Digital Forensics

We're pleased to present one of a unique pair of special issues focusing on digital forensics, an inherently multidisciplinary field that involves aspects of computer science and engineering, signal processing, and criminal justice, to name a few.

Michael A. Caloyannides
Northrop Grumman

Nasir Memon
Polytechnic Institute of New York University

Wietse Venema
IBM T.J. Watson Research Center

To date, work in this area has been very fragmented, so the current issues of *IEEE Security & Privacy* and *IEEE Signal Processing Magazine* represent an effort by the IEEE Computer Society and the IEEE Signal Processing Society to bring the two communities together to understand each other's contributions to the field. This issue of *S&P* focuses on computer forensics, whereas the *SPM* special issue focuses on media forensics.

An Emerging Field

In the years since World War II, computers have slowly but unavoidably become record keepers of human activity. This trend accelerated with the introduction of PCs, handheld devices, the Internet, and the convergence of computing, multimedia, and telecommunications. Today's interconnected world of digital devices presents opportunities and challenges for criminals and investigators, for governments and privacy-conscious citizens, and for commercial and other activity.

Computer forensics is a young but rapidly evolving discipline. Borrowing from principles that have proven themselves in the physical world, it faces challenges that are unique to the cyberspace domain. Here, we'll investigate the ongoing debate about the effectiveness of computer forensics, static versus dynamic analysis, and the legal implications of a fast-moving technological domain.

The Articles

We start by looking at a thought-provoking debate. In "Forensics Is So 'Yesterday,'" Michael A. Caloyannides takes the provocative position that computer forensics isn't effective against antiforensic techniques and thus won't be useful for catching sophisticated criminals and agents. Instead, computer forensics will catch naive crooks who don't know how to hide their tracks and innocent people who don't know how to protect their systems.

In his rebuttal, "Digital Forensics Works," Brian Carrier compares the digital world's forensic processes and challenges with those in the physical world. One important difference is that the laws of nature are constant, whereas the laws in the digital world are subject to change with each new generation of hardware and software. Aside from differences in certainty levels, Carrier argues that digital forensics works and that it's effective in much the same way as physical-world forensics.

Historically, computer forensics has focused on static analysis—that is, the analysis of data from a halted computer system. Although this approach maximizes result reproducibility, it misses dynamic state information, such as processes and network connections, memory-resident malware, unlocked file system decryption keys, or data in output buffers that isn't yet written to file. In "Live Analysis: Progress and Challenges," Brian Hay, Matt Bishop, and Kara Nance explore the challenges and opportunities of live analysis—that is, the analysis of data gathered while a system is operating. The most significant

challenge here is how to gather data without introducing distortions, especially when you must rely on the running system's integrity to execute the data collection software correctly. Even live data collection using specialized hardware comes with opportunities for introducing distortion. Virtual computing presents new challenges and opportunities. On one hand, it enables continuous recordings of a virtual machine's complete state, without running data-gathering software inside the virtual machine itself; on the other, it doesn't entirely eliminate the possibility of distortion or detection by an opponent. Hay and his colleagues summarize the challenges with both real and virtual computing through several intriguing research questions.

As mentioned earlier, the laws of computing are subject to revision with each new generation of hardware or software, and a perfect example is the recent development in Microsoft Office document file standardization. Two competing standards have emerged: OOX (Office Open XML) from Microsoft and ODF (OpenDocument Format) from the Oasis (Organization for the Advancement of Structured Information Standards) consortium. Both OOX and ODF store documents as ZIP files that contain a combination of XML-formatted content and binary objects such as images. Both standards introduce levels of redundancy that can help with the forensic recovery of information from partial or damaged files. In "New XML-Based Files: Implications for Forensics," Simson Garfinkel and James Migletz examine the issues of data and metadata in these document formats for forensic analysis. The prevalence of these file types is still relatively small, but it's increasing rapidly with the deployment of newer software versions.

Besides the technical challenges that get ample attention in this special issue, forensic investigators must also be aware of the legal issues regarding the admissibility of evidence and whether they must have a license before legally performing a forensic investigation. Laws are updated frequently, and in the US, forensic investigator licensing requirements vary with each individual state. In "Overview of Licensing and Legal Issues for Digital Forensic Investigators," Gavin Manes and Elizabeth Downing present an overview of the current US federal rules of evidence that individual states are slowly adopting and of the confusing state of affairs that currently exists with respect to forensic investigator licensing.

As the size of data sets continues to grow over time, so does the challenge of identifying file content. File hashing is a technique that computes one or more cryptographic hashes from a file's contents with a collision-resistant function such as SHA512, SHA256, SHA-1, or MD5. These functions transform an arbitrary-length input into a fixed-length output of 128 to 512 bits. The relatively short outputs allow for compact databases and fast comparisons: when two hash values differ, then you know for certain that the inputs differ, too. In "Hashing and Data Fingerprinting in Digital Forensics," Vassil Roussev presents recent developments that improve file hashing's scalability and increase its applicability. Scalability improves by representing a large number of file hashes in a relatively small amount of memory through a technique based on Bloom filters. Applicability increases by computing hashes over input fragments, so that versions of the same information are identifiable even when mixed with different information. Based on Rabin-Karp pattern matching, the technique first became popular for text identification, but it produces promising results for nontext content as well.

The two issues together represent the state of the art in digital forensics as viewed from two different communities. We hope this leads to increased cross-fertilization among the two communities. □

Michael A. Caloyannides *works for Northrop Grumman and is an adjunct member of the faculty of Johns Hopkins and George Washington Universities in computing security and networks. He has worked as a senior scientist in private industry as well as in a US government agency, the latter of which awarded him its "Scientist of the Year" award. Caloyannides has a PhD in electrical engineering, applied mathematics, and philosophy from the California Institute of Technology. Contact him at micky@ieee.org.*

Nasir Memon *is a professor and the director of the Information Systems and Internet Security (ISIS) Lab (http://isis.poly.edu) at Polytechnic Institute of New York University. His research interests include digital forensics, information security, and data compression. Memon has a PhD in computer science from the University of Nebraska at Lincoln. Contact him at memon@nyu.edu.*

Wietse Venema *is a research staff member at the IBM T.J. Watson Research Center. He is coauthor of* Forensic Discovery *(Addison-Wesley, 2004) and the open source Coroner's Toolkit, and author of the Postfix mail system, the TCP wrapper, and other software. Venema has a PhD in physics from Groningen University. Contact him at wietse@porcupine.org.*