C16 Challenges in Recovering Deleted Data in the Cloud

Robert Jackson, MS, SphereCom Enterprises, Inc, 7900 Sudley Road, Ste 416, Manassas, VA 20109; Richard Austin, MS, Hewlett-Packard, 5555 Windward Parkway, Alpharetta, GA 30004; Martin Herman, PhD, 100 Bureau Drive, MS 2000, Gaithersburg, MD 20899; P.W. Carey, MS, Compliance Partners, LLC, 250 S Grove Avenue, Barrington, IL 60010; and Otto S. Reemelin, MS, CBIZ, Inc, 3101 N Central Avenue, Ste 300, Phoenix, AZ 85012*

After attending this presentation, attendees will have a better understanding of the challenges faced by forensic investigators attempting to recover deleted data and metadata in cloud computing environments.

This presentation will impact the forensic science community by describing research performed by the National Institute of Standards and Technology (NIST) Cloud Computing Forensic Science Working Group, which was established to aggregate forensic science challenges in the cloud environment and to develop plans for measurements, standards, and technology research to mitigate those challenges that cannot be handled with current technology and methods. One of the highest priority challenges is recovering deleted data in the cloud.

Data deletion in the cloud is often based on the deletion of nodes pointing to information in virtual instances. Whether the deletion of the information has been fully achieved needs to be assessed and proven. Pathways for retrieval are dependent on cloud providers offering sufficiently sophisticated mechanisms for access. Recovery of data marked as deleted is difficult since it may get overwritten by another user in a shared virtual environment. The challenge becomes more difficult if the entire virtual environment is also deleted.

Issues in recovering deleted data in the cloud include the huge volume of dynamically and continually changing data; cloud resources previously assigned to the user may be unknown; deleted data are overwritten very quickly; there are multiple locations for any given data (data are moved around multiple servers and storage rapidly); there are likely to be multiple copies of data, leading to multiple deletions of data; and there is uncertainty about the proper owner of deleted data. Other issues include the geographically dispersed "incident scene" and the involvement of multiple organizations in multiple jurisdictions.

One important aspect is the role of end points. End points may contain file remnants, contact information, addresses, etc. that will assist investigators in identifying where the deleted data were stored or processed, and identifying other cloud resources that were provisioned; however, in many situations, such as criminal activity or cyber attacks, the end points that were used by the individual(s) involved will likely be inaccessible to investigators.

This presentation will describe the various characteristics dealing with data deletion in the cloud, the effects of various service and deployment models, technical requirements for recovering deleted data in the cloud, sample use cases that help highlight the challenging issues, and the role of standards and technology. For example, standards are needed for data retention (e.g., backups for live data, Virtual Machine (VM) snapshot images, audit trails, transaction logs) to allow easier development of forensics tools. From a technology perspective, forensic tools exist, but the biggest gap is the "needle in the haystack" issue (i.e., there is such a huge volume of data which is continually and dynamically changing that finding the deleted data, and then being able to attribute it to an individual, is a significant challenge).

Digital Forensics, Cloud Computing, Deleted Data Recovery