Tools

James R. Lyle, PhD, NIST, 100 Bureau Drive, MS 8970, Gaithersburg, MD 20899*

After attending this presentation, attendees will understand some of the limitations and constraints when trying to establish error rates for digital forensic tools.

This presentation will impact the forensic science community by increasing awareness in the community regarding error mitigation strategies that should be used instead of error rates to establish the reliability of digital tools. This presentation will also aid forensic practitioners in recognizing that asking about error rates of digital tools is asking the incorrect question.

Extraction of digital evidence from digital systems is dependent on software to interpret and present relevant data. The courts need assurance that any testimony based on software is scientifically sound and reliable. The *Daubert* guidelines list testing and establishing an error rate as two criteria for the court to consider before deciding admissibility of evidence in court.

In the context of *Daubert*, *error rate* has more the meaning of statistical Type I and Type II errors (i.e., rates of false positive and false negative decisions for matching questions such as: (1) is a sample from a suspect a match to a sample found at a crime scene?; and, (2) is a sample found at a crime scene a match to an item in a data base?)

These matching questions come up in several contexts (e.g., DNA, tool marks, finger prints, etc.). The answer to these questions can exclude a suspect from further consideration or identify a new suspect for closer investigation. Matching questions usually have a random component and can be treated as a statistical hypothesis test with false positive and false negative error rates that can be computed and stated.

These questions are also seen in some situations in digital investigations, such as using cryptographic hashing algorithms to determine if two files match. Hashing reduces an arbitrary length file, possibly quite large, to a short fixed length (128 bits for MD5, 512 bits for SHA3-512) hash value. Error rates can be constructed for hash algorithms: (1) if hashes differ, then the files differ; error rate is zero; and, (2) if hashes match, the chance that there is a hash collision (different files with the same hash) is 1 in $2^{n/2}$, where n is the hash length.

These error rates contribute to satisfying the question (of concern to *Daubert*) of the reliable scientific basis for using hashing to identify file matches; however, hashing illustrates that there are more issues to consider. For digital forensic tools, the implementation must also be considered. Errors in an implementation are not usually statistical in nature and are often triggered by a combination of non-random factors. Unlike elements of DNA within a human population that are stable and change slowly over time, the factors that are relevant to digital evidence change with the pace of technological evolution.

This talk will address how to define error for digital forensic tools, how to use tool testing to identify tool behaviors that are relevant to an investigation, and how to use the knowledge gained from tool testing to mitigate incorrect tool behaviors.

Digital Evidence, Software Testing, Error Rate