Chapter 3

A NETWORK-BASED ARCHITECTURE FOR STORING DIGITAL EVIDENCE

Mark Davis, Gavin Manes and Sujeet Shenoi

Abstract The storage and handling of digital evidence are creating significant challenges for federal, state and local law enforcement agencies. The problems include acquiring and processing massive amounts of digital evidence, maintaining the integrity of the evidence, and storing digital evidence for extended periods of time. This paper describes a network-based storage architecture that helps address these issues. The architecture also supports collaborative efforts by examiners and investigators located at geographically dispersed sites.

Keywords: Storage media, network area storage, storage area networks

1. Introduction

Law enforcement agencies are facing major challenges with regard to the storage and processing of digital evidence [6, 17]. Complex cases are being encountered that require evidence to be extracted from networks, multi-drive computers and sophisticated portable electronic devices [14, 15]. Most cases still involve single hard drives, but hard drive capacities can be very large [1, 11]. In a recent case, the Tulsa (Oklahoma) Police Department's Cyber Crimes Unit seized a personal computer with three 250GB hard drives. New hard drives were purchased to handle the large volume of data. However, the unit's imaging workstations relied on ATA-100 technology, which could not support drives larger than 137GB. New equipment based on ATA-133 technology had to be purchased so that the larger hard drives could be used to process evidence.

The long-term storage of digital evidence is also presenting serious problems for law enforcement agencies [6, 14, 17]. Sometimes, evidence has to be maintained only for the duration of a trial. In other instances, evidence must be stored for the length of the sentence. A recent triple

homicide case in Tulsa involved more than 350GB of digital evidence. The 27 year-old accused received a life sentence without parole, which could require that all the evidence in the case be stored for 50 years or more. Digital storage media degrade over time and few, if any, media can guarantee the integrity of the stored evidence beyond fifteen years [4, 16, 17]. Special environmentally-controlled storage rooms can help extend the life of certain media, but these are very expensive.

Meanwhile, digital media technology is constantly changing. Currently, it is difficult to obtain a 5.25" floppy drive, although it was the primary removable storage medium just fifteen years ago. Evidence stored on an IDE hard drive may not be accessible twenty years from now because the hardware might not be readily available [17].

Evidence handling – especially maintaining the chain of custody – is a strict and meticulous process that requires special consideration with regard to digital evidence [10]. Digital evidence is easily moved and copied, making it difficult to document who had access to the evidence and when the evidence was accessed. Moreover, digital evidence must be protected using physical access controls as well as computer-based access controls [2]. Since most law enforcement agents are not computer security experts, it can be difficult for them to ensure that the integrity of the evidence is maintained.

Digital forensic procedures must also be reliable enough to withstand courtroom scrutiny. Law enforcement agents compute hash values of image files to verify their integrity, but problems arise when the integrity of an image is lost. In such cases, the original storage media must be re-imaged [10, 17]. However, the media may not always be available or it may be damaged or destroyed.

The sheer volume of evidence involved in many cases requires examiners and investigators, who may be at different geographic locations, to cooperate in digital forensic investigations. What is needed is an efficient methodology for storing, moving and examining data across geographic boundaries. The ideal implementation is a centralized repository where evidence is stored and maintained, but which allows the evidence to be securely accessed from remote locations. Furthermore, the system must be technologically transparent and it should eliminate the need for forensic examiners and investigators to perform systems administration duties.

This paper describes a network-based solution for storing and handling large quantities of digital evidence. The design is intended to streamline digital forensic investigations and support the collaborative analysis of digital evidence at multiple locations. To provide a framework for discussing the network-based storage solution, the following

section describes the main technologies for implementing networks with massive storage capabilities.

2. Digital Evidence Storage Networks

Two main technologies exist for implementing networks with massive storage capabilities: network area storage (NAS) and storage area networks (SAN). These technologies are discussed below.

2.1 Network Area Storage

Network area storage (NAS) is a solution for storing massive quantities of data in a centralized location [13]. NAS grew out of the file server concept made popular by Netware and Microsoft's Windows NT server [5]. The realization that comprehensive operating systems were not needed to perform storage functions led to the creation of NAS storage devices. These storage devices, with embedded operating systems, are attached to a network and accessed via standard protocols, e.g., TCP/IP. Access control is typically implemented by a network sharing mechanism similar to Windows shares or Samba shares in UNIX [9].

Due to its ease of use, NAS became a popular digital evidence storage solution. In the late 1990s, some FBI laboratories relied on NAS-based SNAP appliances – small rack mountable devices with proprietary operating systems that contain 250GB to 15TB of storage [8]. However, as the protocols for accessing and analyzing digital evidence became more complicated, a more scalable solution than NAS was deemed necessary.

2.2 Storage Area Networks

A storage area network (SAN) is a segmented area of a network that handles storage and data transfer between computers and storage elements [3, 12, 13]. The SAN model removes storage devices and storage-heavy traffic from general networks, creating a network designed exclusively for storage operations. SANs use fibre channel or fabric networks to implement many-to-many connectivity between servers and storage devices. The network-based architecture of SANs makes them highly configurable and scalable, and able to support redundancy.

The addressing scheme used in a fabric network requires that every network device have a unique world wide name (WWN). A WWN is a 64-bit hexadecimal number coded into each device, similar to a MAC address on an Ethernet network. A logical unit number (LUN) is a name given to a RAID set within a storage array. A software client allows LUNs within the disk array to be assigned to WWNs on the network, enabling a LUN to behave identically to a local hard drive on

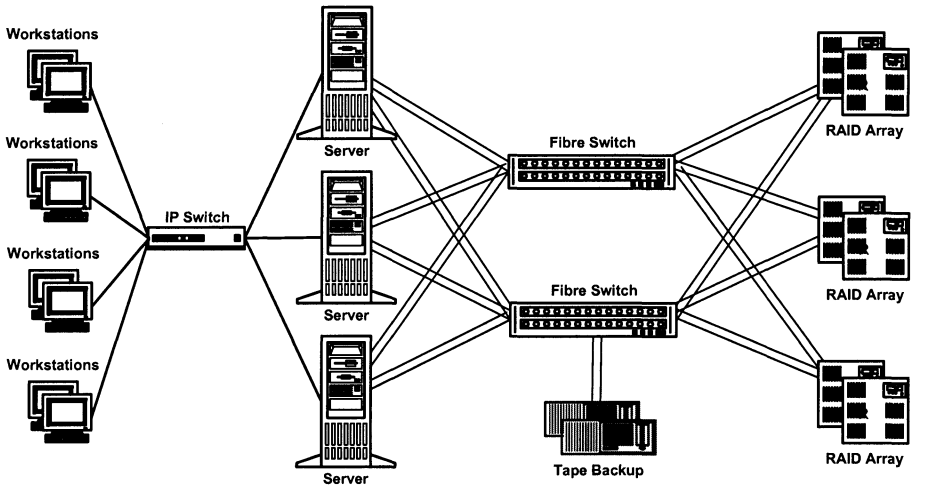

Figure 1. Storage area network.

a computer. LUNs can be re-assigned, unassigned or even increased in size dynamically according to network needs.

Figure 1 shows a typical SAN architecture. RAID disk arrays in the SAN are attached to one or more fibre switches, which in turn are connected to fibre channel cards within computers in the network. Connecting RAID arrays and computers to more than one fibre switch ensures that the SAN and disk arrays have redundant paths to access data during hardware failures. RAID disk arrays speed up data transfer and provide data integrity and redundancy in the event of an accidental loss of digital evidence.

2.3 Combining NAS and SAN Technology

NAS and SAN are similar technologies and either can work well in a given situation [5, 9, 13]. Both technologies use RAID arrays to store data. In a NAS implementation, almost any machine in a LAN can connect to a NAS storage device. However, in a SAN implementation, only computers equipped with fibre channel cards may connect directly to storage devices. A NAS device handles operating system data, backup and mirroring, and data tranfers using operating system metadata (e.g., file name and byte information). On the other hand, a SAN addresses and transfers data by raw disk blocks. A NAS allows data sharing between multiple operating systems (e.g., Unix and Windows NT); a SAN only allows sharing via fibre channel and access is dependent on operating system support. Finally, a NAS typically manages its own file system, while SAN file systems are managed by connected servers.

By combining SAN and NAS technology, LUNs can be shared by user workstations via servers (see Figure 1). This approach is often used by web and file server applications for which availability and load balancing are primary concerns. Usually a portion of the SAN is assigned to a server, which provides services to many clients [3, 12]. The server need not have local storage, resulting in significant cost savings. Server applications, operating systems and storage are easily reassigned, shared or moved from one server to another. For example, if a server fails, its LUN can be reassigned to another server. A LUN can even be assigned to several servers, which means only one copy of the data exists and all the servers would be identical. This also allows many servers and workstations to access and process data simultaneously.

Implementing such a system in a digital forensic environment can drastically improve operational efficiency. Forensic examiners do not have to keep hundreds or thousands of hard drives in evidence storage lockers to preserve evidence. Furthermore, data is transported quickly and easily by reassigning LUNs to different servers [8]. A SAN eliminates the need to manually transport evidence – data is simply assigned wherever it is needed.

The efficiency of a NAS over SAN solution is verified by statistics from the FBI's North Texas Regional Computer Forensics Laboratory (NT-RCFL) [7, 8, 18]. During the four and a half months following the September 11, 2001 attacks, NT-RCFL processed approximately 7.4TB of data using fifteen dedicated examiners. After the NT-RCFL's SAN became operational a year later, an 8.5TB case was processed in one month using only five dedicated examiners. The SAN also helped reduce case backlogs. With its original NAS-based SNAP solution, NT-RCFL had accumulated eight months of case backlog as of September 2001. The NT-RCFL SAN increased data examination rates by a factor of five – the number of examiners fell from fifteen to twelve and the case backlog dropped to just two months.

The NAS over SAN model is an ideal evidence storage solution for a large FBI laboratory, which typically processes and maintains digital evidence at a single location. On the other hand, many federal, state and local law enforcement agencies employ smaller facilities at multiple locations. This requires digital evidence to be delivered, examined and processed at one location, and then physically transported to another location for further examination, presentation or storage. To streamline digital forensic investigations, it is necessary to design a modified NAS over SAN model that facilitates the collaborative analysis of digital evidence at geographically dispersed sites.

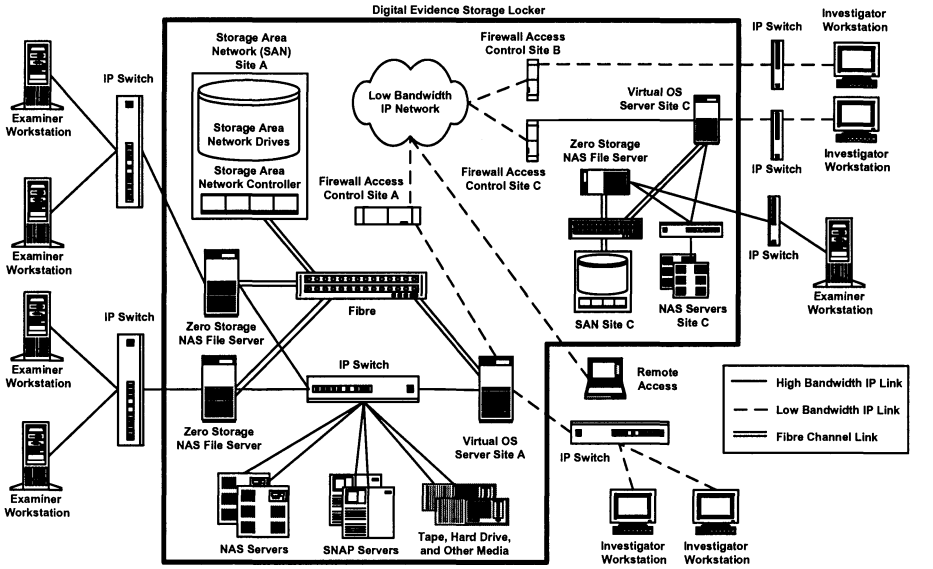

Figure 2. Digital evidence custodian architecture.

3. Digital Evidence Custodian

This section describes the architecture of a digital evidence custodian (DEC), which is intended to streamline investigations and support the collaborative analysis of digital evidence at multiple locations. In many computer crime investigations, the same individual serves as the examiner, investigator and case agent. However, this situation is rapidly changing as caseloads and evidence volume increase, and digital forensic tasks become highly specialized. To streamline investigations, the DEC architecture implements the logical and physical separation of duties of forensic examiners, forensic investigators and evidence custodians.

Figure 2 presents the DEC architecture. Evidence is stored in a digital evidence storage locker (DESL), which primarily uses NAS over SAN technology to facilitate the collaborative processing of digital evidence by examiners, investigators and case agents who may be at different locations. Any storage technology, e.g., NAS, SNAP servers, tape drives or file servers, may be used in a DESL. The DEC architecture eliminates network configuration, administration and maintenance tasks and provides transparency of technology to forensic specialists and agents, enabling them to focus exclusively on case investigations.

Forensic examiner workstations in the DEC architecture (Figure 2) are dedicated computers for imaging storage media. These computers are networked to a zero storage local server in a NAS configuration

to access storage devices located within the internal SAN configuration. The internal NAS and SAN systems comprise the storage locker (DESL).

During an investigation, the DEC dedicates storage space for a case within the DESL to an examiner's zero storage local server. The zero storage local servers share this space with the examiners' workstations, allowing them to image storage media. Depending on the urgency of the case, a forensic examiner may perform analysis functions such as live previews, file filtering and keyword searches on the imaged media. Upon completion of the imaging process, the evidence is stored in the DESL and all access to the evidence is removed from the examiner. This reduces, if not eliminates, the tedious drive-swapping imaging process that is common in digital forensics practice.

Full examination of the evidence is accomplished by assigning the desired section of the DESL to a virtual OS server. The virtual OS server provides access to evidence stored within the DESL and the primary platform for evidence processing. The DEC creates a session on a virtual OS server, assigns permissions to evidence in the DSL, and configures the desired forensic programs and examination environment. The virtual OS server assigns this access in a write-protected mode, allowing traditional examination of digital evidence using forensic software. Alternatively, the virtual OS server may place evidence in a persistent mode, allowing examiners to view and handle evidence as if it were in the original imaged device. Once the examination is complete, access to the evidence is removed from the virtual OS server; this secures the evidence within the DESL, which models a physical evidence custodian and evidence locker.

At DESL locations, access to the virtual OS server is accomplished via secure firewalls and VPN connections over TCP/IP networks (Figure 2). The standard IP network infrastructure provides examiners and investigators from other locations with access to digital evidence and examination reports via broadband or even low-bandwidth modem connections. Digital evidence can be mirrored to other DESL sites to support data redundancy and parallel examinations.

Figure 3 shows a DEC designed to support electronic crimes investigations and digital evidence storage needs of the Oklahoma State Bureau of Investigation (OSBI). OSBI has three main sites (Tulsa, Oklahoma City and Weatherford), each of which could house a full-blown DESL, including a SAN, virtual OS server and digital forensic workstations. Each site would field two to five digital forensic examiners who would serve the entire state of Oklahoma. The three DESLs would be connected by dedicated high-speed Internet2 connections, allowing agents from the three main sites to collaborate on cases. For example, if Weatherford has a small caseload, agents in Weatherford could work on Tulsa cases

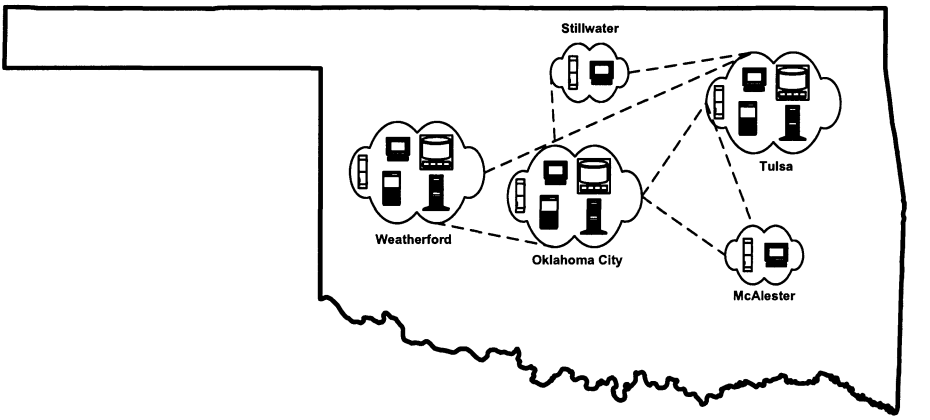

Figure 3. Digital evidence custodian supporting state-wide access.

without making the three-hour trip to Tulsa. Evidence in high priority cases could be processed at all three locations simultaneously. Furthermore, digital evidence could be mirrored at multiple sites to enhance efficiency and support redundancy and disaster recovery efforts.

OSBI agents at other locations in the state could also participate in digital forensic investigations. For example, agents in Stillwater and McAlester (Figure 3) could access the network of DESLs using smaller networks of forensic workstations and servers. Agents at these locations could perform imaging, examinations and report generation, significantly enhancing the overall productivity.

The OSBI DEC could support investigations throughout Oklahoma. For example, law enforcement agents from a rural community with limited expertise and technology could seize storage media, computers or portable electronic devices and send them to a DESL site or to a location with access to a DESL for processing. They could then access the results and investigative reports using standard Internet connectivity.

4. Conclusions

The digital evidence custodian (DEC) architecture is a powerful, yet relatively inexpensive, network-based solution for storing and handling massive quantities of digital evidence. In a typical implementation, evidence is stored in a network of digital evidence storage lockers (DESLs), which use NAS over SAN technology and dedicated high-speed Internet2 connections to facilitate the collaborative processing of digital evidence by examiners, investigators and case agents who may be at different locations. Use of standard IP network infrastructures enables other authorized individuals to access digital evidence and examination reports

maintained in the DESL network via broadband or even low-bandwidth modem connections. In addition to simplifying the tasks of storing evidence and maintaining its integrity, the DEC architecture significantly enhances the productivity of digital forensic investigations by supporting the distributed access and processing of digital evidence.

References

[1] M. Anderson, Hard disk drives – Bigger is not better, New Technologies (www.forensics-intl.com/art14.html), 2001.

[2] B. Carrier and E. Spafford, Getting physical with the digital investigation process, *International Journal of Digital Evidence*, vol. 2(2), 2003.

[3] T. Clark, *Designing Storage Area Networks: A Practical Reference for Implementing Fibre Channel and IP SANs (Second Edition)*, Addison-Wesley, Reading, Massachusetts 2003.

[4] Digital Preservation Coalition, Media and formats (www.dpconline.org/graphics/medfor/media.html).

[5] B. Goldworm, The difference between SANs and NAS, *Network World Storage in the Enterprise Newsletter* (www.networkworld.com/newsletters/stor/2000/1002stor1.html?nf), 2000.

[6] Institute for Security Technology Studies, *Law Enforcement Tools and Technologies for Investigating Cyber Attacks: Gap Analysis Report*, Dartmouth College, Hanover, New Hampshire, 2004.

[7] T. Maiorana, Building community support - The Heart of America RCFL, presented at the *Digital Forensics Working Group Meeting*, Orlando, Florida, March 2004.

[8] C. Mallery, Personal correspondence, North Texas Regional Computer Forensics Laboratory, Dallas, Texas, February 20, 2004.

[9] NAS-SAN.com, Technology overview (www.nas-san.com/differ.html), Zerowait Corporation, Newark, Delaware, 2003.

[10] B. Nelson, A. Phillips, F. Enfinger and C. Steuart, *Computer Forensics and Investigations*, Thompson Course Technology, Boston, Massachusetts, 2004.

[11] D. Orzech, Rapidly falling storage costs mean bigger databases, *CIO Update: Technology Trends* (www.cioupdate.com/trends/article.php/2217351), June 4, 2003.

[12] C. Poelker and P. Nikitin, *Storage Area Networks for Dummies*, Wiley, New York, 2003.

[13] W. Preston, *Using SANs and NAS*, O'Reilly, Sebastopol, California, 2002.

[14] RCFL National Program Office, *Regional Computer Forensics Laboratory Program: Fiscal Year 2004 Annual Report*, Federal Bureau of Investigation, Quantico, Virginia, 2005.

[15] M. Reith, C. Carr and G. Gunsch, An examination of digital forensic models, *International Journal of Digital Evidence*, vol. 1(3), 2002.

[16] K. Shanmugasundaram, A. Savant, H. Bronnimann and N. Memon, ForNet: A distributed forensics network, *Proceedings of the Second International Workshop on Mathematical Methods, Models and Architectures for Computer Networks Security, Lecture Notes in Computer Science, Volume 2776*, Springer-Verlag, Berlin-Heidelberg, Germany, pp. 1-16, 2003.

[17] M. Vatis, *Law Enforcement Tools and Technologies for Investigating Cyber Attacks: A National Need Assessment*, Institute for Security Technology Studies, Dartmouth College, Hanover, New Hampshire, 2002.

[18] R. Voss, Building a team – Chicago RCFL, presented at the *Digital Forensics Working Group Meeting*, Orlando, Florida, March 2004.