Chapter 28

A VIRTUAL DIGITAL FORENSICS LABORATORY

Philip Craiger, Paul Burke, Christopher Marberry and Mark Pollitt

Abstract This paper discusses the concept of a virtual digital forensic laboratory, which incorporates networked examination and storage machines, secure communications, multi-factor authentication, role-based access control, and case management and digital asset management systems. Laboratory activities such as the examination, storage and presentation of digital evidence can be geographically distributed and accessed over a network by users with the appropriate credentials. The advantages of such a facility include reduced costs through shared resources and the availability of advanced expertise for specialized cases.

Keywords: Virtual laboratory, virtualization, storage area network

1. Introduction

The collection, storage, examination and presentation of digital evidence typically occur in centralized laboratories. This is an inefficient model in that law enforcement agencies duplicate resources that are available elsewhere. A modest laboratory can cost tens of thousands of dollars, even more when the costs of computers, storage, forensic tools and training are considered.

The validation of forensic tools also poses problems. Proper forensic procedures require that the tools used in examinations be continually validated. Unfortunately, most examiners may not have the expertise to perform hardware and software validation. Additionally, there is a tremendous amount of duplication if every examiner has to validate the same tools.

Digital forensics laboratories of the future will be "virtual" in nature – they will not be limited by geographic boundaries. This paper proposes the concept of a virtual digital forensics laboratory (VDFL).

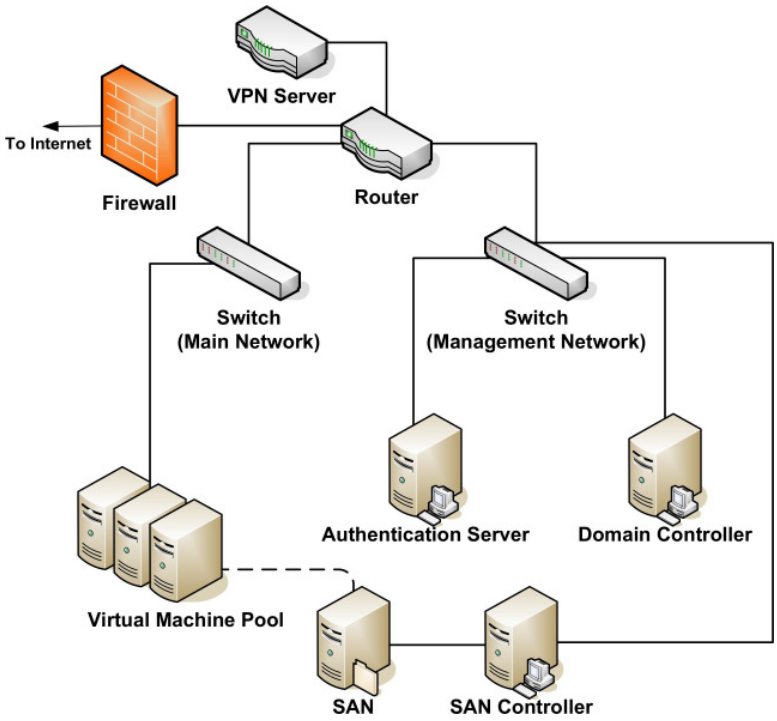

Figure 1. Virtual digital forensics laboratory.

The proposal builds on previous work (e.g., [1]) by decentralizing forensic functionality while providing security and quality management processes. The virtual laboratory incorporates storage area network (SAN) [7, 10] and virtualization [3, 4, 8, 9] technologies, forensic tools and security mechanisms to create a robust alternative to a physical laboratory. Such a laboratory will reduce the duplication of resources and tasks, provide law enforcement agents with cutting-edge tools, resources and expertise, and lower the cost of forensic examinations.

2. Virtual Laboratory Overview

The virtual digital forensics laboratory architecture attempts to leverage state-of-the-art networking and digital forensics technologies to support the acquisition, transportation, examination and storage of electronic evidence. This has resulted in somewhat unique problems that, in turn, require unique solutions.

Figure 1 presents a schematic diagram of the virtual laboratory. The laboratory provides facilities for examining, storing and presenting electronic evidence. The forensic examination component includes hardware, software and processes associated with the extraction, identifica-

tion and interpretation of evidence. The storage component incorporates large-scale, redundant magnetic storage that is logically separable by case. The presentation component includes extracted evidence, reports and other relevant information that may be accessed at various levels of detail by authorized parties (investigators, prosecutors, defense attorneys, judges and jury members).

The virtual laboratory employs a distributed model, which allows the complete functional separation of components. Each component is tied to the other components via a high-speed network. This approach is used very effectively in data centers, where physical and/or logical components are spread across multiple geographic locations.

The first step in defining system requirements is to identify the potential users. The user groups are law enforcement personnel, prosecution and defense attorneys, judges and jury members. Role-based access control allows each user to have the appropriate degree and type of access. For example, a law enforcement agent (examiner) might be granted full read/write access to each subsystem to perform imaging, upload images to long-term storage, conduct examinations of the evidence, and output intermediate results and final reports. Attorneys would not need access to the raw data or to examination tools, only to the intermediate results and final reports. Judges and jury members would only be able to view the final reports containing evidence prepared for trial.

Several other issues must be considered when specifying the system requirements. System performance is important in a distributed system because data has to travel over greater distances and on slower links than in a centralized facility. Security is equally important. Imaging, examination, storage and presentation involve machines in multiple locations, but the level of assurance provided must be just as high as that in a physical laboratory. Another issue is system management – every system component in every location must be maintained by an administrator who has the appropriate technical expertise and qualifications. The final issue is to create a system that is transparent to end users. Indeed, the system should look, feel and react as if it is a local examination computer.

System Performance A virtual system must provide the responsiveness and efficiency of a physical system. Even when high-speed networks are used, the transfer rates of forensic data to a remote location are much slower than the data transfer rates within an average computer, i.e., network speed is much slower than computer bus speed. Furthermore, in a multi-user environment, each data processing component must be capa-

ble of supporting multiple users reliably and with a satisfactory level of responsiveness.

Security Security is paramount for a distributed forensic system. The legal requirements for evidence handling must be met, and there should be explicit guarantees about the confidentiality, integrity and availability of all data. The system must incorporate a strong user authentication mechanism, ideally one that relies on multi-factor authentication. The system must logically separate users to ensure that forensic data can only be accessed by authorized individuals. A logging system must be in place to provide accountability for user actions and to track system use because of the sensitive nature of the data and the applicable legal requirements (e.g., evidence handling and chain of custody).

System Management System management is not a major concern for the typical forensics examiner. A distributed, multi-user system, however, requires a system administrator with demonstrable technical skills to create users, set access controls, maintain performance requirements, troubleshoot network problems and ensure system integrity.

Usability Ease-of-use is also important. A distributed architecture leads to increased complexity for users; this must be addressed by making the system appear cohesive and familiar to all types of users, and by providing adequate documentation and support to ease the transition to a virtual laboratory. Additionally, the system should facilitate information sharing and collaboration.

3. Virtual Laboratory Architecture

We have developed a prototype virtual laboratory that meets the requirements described above. However, it is only the laboratory users who are distributed; all the hardware and software components are currently situated in one geographic location. We chose to use a single location initially in order to evaluate the interactions between the hardware and software components, which would be more difficult to accomplish in a distributed environment. After the components are tested individually and as a complete system, the components will be distributed and the virtual laboratory will subsequently be tested.

Figure 2 presents the architecture of the virtual laboratory. The examination system uses a virtual machine pool with a variety of operating systems and complete user separation. SAN technology is used for local data storage because it can reliably store case data and is designed to support multiple users simultaneously. Authentication and security

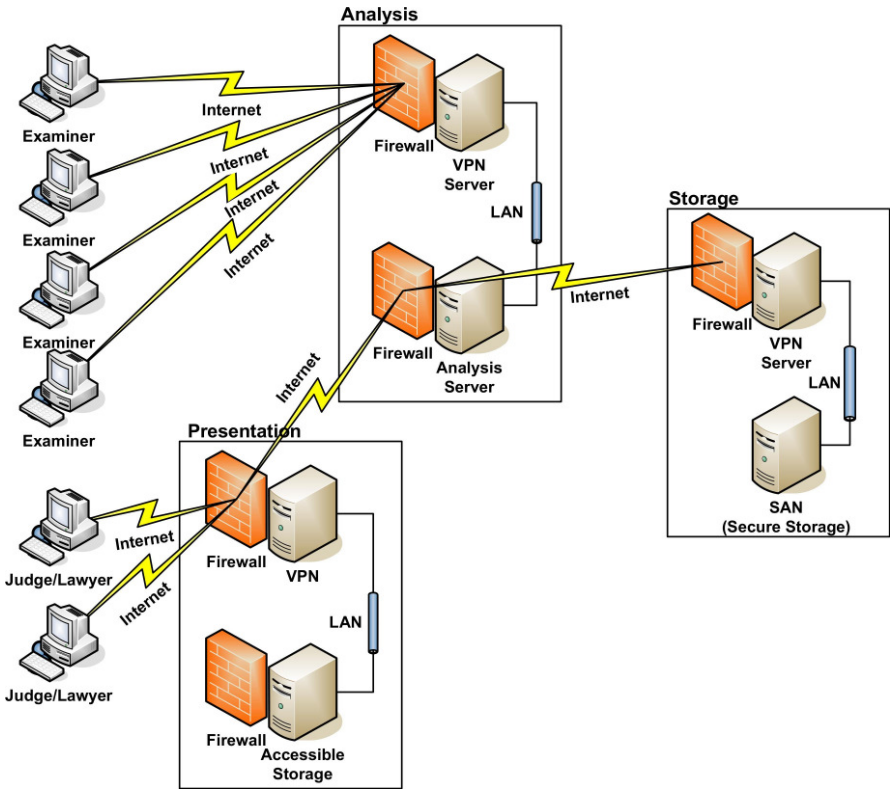

Figure 2. Virtual digital forensics laboratory architecture.

are implemented via a virtual private network (VPN) bound to a multi-factor authentication system. The next few sections discuss the design themes pertaining to the prototype laboratory.

3.1 Internal Network

A hardware-based firewall is used to control and filter inbound and outbound traffic. A gigabit Ethernet switch is used for internal network traffic. Multiple virtual LANs (VLANs) are employed on the switch to logically and securely segment management and user traffic. The switch also provides access control. Specifically, every port is configured to only accept connections from predefined hardware; this prevents a rogue user with physical access from connecting to the internal network.

3.2 Authentication and Access Control

The virtual laboratory uses multi-factor authentication and multiple access control techniques. Remote access to the network is controlled via

two methods. One is a hardware-based VPN appliance that uses two-factor authentication (username/password and hardware token). The other is IP-based authentication that allows access to users from a predetermined list of static IP addresses. Using these two methods in combination reduces the potential for system compromise [5].

Users connected to the internal network access their personalized workspace (e.g., user preferences, software and casework) using the same two-factor authentication mechanism. User authentication and access control are handled by a central management (LDAP) directory running on a non-virtualized system.

3.3 Virtualization

The VDFL has a pool of physical servers that host a number of virtual machines (Figure 2). These virtual machines are used by examiners to operate on the data stored in the SAN. Virtualization software was chosen instead of separate physical machines for several reasons. Virtualization allows for the logical separation of users. It reduces costs by consolidating hardware; also, deployment time is decreased because the virtual machine hardware is standardized. Moreover, virtualization allows administrators to move active virtual machines in the pool of host servers in real time to improve system performance and reliability.

3.4 Forensic Tools

Examiners using the virtual laboratory would have access to a variety of commercial and open source forensic tools running under virtual machines. Access would be available to multiple operating systems and related tools based on examination needs. Note that access control is role-based and determined by user credentials, not by the particular operating system or tool used. Costs would be reduced because forensic tools would be shared.

3.5 Storage

One of the principal issues is to provide an adequate storage pool for users. Another is to maintain satisfactory throughput for multiple simultaneous users. SAN technology was selected for its speed and reliability. Throughput is provided by a fibre channel connection running at 2 Gbps, which offers dedicated bandwidth for multiple users working with large data sets.

The SAN is partitioned into various logical storage roles to include virtual machine storage and case data (raw data, intermediate results and presentation data). Logical separation of storage allows access con-

trols to be applied through the management directory. Law enforcement agencies are often required to retain case evidence for decades. Providing long-term storage is a goal for the virtual laboratory. However, using the SAN for this purpose should be considered carefully along with the possibility of using other archiving technologies.

3.6 Internal Network Security

Security is an integral part of the virtual laboratory architecture because of the sensitivity of the data and the requirements imposed by federal and state laws. Every component must be secured using applicable procedures and best practices [6]. Also, extensive logs must be maintained of user access and resource utilization. The current design uses segmented logging channels (separate from those used by the main network traffic) to send data to a central logging server.

4. Challenges

Several challenges were encountered while designing and implementing the virtual laboratory. A major challenge was the inability of the current configuration to support the uploading of large data sets. This was largely due to the limited bandwidth available for consumer-level Internet connections. A promising solution is to use a high-speed network (e.g., Internet 2 Abilene/Florida LambdaRail [2]).

Another challenge is posed by popular forensic examination suites that rely on physical hardware locks (typically a USB key or "dongle"). Mapping and managing these dongles to individual virtual machines are problematic due to various physical and logical constraints (e.g., mapping identically-named devices to their virtual machines and coping with physical limitations of the available ports to connect these devices). A promising solution is to use special connectivity software that would allow a dongle to be remotely connected to a local virtual machine. This software would also enable users to use their own dongles and the virtual laboratory to host dongles.

Ultimately, however, the principal challenge is to address cultural and political barriers to the use of hardware and software resources located outside the local law enforcement agency jurisdiction. Furthermore, even if state-of-the-art technologies and best practices are employed to maintain the confidentiality, integrity and availability of digital evidence in the virtual laboratory environment, questions will persist. All the stakeholders – law enforcement agents, attorneys, judges and juries – will have to be educated about the benefits of the laboratory and the proper use of its facilities.

5. Conclusions

A virtual digital forensics laboratory is not limited by geographic boundaries – it decentralizes forensic functionality while providing security and quality management processes. As such, it would reduce the duplication of resources and tasks, provide law enforcement agents with cutting-edge tools, resources and expertise, and lower the cost of forensic examinations. Practically every crime now involves some aspect of digital evidence and the volume of digital evidence is growing faster than the ability of law enforcement to process it. Several challenges remain to be addressed before virtual digital forensics laboratories can become operational, let alone thrive. Nevertheless, these facilities may be the single best hope for law enforcement agencies to cope with the deluge of digital evidence.

Acknowledgements

This research was supported by the Electronic Crime Program of the National Institute of Justice under Contract No. 2005-MU-MU-KO44.

References

[1] M. Davis, G. Manes and S. Shenoi, A network-based architecture for storing digital evidence, in *Advances in Digital Forensics*, M. Pollitt and S. Shenoi (Eds.), Springer, Boston, Massachusetts, pp. 33–42, 2005.

[2] Florida LambdaRail, Florida's Research and Education Network (www.flrnet.org).

[3] International Business Machines, IBM Systems Virtualization (Version 2, Release 1), Armonk, New York, 2005.

[4] N. McAllister, Server virtualization, *InfoWorld*, February 12, 2007.

[5] V. Mukhin, Multi-factor authentication as a protection mechanism in computer networks, *Cybernetics and Systems Analysis*, vol. 35(5), pp. 832–835, 1999.

[6] National Security Agency, NSA Security Configuration Guides, Fort Meade, Maryland (www.nsa.gov/snac), 2005.

[7] B. Phillips, Have storage area networks come of age? *IEEE Computer*, vol. 31(7), pp. 10–12, 1998.

[8] A. Singh, An Introduction to Virtualization (www.kernelthread.com/publications/virtualization), 2004.

[9] M. Stockman, J. Nyland and W. Weed, Centrally-stored and delivered virtual machines in the networking/system administration lab, *ACM SIGITE Newsletter*, vol. 2(2), pp. 4–6, 2005.

[10] J. Tate, F. Lucchese and R. Moore, *Introduction to Storage Area Networks*, IBM Redbooks/Vervante, Rolling Hills Estates, California, 2006.