Virtual Laboratory Environments: Methodologies for Educating Cybersecurity Researchers

Kara Nance\textsuperscript{a}, Brian Hay\textsuperscript{a}, Ronald Dodge\textsuperscript{b}, Alex Seazzu\textsuperscript{c} and Steve Burd\textsuperscript{c}

\textsuperscript{a}Department of Computer Science, University of Alaska Fairbanks
\textsuperscript{b}Department of Electrical Engineering and Computer Science, U.S. Military Academy
\textsuperscript{c}Anderson School of Management, University of New Mexico

Abstract

In an ongoing effort to improve the educational experiences for their students and to open the door to research opportunities, many institutions are investigating innovative methodologies to provide hands-on learning and research environments. The application of virtualisation technologies to the study of computer security and other academic disciplines has had the most significant impact through the development of specialised laboratories utilising workstation or server based virtualisation. While these labs vary greatly in configuration and scope, they share a common purpose; to provide scalable infrastructure solutions to support cybersecurity research and education, training, and awareness. The innovation and variety associated with these labs are remarkable, with the additional challenges and opportunities of each deployment providing a rich foundation for future development and extension to a wider audience. The common component employed in each program is the extensive use of virtualisation. This paper investigates three unique example implementations of these environments which represent the continuum from local to remote access. Findings include the significant amount of resources required to initially create a virtual research environment, the administration requirements, and the advantages of leveraging the knowledge developed through several years of testing and application to reduce the resource requirements and financial investment.

Key words: Cybersecurity, security education, virtualisation, VMware, virtual laboratory

Introduction

Cybersecurity is an area of significant investment by both government and industry for over a decade. Recent government-led efforts have included:

advance the practice of cryptography, authentication, and security within desktop computing environments.

- Department of Defense Cyber Subcommand (U.S. Department of Defense, 2009) Created to ‘better coordinate the day-to-day defense, protection and operation of military computer networks.’
- Department of Homeland Security Cybersecurity Hiring Plan (U.S. Department of Home Security, 2009) Authorizes up to 1,000 positions over three years to fill critical cybersecurity positions including cyber risk and strategic analysis, cyber incident response, vulnerability detection and assessment, intelligence and investigation, and network and systems engineering.

The latter announcement was met with a tremendous amount of speculation about whether 1000 qualified applicants exist (Schneier, 2009; Cringely, 2009; Spafford, 2009).

To expand research and teaching efforts in cybersecurity, the U.S. government established the National Centers of Academic Excellence in IA Education and Research, the Department of Defense Information Assurance Scholarship Program, and the National Science Foundation’s Scholarship for Service Program 10 years ago. There are currently 96 centres of academic excellence, each of which has created or expanded degree programs and research efforts in cybersecurity to develop new knowledge and to increase the supply of cybersecurity researchers and professionals.

Educational institutions (as well as training organizations) are faced with the difficult challenge of providing infrastructure to meet the daunting needs and requirements for cyber training, education and research. An evolving trend in cyber education is the increasing reliance on hands-on educational components to facilitate learning and assessment (Hoffman et al. 2005; Conklin, 2006). As in many areas of study, cybersecurity research and education programs that engage students in practical application of theory and standards are considered highly effective (Hoffman et al. 2005; Conklin, 2006; Greenberg et al. 2003; Mountain, 2004). One of the most effective supporting mechanisms for providing hand-on exercises and supporting cybersecurity research is virtualisation technology.

Virtualisation enables a user to simulate an entire network of computers and their installed software on a single physical machine. The simulated computers, called virtual machines (VMs), are compartmentalized and function exactly as a system running on physical hardware. VMs can be configured to connect to one other over isolated virtual networks thus enabling users to experiment with a wide range of security configurations and tools while ensuring that production networks are unaffected.

In much the same manner that numerical simulation models can be applied to a variety of disciplines, virtualisation applications are not limited to cybersecurity. VMs and networks can also be used to support several other computer science fields. For example, VMs can be used to implement cluster algorithms without requiring multiple physical systems. VMs can also be used to research communication protocols, social networking, and computer-supported collaborative work. In all of these application areas, virtualisation enables researchers and students to construct, manipulate, modify, and delete test beds of networked computers and software without having to modify physical networks, hardware, and software. Much as numerical simulation has enabled rapid advances in areas such as biochemistry and aviation, virtualisation can speed knowledge creation and related educational activities in any field of study partly or solely concerned with computing hardware, software, and networks.

This paper investigates the virtual laboratories at three academic institutions that have been designated by the National Security Agency and the Department of Homeland Security as National Centers of Excellence in Information Assurance Education. These institutions are not the only security labs that are using virtualisation, but they represent a collection of distinct approaches which can be used as foundations for a comparative analysis with respect to research opportunities and challenges. After outlining the configurations of the
laboratory environments, we investigate the impact of their design and process decisions with respect to lab isolation before expanding to more general discussions about educational enrichment from both a student and instructor perspective and research implication and benefits of using virtual laboratories.

Virtualisation overview

Virtualisation was first widely employed by IBM in the 1960s to enable a single mainframe computer to host multiple operating systems and their users. The idea fell out of favour in 1970s as computers became less expensive and as single-user computers emerged. Virtualisation reappeared in the 1990s and found wide application in the 2000s as a way to consolidate many servers on fewer computers, thus reducing support costs and providing deployment scalability and flexibility. Virtualisation also found use in software development for testing new software on multiple operating system and hardware platforms. As virtualisation technology improved it found increasing application as a key component of cloud computing models and in educational settings to support distance learning and advanced study in computer science.

Under virtualisation, a single physical computer (the host) simulates the hardware of one or more other computers, the VMs. The host computer dedicates a portion of its own hardware resources to each host computer including processors, memory, disk storage, and input/output devices. The ‘disk’ of a VM is simply a large file on the host. When the VM is active an operating system and other software are installed on its ‘disk(s)’.

Other host files describe the configuration of the VM including its processor type(s), allocated memory, installed operating system, and connected input/output devices. The files that describe VM configuration and disk contents can be copied to create VM clones or moved across computer networks to redeploy VMs. With the right software, a VM can run a different system than the host computer and the host computer can simulate hardware devices that may or may not actually be installed on the host computer. Remotely-located users can interact with VMs via Web browsers and other software as if they were sitting at the keyboard and display of a physical machine. Complex software on the host supports communication with users provides needed hardware simulation, and acts as a ‘container’ for the VMs. Examples of such software include VMware Workstation, VM Infrastructure, and Microsoft Hyper-V.

Multiple VMs on a host can communicate with each through a virtual network and with the outside world via the host’s network interfaces. In education and testing environments, the VMs may be disconnected from the outside world to provide an isolated network for experimentation. This configuration is particularly valuable when the VMs and their virtual network are used to test ‘dangerous’ software such as computer viruses and worms.

Sample virtual laboratories

The following includes a brief description of each of the virtualized lab environments at the three selected institutions, the United States Military Academy at West Point (USMA), the University of New Mexico (UNM), and the University of Alaska Fairbanks (UAF). A more in-depth discussion of the laboratory configurations can be found in Nance et al (2009). These institutions began their collaborative work based on their common interest in virtual research environments and different approaches implemented to solve similar challenges.
USMA first used virtualisation to support cybersecurity education in 2001 (Schafer et al. 2001; Hoffman et al. 2003). The lab, the Information Warfare Analysis and Research Lab (IWAR), was initially created to incorporate multiple operating systems in the curriculum without supporting multiple systems or reboot systems during class. As virtualisation technology matured and systems became more powerful, the environment expanded to include multiple VMs for each student, interacting in an isolated network environment to provide direct experience with security technologies at a previously un-attainable level. The lab continued to evolve with virtualisation improvements, incorporating a robust backend network to serve as a ‘firing range’ for security tools. Students access the resources through local workstations managed in a dedicated environment, not connected to any production networks.

Originally deployed in 2004, UNM’s virtual laboratory (VLAB) was part of a strategic initiative at the Anderson School to expand computing capabilities and services beyond its physical labs. Currently the VLAB
contains 42 rack mounted personal computers connected to KVM switches, network switches, a storage area network (SAN), and uninterruptible power supplies. Users access virtual laboratory workstations via a Web-based interface on a small dedicated web server that displays each machine as an icon with availability status. When a user clicks on an available system, the web server initiates a Microsoft Remote Desktop (RDP) connection from the user’s computer directly to the chosen workstation. VMware workstation is installed on all VLAB machines and it is heavily used within cybersecurity and other courses for classroom exercises and student projects. The SAN stores VMware operating system images which are accessed via a mapped drive letter in the workstations. In addition to its computer security courses, the VLAB was adopted by other management disciplines to leverage its capabilities. These include remotely accessing statistical software in support of Marketing Research courses, configuring and managing database servers for Information Systems courses and delivering financial simulations for accounting courses.

The Advanced System Security Education, Research, and Training (ASSERT) Lab at UAF has been continually evolving since 2001 (Hay and Nance, 2006; Nance and Hay, 2008). Virtualisation was used throughout this process as a means to address the challenges that were faced, including the issues of physical space to host large numbers of students, and the ability to provide the same lab environment to distance education students that were available to local students. The current version of the lab utilizes VMware Infrastructure as the virtualisation suite, and includes four Dell 2950 8-core servers, supported by a Fiber Channel SAN on which the VM images are stored. The lab is managed using VMware Virtual Center running on a separate host.

Users make a connection to the Virtual Center server, either using a standard web browser or the Virtual Infrastructure Client software. Interaction with the VM has the same look and feel as a Remote Desktop or VNC session as shown in Figure 2. VMs in the ASSERT Lab are typically confined to one or more virtual networks, although physical networking components have been added to the lab in order to meet the needs of some users.

**Lab isolation**

The IWAR Lab at the U.S. Military Academy uses physical isolation to ensure separation from the campus network. The isolated network consists of student systems hosting a suite of VMs that can interact with each other as well as an enclave within the IWAR network hosting production style services on a variety of system and operating system architectures. Each student desk in the IWAR lab has an associated second computer that is connected to the campus network. This enables security students to have access to the Internet for research as well as allowing non-security courses to use the room. A KVM switch allows students to switch between machines on the campus network and the IWAR network.

UNM’s VLAB enforces lab isolation with firewall rules and VMware settings. VM network adapters are configured as ‘host-only’ isolating the traffic generated into a local virtual network. Firewall rules restrict incoming and outgoing traffic to/from the physical machines to hosts within the UNM network. This configuration enables students to upload and download homework notes and results to their UNM account.

Lab isolation in the UAF ASSERT Remote Access Lab is enforced in the hypervisor configuration, rather than relying on physical isolation of the systems. This approach was chosen at UAF to allow remote users to interact with their VMs, while not requiring that the VMs have any network pathway to the external environment (even through an intermediate host). All VMs in the ASSERT lab are confined to virtual networks (defined in the Virtual Center environment), or to physical networks that are air-gapped from any external network. Remote lab users make a connection to Virtual Center, which allows the user to interact with their VMs in a manner similar to Remote Desktop or VNC, with the difference that the hypervisor
facilitates the interaction with the VM, rather than using a network connection as would be the case for more traditional remote desktop like functionality.

**Educational experience – instructor perspective**

Before individuals can set up and use virtual environments, they need to become comfortable with virtualisation. To accomplish this, each lab has set up methods to train students and researchers in virtualisation.

The IWAR lab provides a rich suite of exercises that range from introductory concepts to detailed malware analysis and incident response. Courseware development has revolved around the use of both virtual only labs and labs incorporating the production enclaves in the IWAR. The United States Military Academy has two core IT courses (freshman and junior level) that service 500 students each semester. The use of the IWAR lab is not suitable for these students. Instead, a series of labs were created using three and four VMs to provide basic security awareness education. These labs can be run on student laptops or on classroom computers. Instructors from the two courses played an important role in the development of the content to ensure the direct support of the respective course requirements. The IWAR lab has also continued to grow to support additional courses. In 2005, the lab was used to support a new digital forensics course. Again, in 2006 the forensics course was expanded to incorporate network analysis.

The instructors using the IWAR are typically responsible for courseware development and enhancement of existing labs. They are all well versed in the functionality of VMs (using VMware workstation) and the integration into the IWAR environment. At UNM, each instructor designs how the virtual lab is used to support their courses. VLAB support available for courses ranges across a broad spectrum of intensity and technical complexity levels. At the simplest level, an instructor may use the VLAB because software needed for course assignments is installed on those workstations. At the other end of the spectrum, cybersecurity instructors customize VMware VMs to implement complex labs that incorporate isolated networks, multiple operating systems, and attack/defend scenarios. Instructors are trained in the use of the VLAB by technically-oriented faculty members as well as the department’s IT staff. Faculty members support one another in course design and VLAB applications. IT staff focus on client and network connectivity for end-users and also troubleshoot any hardware and software problems with the remote hosts.

In the ASSERT Lab, instructors and researchers are free to design and implement VMs and network environments appropriate to their projects, although this can be accomplished in many ways. In some cases this involves a request for the lab administrator to deploy VMs from the selection of prebuilt VM templates that already exist in the lab. As the needs and the requirements become more complex, the instructor or researcher tends to take an increasingly active role in the configuration. New VMs can be deployed and configured in the lab environment itself for users who have the appropriate permissions to do so. Individuals can also build VMs on their own systems (e.g., using VMware Workstation), then import their VMs into the lab. This can be a particularly useful approach when using a physical host as the basis for a research experiment, as a P2V (Physical to Virtual) tool can be used to create a VM from an existing physical host. In some cases a class may involve the deployment of entire networks of VMs by students. Students with appropriate permissions can create new VMs themselves, or alternatively can configure baseline images with the hardware and software necessary for their particular environment.

Once a scenario has been developed, a script is usually added to the lab environment to automate the deployment process. For example, if an individual creates an exercise environment consisting of four VMs and two virtual networks, the script could deploy that exercise, including creating VMs from the templates,
creating virtual networks, and assigning user permissions, for a list of users to read from a file. In this way a scenario can easily be deployed to an entire target group automatically. The majority of these scripts are written in Perl using the VMware VIX Perl API.

The ASSERT Lab supports researchers with a wide range of experience in virtual infrastructure use and development, including those who want to use the lab without learning about the implementation details. Assistance for individuals is provided by a faculty member and the IT administrator in the Computer Science Department. While the scripting of tasks and management of the underlying physical infrastructure is performed almost exclusively by these personnel, training and support is available to allow other researchers to perform a wide range of configuration and management tasks. In addition, the ASSERT Lab includes a wide selection of baseline VMs (known as templates), from which individuals can deploy VMs (and even create new, more specialized templates) as needed.

**Educational experience – student perspective**

In addition to the resources needed to aid instructors and researchers in the creation of educational experience, there is also a need to assess the educational experience from the student perspective. Lab utilization models can be clustered into three groups as defined by Savery (2005):

1. Blended learning where instruction is primarily face-to-face with lab experiences to supplement the classroom learning.
2. Hybrid format where the class meets face-to-face but the majority of the activities are online.
3. Online format where there is no face-to-face component and all learning takes place in the online environment.

Lab utilization at USMA, UAF, and UNM covers all three models with USMA primarily employing blended learning and UAF and UNM employing the hybrid format for some courses and an online format for others. Each lab utilization model presents its own opportunities and challenges with concomitant impact on the student learning experience, some of which are discussed below.

**Accessibility and availability**

Physical labs require that a student be geographically bound to the learning environment. In addition, many physical labs are staffed and only open at specific times. In contrast, remote access labs are generally available around the clock thus increasing availability. Accessibility is improved if the student can connect from any geographic location, as is the case with the UNM and UAF labs. In a 2007 survey of students using the UNM VLAB, 95% of respondents considered the VLAB a valuable addition to school computing resources and the same percentage were able to almost always or usually access the VLAB systems when needed. 63% of respondents considered the VLAB much more or somewhat more convenient than UNM's physical lab. Survey responses also showed that students accessed the lab from many locations (e.g., work, home, and while travelling) and at a wide range of times. These results are consistent with those reported by other researchers (Canfora et al. 2004; Cooper et al. 2004). The ASSERT Lab has generated similar positive comments about the associated flexibility and reduction in competition for lab seats. Usage statistics indicate the access is distributed around the clock with peak times in late evening when physical labs are closed. An unanticipated and interesting usage statistic indicates that a large number of students log in remotely from a common physical lab in the same building in order to interact with each and work together as they use the remote environment to work on homework assignments.
Providing technical support

Students using remote labs invariably encounter situations in which they need technical assistance. Situations may include routine questions about using installed applications as well as issues specific to remote labs including connectivity and performance. Physical labs usually have dedicated personnel present to deal with such situations. Providing similar service levels to remote users also requires access to support personnel. Staffing requirements can increase due to a greater number of ‘open hours’, additional technical complexity associated with remote access, and the inherent inefficiency remote vs. local ‘over-the-shoulder’ assistance. In the previously mentioned survey of UNM remote lab users, greater access to technical support was cited as concern by many of the students who provided written comments. The concern was greater among students studying non-technical subjects but was also expressed by some technically-oriented students. At the UAF ASSERT Lab many students tend to log in remotely from physical labs, conceivably to take advantage of the assistance available in the lab and proximity to instructors. However, the ASSERT Lab also offers the ability for multiple users to view and interact with any given VM simultaneously, thus allowing instructors and technical support staff to view the student’s environment as they would if they were physically located in the same lab.

Collaboration Support

In a traditional face-to-face class setting, the instructor and students meet at the same place at the same time. The dynamics of this sort of interaction vary greatly from the online classroom where the instructor and student meet at the same online place, but at different times. There are associated implications when students are isolated from their peers and their instructor as they frequently are in remote labs, which can be magnified when the lab is used as part of a hybrid or online learning experience. Bridging the social isolation issues cause new challenges for instructors in an e-learning environment. When an instructor teaches a face-to-face class or interacts with students in a physical lab environment, the nonverbal cues provided by students can provide the instructor with valuable information about the student’s competency in the class. When the student is not visible, the instructor faces new challenges. Lack of interaction with a student can mean that the student fully comprehends everything, that the student is hopelessly confused and doesn’t know what questions to ask, or there could be a myriad of other motivating factors. The method of teaching shifts from proactivity on the part of the student when issues are encountered to proactivity on the part of the instructor in order to identify students who are facing challenges.

Instructors teaching hybrid and online courses at UAF and UNM have addressed the need for remote student-to-faculty interaction in several ways. Some UNM instructors at UNM use email as the primary means of interaction, sometimes with specific guarantees for turnaround time. When combined with instructor access to student lab files, email provides an effective asynchronous method for supporting student learning. However, students are sometimes frustrated when they’ve allocated a block of time for a remote lab exercises but cannot use the entire block due to ‘hitting a wall’ that requires instructor assistance to bypass. Another asynchronous method that can partly address this issue is a frequently-asked-question (FAQ) web page or web-accessible message log. In either case, questions to the instructor and responses are made available to other students online, which reduce repeated questions and provides more timely answers.

Some UNM instructors, especially those teaching online courses, employ synchronous methods including virtual office hours via online chat rooms and instant messaging responses with specific time windows. These methods improve collaboration among students and instructors though at the cost of reduced schedule flexibility for both. However, for complex lab assignments requiring considerable student-instructor interaction, providing at least some synchronous interaction may be the only feasible way to achieve the associated learning objectives.
Another important issue to consider in remote lab design and utilization is support for collaboration among students. Learning effectiveness and efficiency can be enhanced when students interact with one another to exchange ideas and solve problems. Collaboration among students is facilitated in most traditional labs by physical proximity. Modern furnishing and layouts for physical computing labs enhance collaboration by clustering groups of workstations and furnishings.

Collaboration among students is a challenge when students use remote labs from isolated locations. As noted by Ma and Nickerson ‘[i]t may be that students using remote laboratories will find different ways of collaborating, and the mode of collaboration they choose may affect what they learn from the laboratory experience’ (Ma and Nickerson, 2006). Anecdotal reports by students to UNM faculty show that students use a variety of methods to communicate with one another while working remotely including email, chat rooms, instant messaging, and cell phones. None-the-less, collaboration support technologies could be incorporated into remote labs including methods for checking who else is working remotely combined with direct communication methods such as messaging windows and workstation-based audio- or video-conferencing. Though none of the remote labs described here currently provides these capabilities, they are being investigated as near-term enhancements to better support collaboration among students and instructors.

Faculty at Regis University in Denver have incorporated virtual world environments like Second Life and OpenSim into their remotely accessible teaching labs. For example, students working in a team can meet in Second Life for group meetings and demonstrations, and even for basic interaction with physical and virtual systems. Further development along these lines could conceivably provide students with the ability to manipulate virtual systems from the hardware to the software level in a manner analogous to that encountered in the physical world.

Student Learning Outcomes

The ultimate value measure for any educational technology is impact on student learning. Unfortunately, scientifically-valid evidence of that impact is difficult to obtain due to differences in lab technology, its application, and student characteristics. These difficulties are pronounced among the three labs and schools described herein making inter-institutional comparisons difficult. However, within each school, comparisons of learning outcomes before and after technology changes are possible. Also, the application of the technology to cybersecurity education and training standards in all three CAE schools provides some basis for comparison. Informal assessments by faculty and students at all three institutions provide overwhelmingly positive feedback about the impact of virtualisation and remote access labs on the learning experience. Virtualisation has enabled instructors to assign more complex and in-depth lab exercises which students routinely praise for their usefulness in tying theory to practice, while greatly reducing infrastructure requirements. The remote lab has increased student scheduling flexibility and reduced travel to/from campus, enabling students to apply more time to the more complex assignments. Rising employer satisfaction with new student hires is another indication of improved learning outcomes.

Research implications and benefits

The capabilities of virtual laboratories that support educational activities are also well-suited to supporting research activities. Specifically the environments described in the previous sections can provide researchers with:

- A remotely accessible environment to conduct experiments, analysis and produce reports.
An environment that can be isolated both in terms of protection for outsiders (consider programming code that could escape ‘in the wild’) and restrict public access to intellectual property and methodology during its development phase.

The ability to quickly deploy and redeploy computing resources without having to purchase, install or reconfigure hardware, software and networking.

Support for capturing snapshots or moments in time of complete computing research environments for playback repeated experimentation from a defined starting point using different parameters.

Using a virtual lab such as those described in this paper researchers can quickly deploy and use small- and even medium-scale experimental environments with a handful to hundreds of concurrent VMs. The results of this small- to medium-scale work can then be used as to design experiments to run effectively on large-scale test beds, such as GENI (2009) or the National Cyber Range (2009), if necessary.

An additional benefit of using remotely accessible virtualized environments, such as those at UNM and UAF, is that research teams can share isolated test beds without the requirement that the researchers in the same geographic location. Lab users can access lab resources, gather experimental results, and even share the same view of the environment from anywhere in the world, allowing teams to be comprised of those researchers most suited to working on a given project, rather than those who are simply located at a particular institution.

By using virtual lab environments, particularly those which are deployed as a network accessible resource rather than something the researchers must first construct, researchers can focus on the important research questions within their project in a collaborative manner, rather than the mechanics of constructing a suitable experimental environment.

Conclusions

These virtual labs are each the results of several years of effort, in which capabilities have evolved as new challenges and student demands have been identified. However, the effort needed to replicate any of these environments is significantly less, and as such similar labs could be deployed in a short time frame to meet similar needs at other institutions. The technical challenges are solvable with off-the-shelf technology though securing and isolating lab infrastructure requires significant and ongoing effort.

Each institution’s methodology presents a viable approach to the deployment of a virtual lab for other institutions based on their own mission, resources, and capabilities. As with any infrastructure decision, typically a solution that provides the greatest flexibility is preferred. However, in many cases the more flexible a solution is, the more resource, deployment, management, and maintenance challenges exist.

Virtualisation and remote access technologies, individually and in combination, are powerful tools for teaching cybersecurity and many other topics. Instructors can design and deploy complex exercises and assignments employing a variety of virtualized hardware platforms, operating systems, and networks. Students can manipulate these environments over days or weeks from many locations. Researchers or teams of researchers can conduct experiments using virtualized systems that can be rapidly deployed at a fraction of the cost of physical systems being virtualized. Virtual labs provide an opportunity to reinforce theory with practice that enhances the educational experience for students, faculty, and researchers, and opens the door for a new generation of virtualisation researchers to make significant progress in cybersecurity and other research areas.
References


