THE JOURNAL OF DIGITAL FORENSICS, SECURITY AND LAW

Journal of Digital Forensics, Security and Law

Volume 6 | Number 3 Article 5

2011

A Survey of Contemporary Enterprise Storage Technologies from a Digital Forensics Perspective

Gregory H. Carlton
California State Polytechnic University

Joseph Matsumoto
California State Polytechnic University

A Survey of Contemporary Enterprise Storage Technologies from a Digital Forensics Perspective

Gregory H. Carlton
California State Polytechnic University
ghcarlton@csupomona.edu

Joseph Matsumoto
California State Polytechnic University
jmatsumoto@csupomona.edu

ABSTRACT

As the proliferation of digital computational systems continue to expand, increasingly complex technologies emerge, including those regarding large, enterprise-wide, information storage and retrieval systems. Within this study, we examine four contemporary enterprise storage technologies. Our examination of these technologies is presented with an overview of the technological features of each offering and then followed with a discussion of the impact of these technologies on digital forensics methods, particularly regarding forensic data acquisition. We offer a general opinion concerning a recommended data acquisition method when faced with the task of obtaining a forensic image of data contained within these technologies, we discuss limitations of our study, and lastly, we suggest areas in which additional research would benefit the field of digital forensics.

1. INTRODUCTION

The evolution and adoption of contemporary enterprise data storage technologies provide challenges to traditional approaches for computer forensic data acquisition (Mohay, 2005). Some of these technologies invalidate the once simple relationship between a single storage device or small group of storage devices and a volume of data presented to a computer or server. In the absence of this simpler relationship, it will be increasingly difficult to follow traditional forensic data acquisition procedures in obtaining a forensically sound image of the data (Carlton, 2007). This article explores some of these technologies and provides an overview of the challenges they may pose. Alternative approaches for dealing with the challenges are also explored.

2. OVERVIEW OF TRADITIONAL FORENSIC DISK IMAGE ACQUISITION METHODOLOGY

The traditional digital forensic disk image acquisition processes provide the forensic examiner with physical access to the computer systems, including any disk storage media. In these situations, forensic examiners have access to the

physical disk drives and are able to acquire static, or dead, forensic drive images from each of the physical disk drives. In situations involving servers, the traditional dead acquisition utilizes a standard server shutdown procedure. A hardware or software write blocker is used to prevent altering data on the evidence drive(s) and a bit-stream image of the evidence drive is made (Lessing & von Solms, 2008). The traditional forensic data acquisition approach for servers utilizing an array of multiple physical disk drives into a logical volume, RAID configuration, provides the forensic examiner with two alternatives for static data acquisition. He or she may statically image each of the individual physical drives and then reassemble them into a logical disk volume using their forensic analysis tools, such as EnCase. In situations where the forensic examiner is not able to reassemble the individual disks into a logical volume, the forensic examiner may statically image the logical RAID volume from the server.

An alternative to static data acquisition is a live disk image acquisition. While a static data acquisition is arguably preferred from scientific and legal perspectives, practical objections based on temporal and fiscal factors frequently restrict the static data acquisition of servers. Citing these objections as a justification for best evidence, a live acquisition alternative may be necessary when the target server cannot be taken down for forensic drive imaging. Additionally, a live acquisition is a viable alternative when the forensic examiner does not have physical access to the evidence storage media. Live acquisitions typically utilize the use of a software agent or acquisition tool on the target. This may require the installation of the acquisition software on the host server after the fact, unless it was proactively installed. This technique introduces an alteration to the data that would need to be documented, and accepted as best evidence. A forensic drive image obtained in this manner would be similar to the logical drive image described above as it is acquired from the host end and not directly from the disk drives themselves. Tools such as EnCase Enterprise Edition can be used to perform a live acquisition when static imaging is not possible (Guidance Software, 2011). An additional challenge to this live acquisition method concerns the realization that data is dynamic during the live acquisition process; therefore, the ability of validating the acquisition process through duplication will not produce matching hash values.

The static and live data acquisition methods have been recognized in the practice of digital forensics for over half of a decade. While both of these methods have their merits and limitations, these traditional forensic data acquisition methods were based on assumptions that suspect computer systems consisted of physical computational devices and physical storage media. In the following section, we discuss aspects of contemporary storage technologies that do not necessarily hold these assumptions.

3. DISCUSSION OF CONTEMPORARY ENTERPRISE STORAGE TECHNOLOGIES

As advances in computational hardware, software, and information storage and retrieval technology continue to develop, numerous alternatives are emerging into the marketplace that challenge the traditional notion of physical computational devices and physical storage media. Within this section, we discuss four emerging storage technologies targeted at the enterprise storage market, namely: enterprise storage arrays and logical volumes, automatic storage tiering, data deduplication, and thin provisioning. While these technologies differ in their approach regarding information storage and retrieval, each provides specific challenges for digital forensic examiners.

3.1 Enterprise storage arrays and logical volumes

Enterprise storage arrays and logical volumes focus on storage technologies commonly found in enterprise storage area network (SAN) storage arrays. The use of large enterprise storage arrays introduces multiple challenges to the traditional forensic disk image acquisition where direct access to the physical storage media is separated by an additional level of abstraction. These logical storage arrays can consist of hundreds of physical disk storage devices that are used in the creation of logical units (LUN). LUNs are logical partitions of redundant array of independent disks (RAID) groups from the storage system (EMC Corporation, 2006). Hosts deriving storage from the storage system see a LUN as they would an individual disk drive. The host is unaware of the exact physical makeup of the LUNs delivered by the storage array. A LUN may consist of a single physical drive or many drives in RAID1, RAID5, RAID10, or other arrangement. This additional level of abstraction raises the following challenges to the process of forensically acquiring data from a server utilizing enterprise storage arrays:

1. Negative impact of downtime – The multi-user nature of servers frequently result in financial and legal objections to taking a server offline for the period of time necessary for forensic examiners to obtain a static image of the data. These objections are likely to be compounded in configurations utilizing enterprise storage arrays, as multiple servers may share the storage arrays. Given the multi-user nature of servers and the impact on multiple servers, it may not be feasible to take the server down the time necessary to perform a traditional static drive acquisition due to the broader impact of the downtime.
2. Additional complexity - The fact that the association between physical drives and the server is not as straightforward adds further complexity. It would take collaboration with the storage system administrator to identify all involved storage devices as well as the specific RAID algorithm and parameters in order to access the

drives for forensic imaging and later reassemble the data into a usable form. In addition to the challenge presented by this additional technical complexity, additional legal complexity will likely surface in situations where physical storage devices are logically shared among different legal entities.

3. Broader impact of shared storage - Even if the server can be taken out of service, the storage devices (disk drives) might not be dedicated to just the server in question. This means that additional services would be impacted if drives were physically removed for forensic disk imaging.

Despite the challenges identified above, it may still be feasible to perform a physical forensic drive acquisition in situations where a manageable number of drives are involved and where they can be physically accessed. A logical acquisition of the LUN itself is also still possible. The remaining storage technologies that will be discussed will build on top of the concept of LUNS in presenting a logical disk or storage volume to server hosts, and present additional challenges which may prohibit the more direct forensic imaging of physical disk drives.

3.2 Automatic storage tiering

Storage and server administrators have been using a tiered approach for storage management for many years. Administrators would allocate storage to servers based on performance and capacity needs, and this could be done on a per-LUN basis as required. Tiering involves allocating LUNs from various groups of storage devices which are grouped by criteria, such as, differing performance levels and capacities. For example, one might configure three tiers using these criteria: a capacity tier based on low-cost, high- capacity SATA drives, a performance tier based on faster, fiber channel drives, and even an ultra-high performance tier based on solid state drives. In this scenario, lower priority or bulk storage applications would be allocated LUNS from the lower tier, while higher priority applications would require storage from the high performance tier or ultra-high performance tier. The manual allocation of LUNS from different tiers of storage does not necessarily introduce additional significant challenges to forensic imaging.

However, unlike the manual allocation of LUNS described above, automated storage tiering introduces significant changes to the scenario. Automated storage tiering allows a storage system to automatically identify “hot data” and move that data to higher performing data storage devices (Feresten, Freeman, & Woods, 2011). This simplifies storage data management by eliminating or reducing manual storage management, and it has the potential to reduce storage costs by using low-cost storage where permitted while intelligently moving high-use data to faster storage devices.

When automated storage tiering is utilized, there are two important differences to consider regarding data movement. First, data are moved between storage tiers automatically as dictated by the storage system for optimized performance. Second, the entire volume or LUN is not necessarily moved. Individual blocks or predefined chunks of data are moved between storage tiers. The storage system maintains a bitmap record of the physical location of all the data for each LUN (Hernandez, 2011). The movement of data is done transparently to the host server. The unit size of data moving between the different storage tiers is specific to a vendor's own implementation. NetApp's implementation allows for the movement of data as small as a 4k block between storage tiers (Feresten, Freeman, & Woods, 2011). EMC moves data between different storage tiers in 1GB chunks (Hernandez, 2011).

Automated storage tiering introduces the variable of dynamic change in the physical location of data across a number of storage devices. This eliminates having a fixed, defined space on discrete disk drives where all data for a given LUN may reside, thus making it impractical, if not impossible, to directly acquire individual forensic disk images to reconstruct a logical disk volume. As a result of this limitation, it is then necessary for the forensic examiner to rely on the acquisition of a logical disk image from the server's view of the LUN or logical disk volume.

3.3 Data deduplication

Data deduplication is another storage technology that presents additional challenges to the traditional forensic disk image acquisition process. Data deduplication is a storage compression tool used to reduce storage capacity requirements, and it is used in both inline storage (i.e., online, secondary storage) and backup storage functions. Data deduplication in disk storage refers to the use of an algorithm that searches for duplicate data, and then removes the duplicates. Duplicate data is replaced with reference pointers to a single copy of identical data (Freeman & NetApp, 2009). Data deduplication is generally done at a block or file level and involves processing the data with a hash algorithm such as SHA-1 or MD5 to generate an index value for each block or file (Bigelow & Hawkins, 2008). When new data are being written, the index value created is compared to a maintained bitmap or index table for existing data to determine if the new data already exists. New data is written to the logical storage device, while duplicate data is simply accounted for in the bitmap or index table without writing another instance of the data to the logical storage device (Hernandez, 2011). In addition to the inline technique of data deduplication described above, these algorithms can work a post-process. When data deduplication is configured as a post-process, data are written to the target LUN without regard to duplication, and a scheduled process periodically evaluates all newly written data, removing duplicates to reclaim the space while placing the appropriate pointers to the retained single copy.

It does not seem feasible to acquire a forensic image from disk drive where data deduplication is utilized, as it is unrealistic to expect a forensic examiner to reconstruct data from this image. Reconstructing the data from the image would require, at a minimum, a complete understanding of the applicable vendor's specific deduplication implementation. An additional obstacle is the bitmap index table used to identify deduplicated files or blocks may not reside on the same disk drives associated with the LUN. This bitmap index is necessary to identify or reconstruct all of the missing pieces of data. Based on the unlikely probability of successfully analyzing data from a physical image of a deduplicated disk, once again, we recommend forensic examiners obtain an image of logical volumes for deduplicated disks.

Since data deduplication is based on the use of hashing algorithms to identify duplicate data, there is also the theoretical possibility of hash collisions resulting in the incorrect deduplication. The non-duplicate data potentially could be lost and replaced by the misidentified duplicate. Storage vendors should have sufficiently mitigated this risk, but it is important for a forensic examiner to be aware of this potential, at least to defend the validity of any forensic evidence obtained from a LUN where deduplication is in use. The forensic examiner should also be aware of the level (block or file) at which any deduplication is being performed.

3.4 Thin provisioning

Thin provisioning is a storage technology used to optimize the efficiency of storage capacity usage in enterprise storage systems. Floyer states, "With thin provisioning, a storage administrator allocates logical storage to an application as usual, but the system releases physical capacity only when it is required. When utilization of that storage approaches a predetermined threshold (e.g. 90%), the array automatically provides capacity from a virtual storage pool which expands a volume without involving the storage administrator" (Floyer, 2009). Thin provisioning initially allocates only a portion of the physical space for a LUN. The storage system maintains a map of allocated physical storage which may be stored in the storage system's cache memory or private disk (Hernandez, 2011). Without access to this mapping facility of physical storage, the correct physical storage devices cannot be identified; therefore, forensic examiners will find it practically impossible to identify all physical storage devices necessary for forensic images. Even greater challenges than identifying the appropriate physical storage devices for the forensic examiner are the problems of identifying the specific allocation blocks or physical sectors associated with the thin provisioned volume and the task of reassembling the logical disk volume.

A technique for obtaining the contents of a thin provisioned logical volume is to generate a host-based dump of the thin provisioned logical volume. This host-based dump would read all of the data from the target logical volume allocated to physical storage devices and output a bit-stream copy of this data to a designated

image file. This is necessary, as the host servers are unaware that only a portion of the logical volumes' space resides on physical storage. This is accomplished by the storage system providing the host server with null characters (i.e., binary 0s) for the remaining non-provisioned space of the LUN. This technique effectively spoofs the host's view of the LUN for the unallocated space (Hernandez, 2011). Often, the server's view of its storage and the storage system's view of the storage associated with a given server are quite different.

Some storage systems provide the capability to add additional disk drives and then rebalance allocated space of existing LUNS across the added devices. This is can help spread the IO workload across the entire group of disk drives. This is particularly important when the data are striped across a group of drives. This results in shuffling data across the physical drives providing storage for a LUN, creating another circumstance in which significant data are being moved across storage devices by the storage system without any host awareness.

Some implementations of thin provisioning include a facility for reclaiming unused space. This space reclamation provides another challenge for forensics examiners by removing unallocated clusters. Jooss explains, "Space Reclamation is the process of allowing the storage system to free the blocks no longer used by the host operating system" (Jooss, 2008). The space reclamation process may vary across vendor implementations, for example, some vendors support host-based processes to identify unused space and communicate using T10 industry-standard, SCSI commands including UNMAP and WRITE_SAME unmap. This allows the host operating system to communicate unused logical blocks that can be reclaimed (EMC Corporation, 2011). For NTFS file systems, the SDELETE utility can be used with the "–c" option to overwrite space occupied by deleted files with zeros. A process on the storage system will then release space identified by the host process as unallocated, as well as, any other space containing all null characters, indicating unused space (Hernandez, 2011). For the forensic examiner, the completion of this process is likely to further limit the potential to obtain evidence from unallocated clusters.

In review, we find that the current best practice for acquiring a forensically sound image of a thin provisioned volume is to acquire a logical image of the LUN. The forensic examiner should understand that space reclamation has the potential to also remove traces of prior deleted data as well. At the present time, space reclamation is still a fairly new technology and is not as widely available as the other storage technologies discussed here; therefore, while this reduces the probability that forensic examiners will encounter this technology in the field at this time, it also increases the probability the specific forensic examiner that encounters thin provisioning at this time will be unaware of the technology and its challenges. As with many emerging technologies, we anticipate that thin provisioning will become more widely implemented in the future, thus raising the probability of this being a more significant concern for forensic examiners.

4. SUMMARY OF FINDINGS

Moore's Law is frequently cited to document the rapid growth of performance and capacity within the field of computational and digital devices, as "Dr. Gordon Moore, then a researcher at Intel, hypothesized that computer processing performance would double every eighteen months" (Valacich & Schneider, 2010). Given this rapid growth rate, combined with the reduction in costs associated with storage technology, it is understandable that we are experiencing increasing levels of complexity and capacity in contemporary, enterprise-wide, information storage and retrieval systems. There is a direct relationship between the combined complexity and capacity of these storage technologies and the challenges forensic examiners encounter regarding obtaining useable forensic images of the data contained within these technologies. We have evaluated four, non-mutually exclusive, enterprise storage technologies, and reached an opinion regarding the current best practice for acquiring forensic images of data contained within these technologies.

Each of the enterprise storage technologies presented above pose added challenges to traditional forensic data acquisition methodology. Digital forensics researchers and practitioners should be aware that it is also possible for many of these storage technologies be used in conjunction, further adding to these challenges. As a matter of practicality, we find that the best practice for forensic examiners to utilize when tasked to obtain a forensic image of data stored within configurations based on these enterprise storage technologies is to rely on a logical acquisition by taking a forensically-sound, bit-stream image of the LUNs themselves, and not the physical disk drives. As overview of the basis for reaching our opinion is presented below:

The first set of factors supporting this opinion is based on matters of practicality. While a static or dead acquisition of the target LUN can be done from the target server, or a live acquisition can be performed, either of these is likely to result in an unacceptable amount of downtime. The static acquisition will result in extended downtime for the associated server. This may be deemed unacceptable, forcing the use of a live acquisition process. However, even a live acquisition has the potential to impact the overall performance of a server should a full acquisition of any disk volumes be necessary. There is still the I/O burden associated with a full drive acquisition, and this should be taken into consideration when deciding on the best approach for data acquisition.

In addition to the amount of time necessary for either a static or live acquisition described above, contemporary enterprise computational models are likely to consist of multiple servers, either actual physical servers or virtual servers, and multiple legal entities may be serviced by these physical or virtual devices. It becomes increasingly difficult to define the bounds of legal access to forensic storage targets when multiple servers share virtual portions of physical storage devices, especially when multiple legal entities are involved.

In addition to the practical factors presented above, the four contemporary enterprise storage technologies evaluated utilize virtual storage techniques that make reconstructing logical volumes from physical devices improbable for forensic examiners.

While much the focus of this article has been to identify potential challenges presented by newer enterprise storage technologies, there is also the potential for them to provide some assistance to the forensic examiner. Many enterprise storage systems provide the capability to produce a snapshot or clone of a LUN. This capability has the potential to provide temporal duplicates of data from logical volumes. An additional advantage of creating a snapshot or clone from the storage system is that this process generates a relatively low processing burden on the host server. The snapshot or clone can be done with the server online, offline, or even powered off since it is performed by the storage system itself. Also, this technique can be performed at a specific instant for the entire LUN, addressing potential concerns about changes to a live file system during the time required to take a forensic image of a volume. The snapshot can be created and then presented as a LUN on a different server, thus providing a configuration that can then be used by a forensic examiner to create the logical forensic disk image without an impact on the original host server. In situations where snapshots or clones of LUNs are available, we suggest acquiring the snapshot or clone as an alternative instead of acquiring a logical image of the LUN, as this provides the best method to obtain a logical, forensically sound, data acquisition.

However, forensic examiners should realize that the combination of thin provisioning and space reclamation introduces a potential for non-recoverable deleted data. This should be of concern to forensic examiners, as these technologies become more widely implemented.

5. LIMITATIONS

Within this study, we have identified and discussed four technologies available in contemporary, enterprise storage systems. These four technologies were selected from our survey of the currently available offerings and emerging trends within the enterprise storage system market. While we did not knowingly omit other enterprise storage systems technologies from inclusion within this study, the possibly exists that other relevant technologies either currently exist or are emerging within this market. The reader should also be aware that, given the rate of technological advances, it is likely that additional technologies will enter the marketplace in the future.

There is, however, an entire facet to computing infrastructures, including complex and remote storage systems that we did not address in this study, namely cloud computing. While we do not directly address cloud within this study, it is helpful for the readers to recognize that the enterprise storage facilities described within this paper are also applicable to cloud computing. Cloud computing encompasses many aspects of computing environments, including applications, servers, storage

systems, and network devices. Additionally, cloud computing may be implemented for a specific application, for an entire enterprise level, or anywhere between those two extremes. We thought that a discussion on cloud computing within this paper would distract from the important forensic challenges concerning enterprise storage facilities; therefore, the primary author of this study, along with a coauthor, have addressed forensic challenges to cloud computing in another study, and we would like to refer interested readers to this work, titled, *A survey of cloud computing challenges from a digital forensics perspective* (Carlton & Zhou, in press).

The purpose of this study is to provide computer forensics practitioners and researchers with an introductory level of understanding of the technologies currently available in enterprise storage systems, and then to identify challenges forensic examiners will likely face when tasked with acquiring digital evidence from systems using these technologies.

Our evaluation methods used within this study are based on research of secondary data, and much of these data were provided to us directly by technology vendors. We did not develop any primary data within this study, nor did we conduct experiments using these technologies; therefore, a limitation exists based on the extent to which the secondary data provided to us are accurate. Similarly, given the for-profit motives of the technology vendors, a limitation exists regarding the possibly of exaggerated claims of technical performance without verified, independent testing in controlled laboratory conditions.

Lastly, the information presented within this study is based on our understanding of the data from a technical perspective, as well as, our understanding of current digital forensics data acquisition and analysis methodologies. We are not attorneys; therefore, we are not providing legal advice in our opinions. Rather, our opinions are presented to offer guidelines for forensic examiners, thus allowing them to consult with attorneys from an informed perspective.

6. CALL FOR ACTION

As indicated within the section on limitations, this study was conducted from secondary data. A more thorough study utilizing laboratory experiments of data stored within the technologies identified within this study would be helpful to the digital forensics community. Studies involving experiments with known, in-tact data, deleted data, and otherwise concealed data acquired by an array of methodologies, such as, static, physical acquisition, live acquisition, and static, logical acquisition from both the server's perspective and from the perspective of the LUN would be of particular interest.

Also, more research would be beneficial pertaining to the exploration of the use of snapshot or clone technologies, specifically addressing forensic considerations. Areas of particular interest include forensically sound data acquisition guidelines of snapshot or clone LUN configurations and forensic data analysis of snapshot or

clone LUN configurations, especially techniques for virtual restores or protected environment restores.

An additional topic of research that would enhance the understanding of enterprise storage technologies from a digital forensics perspective would be to identify markers within the data that would help identify the storage technologies in use. For example, as described above in section 3.4 Thin Provisioning, some technologies remove unallocated data; therefore, it would be beneficial to a forensic examiner to know that this technology is in use within an image he or she is analyzing. Identifying these markers, if they exist, within the data that identify the storage technology in use will provide a significant contribution to the understanding of enterprise storage technologies from a digital forensics perspective.

REFERENCES

Bigelow, S. J., & Hawkins, J. (2008, September). *data deduplication (Intelligent compression or single-instance storage)*. Retrieved May 26, 2011, from SearchStorage.com: http://searchstorage.techtarget.com/definition/data-deduplication

Carlton, G. H. (2007). A Protocol for the Forensic Data Acquisition of Personal Computer Workstations. *(Doctoral dissertation). (UMI No. 3251043)* .

Carlton, G. H., & Zhou, H. (in press). A survey of cloud computing from a digital forensics perspective. *International Journal of Interdisciplinary Telecommunications and Networking* .

EMC Corporation. (2011). *EMC Symmetrix VMAX Virtual Provisioning Space Reclamation and Application Considerations.*

EMC Corporation. (2006). *EMC Virtual LUN Technology, A Detailed Review.*

Feresten, P., Freeman, L., & Woods, M. (2011). *The NetApp Virtual Storage Tier.* NetApp.

Floyer, D. (2009, October 9). *Thin Provisioning*. Retrieved May 30, 2011, from wikibon.org: http://wikibon.org/wiki/v/Thin_provisioning

Freeman, L., & NetApp. (2009). *Looking Beyond the Hype: Evaluating Data Deduplication Solutions.*

Guidance Software. (2011). *EnCase Legal Journal.*

Hernandez, N. (2011, 5 25). EMC Technnology Consultant. (J. Matsumoto, Interviewer)

Jooss, R. (2008). *Technical Report: Thin provisioing in a NetApp SAN or IP SAN Enterprise Environment.* Network Appliance Inc.

Lessing, M., & von Solms, B. (2008). *Live Forensic Acquisition as Alternative to Traditional Forensic Processes.* Mannheim, Germany.

Mohay, G. (2005). Technical challenges and directions for digital forensics. *First international workshop on systematic approaches to digital forensic engineering* (pp. 155-161). Taipei, Taiwan: IEEE.

Valacich, J., & Schneider, C. (2010). *Information Systems Today, Managing in the Digital World* (4th ed.). Upper Saddle River, NJ: Prentice Hall.