Emulytics™ at Sandia National Laboratories

V. Urias, B. Van Leeuwen, B. Wright, W. Stout
Sandia National Laboratories[1]
Albuquerque, New Mexico
{veuria, bpvanle,bjwrigh,wmstout}@sandia.gov

ABSTRACT

Testing, evaluation, training, and characterization of large, complex information systems is difficult, expensive, and time consuming. Sandia National Laboratories has pursued research over the last 10 years to develop tools and technologies to help address these issues. Enter Emulytics™.

The goals of this paper are to: (1) describe "Emulytics™" (emulation + analytics); (2) bring awareness to some of the current research efforts at Sandia and how they have been employed for a variety projects; and (3) share some lessons learned and success stories.

The discussion of Emulytics™ research begins by outlining the methodology for successfully creating models of large, complex systems using a variety of techniques. These systems include the blending of simulation, virtualization of hardware and software, emulation of devices, and direct deployment of actual hardware and software – that is, live-virtual-constructive (i.e., real-emulated-simulated) environments. Next, the discussion covers the building blocks used to create immersive, high-fidelity, system emulation environments to perform testing in areas such as: test and evaluation of security architectures and security devices; personnel training; course of action (COA) analysis; and application performance. Finally, we discuss some of the current research efforts, deployments and challenges.

ABOUT THE AUTHORS

Vincent Urias is a Principal Member of Technical Staff at Sandia National Laboratories, where he has spent the last twelve years conducting cyber security research and development. His research areas include cyber test-bedding, cyber modeling and simulation, as well as cyber analytics, cloud computing, and networking.

Brian Van Leeuwen is a Distinguished Technical Staff Member at Sandia National Laboratories. For 15 years he has been involved in the analysis of secure information systems using modeling and simulation methods. He received his Master of Science degree in Electrical Engineering in 1989 from Arizona State University, Tempe, Arizona.

Brian Wright has been a member of Sandia's Technical Staff since 2010, where his focus has been in cyber testbedding, cyber modeling and simulation, and analytics. He holds a Master's degree in Computer Engineering from the University of Illinois – Urbana/Champaign.

William Stout is a Senior Member of Technical Staff at Sandia. His research interests include emulation platforms, network virtualization, software-defined networking, and cyber-security systems design and assessment. He holds a Master's in degree in Computer Engineering from the Air Force Institute of Technology.

[1] *Sandia is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the United States Department of Energy under contract DE-AC04-94AL85000.*

Emulytics at Sandia National Laboratories

B. Van Leeuwen, V. Urias, W. Stout, B. Wright
Sandia National Laboratories
Albuquerque, New Mexico
{bpvanle,veuria,bjwrigh,wmstout}@sandia.gov

INTRODUCTION

Networked information systems play a key in role supporting critical government, including military and private information systems. Many of today's systems have strong dependencies to secure information exchange among geographically dispersed systems. As the systems become increasingly dependent on the information exchange they also become targets for exploitation. Operators of the information systems recognize the need to secure these systems but securing the systems becomes an increasingly daunting task. Securing these information systems is not only creating secure system architectures and secure system configurations but also heavily relies on well trained defenders of the systems. Thus there is a need to for flexible cyber security training, testing, and analysis platforms that can replicate information systems with high levels of realism to enable training and analysis.

Currently cyber defender training and system analysis is performed either on operational systems, some limited testbed, or on simulated models of the system of interest. Analysis and training on operational systems is limited to the most benign levels since any disruption to the operational has potentially severe consequences. Testbeds for analysis and training are typically expensive and time-consuming to construct and deploy, are single-purpose, and difficult to maintain. Testbeds are typically limited to small subsets of the system of interest and thus limited in the level of realism when compared to the operational system. Another option is the use of modeling and simulation for analysis and training. In many cases, the modeling and simulation program code needs to be developed to simulate the system and devices in question or extensions need to be made in order to answer specific questions. These (sometimes buggy) simulation codes typically do not depict an accurate picture of the system. To increase simulation result accuracy, models have to be extended and validated. This process can be time consuming and inefficient [2,3]. Thus there is a need for the ability to rapidly create high-fidelity replications of information systems for cyber training and analysis. Sandia National Laboratories has undergone multi-year research projects in the development of new strategies and methodologies that enable researchers to quickly and accurately model information systems hosts and networks of interest for cyber analysis and training.

METHODOLOGY

The research and development described in this paper set out to create a capability to effectively instantiate representations of networked information systems that enable cyber analysis and cyber training with high-levels of fidelity and realism. The capability includes the development of a flexible analysis platform that can replicate operational networks and the environment in which it operates. The capability will enable cyber analysis and cyber training. More specifically, the capability will enable understanding and planning of cyber operations, evaluate the effectiveness of deployed defense strategies and technologies, and determine effectiveness against expected cyber-attack approaches. The methodology of developing the cyber analysis and training capability includes asking numerous questions, such as:

- Can data obtained from real-life cyber incidents be leveraged in the cyber analysis capability and platform to create more-realistic and real-time training scenarios?
- How capable is the platform in configuration and deployment of new cyber experiments? How quickly can experiments be designed and implemented (i.e., machine speed vs. human speed)?
- How faithful is the capability and platform in representing and evaluating cyber security technologies?
- What is the process for effective training and equipping of the cyber analysts with new approaches, tactics, techniques, and solutions?

- Does the capability include methods or algorithms for scoring and measuring the effectiveness of the approaches, tactics, techniques, and solutions under evaluation?
- What is the scalability of the system-under-study through deployments on the platform? Can the capability and platform replicate systems at desired scales?
- Can multiple information system applications be deployed and have faithful interoperability with other systems and applications? Will the capability and platform accurately represent the operation of mission critical applications and the impacts to it from the approaches, tactics, techniques, and solutions under evaluation?
- Can technology and device specific cyber training and testing be performed? Consider IPv6 and wireless communications? Are mobile communications faithfully represented in the capability and platform?
- In cyber training scenarios, can the defender's actions be observed, assessed, and replayed?
- Will the cyber analysis capability and platform enable analysts and commanders in understand and quantifying the effects of their decisions in executing a plan?

The research and development team determined that an emulation and analytics capability developed at Sandia National Laboratories (Sandia) was directly applicable to meeting the cyber training and analysis objectives. Sandia has an art and science of modeling, simulating, emulating, instrumenting, and analyzing large-scale networks of engineered and human-coupled subsystems that have significant dependencies on cyberspace capability it named Emulytics™ (Armstrong and Rinaldi, 2010). Emulytics™ is derived from the following terms - emulative network computing + analytics. Emulytics™ includes the following objectives:

- Large-scale, vastly heterogeneous networked systems,
- Integrated systems that can be configured and used both for controlled experimentation and interactive exploration of system behavior,
- Components may be real, emulated, or simulated,
- Includes network(s) creation, management, and instrumentation,
- Includes large HPC platform management and monitoring,
- Includes data extraction and warehousing, and
- Includes analysis and result visualization.

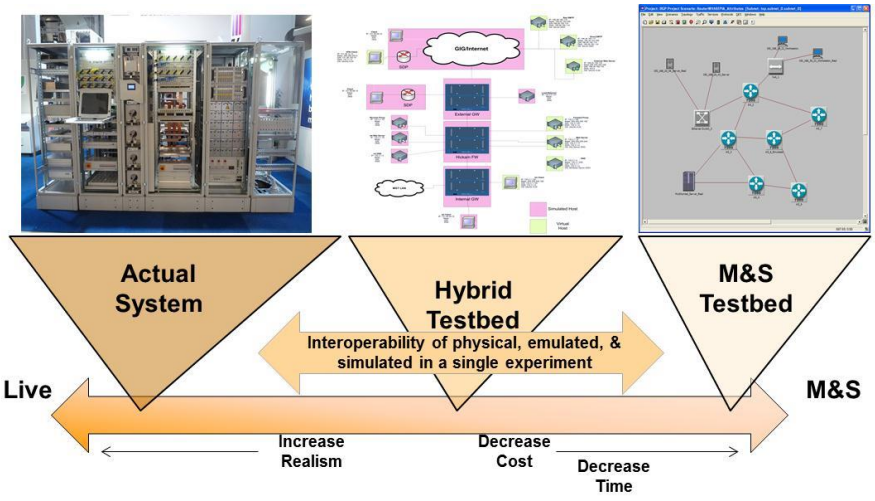

Figure 1. Emulytics™ platform employs live, emulated, and simulated models.

Emulytics™ provides the capability to combine real, emulated, and simulated devices, i.e., live, virtual and constructive (Van Leeuwen, et al, 2010), into a single system-level experiment to enable system-wide cyber training, while answering a wide variety of cyber-related questions. Unique capabilities of Emulytics™ include:

- Mechanisms to rapidly specify and deploy complex networked information systems of routers, switches, hosts, services, and applications.

- Extensive protocol support for network devices such as switches and routers.
- Instantiate 10,000's hosts, such as servers or workstations, in high-fidelity. Currently supports Windows and Linux operating systems but can be extended to support a greater variety of systems and devices including VoIP phones, printers, etc.
- Instrumentation at the hosts and network layer to capture, in high-fidelity, data describing system operation.
- Create complex scenarios (e.g., of deployments, intrusion attempts, user impact, etc.) that can be scripted for execution within the experimental platform.
- Incorporate application-layer overlay systems such as those used for Supervisory Control and Data Acquisition (SCADA) (Urias, et al, 2012).
- Represent mobile communications and their interoperability with fixed-networked systems.
- Represent the latest and upcoming security approaches such as Moving Target Defenses (MTD).

In general, an Emulytics™ solution includes obtaining operational system devices and configurations (router, switches, firewalls, security appliances, etc.) and deploying networked endpoints (e.g., Windows, Linux hosts or servers) that represent their operational system. An Emulytics™ platform will include instrumentation, data collection, and analysis backend capability that can digest the unstructured data produced by network devices, applications, hosts, and network defense tools to enable key aspects for training. The platform is capable of adequately representing the operational system so that cyber red teams and blue teams can exercise their techniques and develop tools, tactics and procedures.

The remainder of this research paper will describe the employment of Sandia's Emulytics™ to create a cyber-training and analysis environment. An example scenario is presented that describes the level of device and system fidelity realism that can be achieved with Emulytics™.

MODELED SYSTEM DEPLOYED IN EMULYTICS TESTBED EXPERIMENT

This section outlines how a system model is deployed using Emulytics™. The details are presented as part of an experiment description used for a cyber-training activity. The Emulytics™ methodology incorporates system modeling using emulation, physical hardware, and extensive virtualization. The modeling capability proves effective at incorporating necessary levels of realism for analysis. The system-level modeling includes distributed, replicated subsystems to create experiments of increased scale while maintaining high-levels of realism. The modeling capability includes instantiation of real applications and services running on virtualized hardware producing actual system transactions and network traffic. The experiment includes instrumentation, and data analytics. The system-level model incorporates many servers and workstations hosting actual applications and network services. Connectivity is provided by various types of network devices comprising LANs and WANs.

The example networked information system created with Emulytics™ includes a global Internet-like network with multiple cities having cyber cafes, an enterprise system employing a DMZ between it and the Internet, and a SCADA system managed by the enterprise system. Details of each sub-system are described below.

Global Internet-like System

The example network system includes a representation for global connectivity. The global Internet-like network includes Internet service provider (ISP) router representations in cities located around the world as shown in Figure 2. The ISP routers are configured as autonomous system and peer using BGP protocol. Each city's ISP router is connected to a distribution-like network comprised of routers using OSPF and has connectivity to business representations or cyber café representations. In our example scenario multiple cyber café locations include hosts that cyber red teams can use for their reconnaissance and exploit launching points.

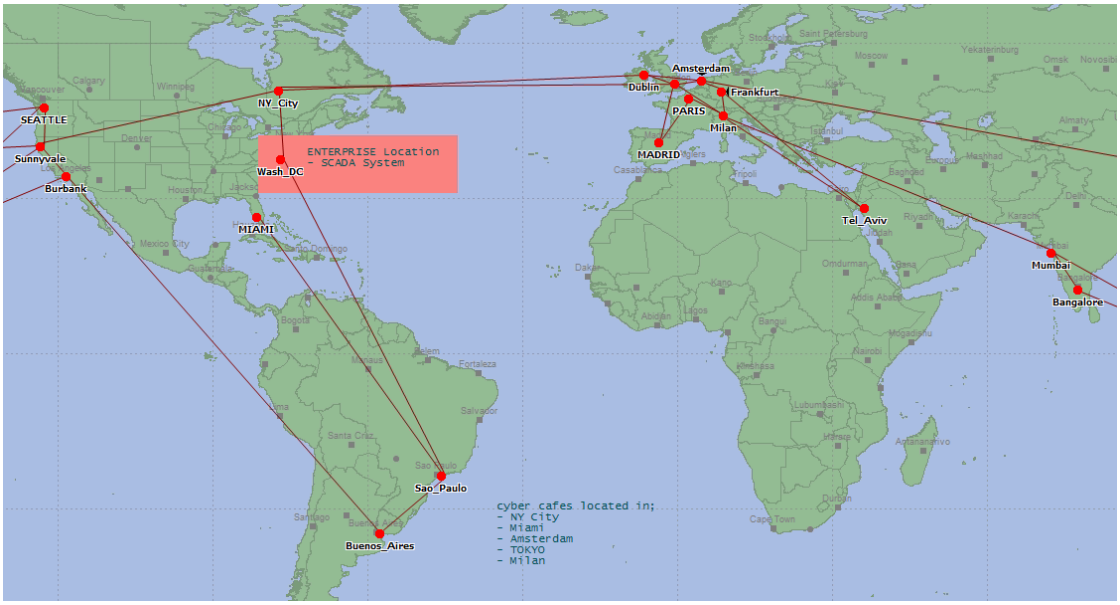

Figure 2. Global Internet and Cyber Cafés

Enterprise Networked Information Systems

In our example system, an enterprise network is represented. The enterprise network representation is located in Washington D.C. and is connected to the Washington D.C. ISP router via an intermediate router as shown in Figure 3. Connectivity between the enterprise network and Internet-like network is through a firewall and the network includes a demilitarized zone (DMZ) that includes security services and other network gateway functions such as mail servers, dynamic name service (DNS), and proxies. Additionally, several enterprise areas are included each having approximately 100 end points representing actual workstations, servers, printers, etc. In the Emulytics™ platform these end points are actual Windows or Linux hosts each configured with unique IP addresses and hosting specific applications. These end points will respond to network reconnaissance and mapping tools such as NMap. The response from the modeled system is similar to the response expected from a live operational system.

In our example training system, the enterprise network is also connected to a supervisory control and data acquisition (SCADA) system. The SCADA system could be, for example, an industrial assembly system or possibly a power distribution system. The enterprise network connectivity to the SCADA system network is via a firewall and includes an enterprise/SCADA interface subnet. In this architecture, the SCADA network can be accessed from the enterprise network through the SCADA/enterprise interface subnet.

For our example training scenario, the enterprise DMZ includes several hosts that incorporate vulnerable operating systems (OS) that can be used by red teams in their training exercises. These vulnerable hosts are included to provide pivot points for the red teams. The pivot points are necessary since a training exercise is limited in duration and for the red team to make progress in the allotted time, pivot points can be used to make progress towards some objective.

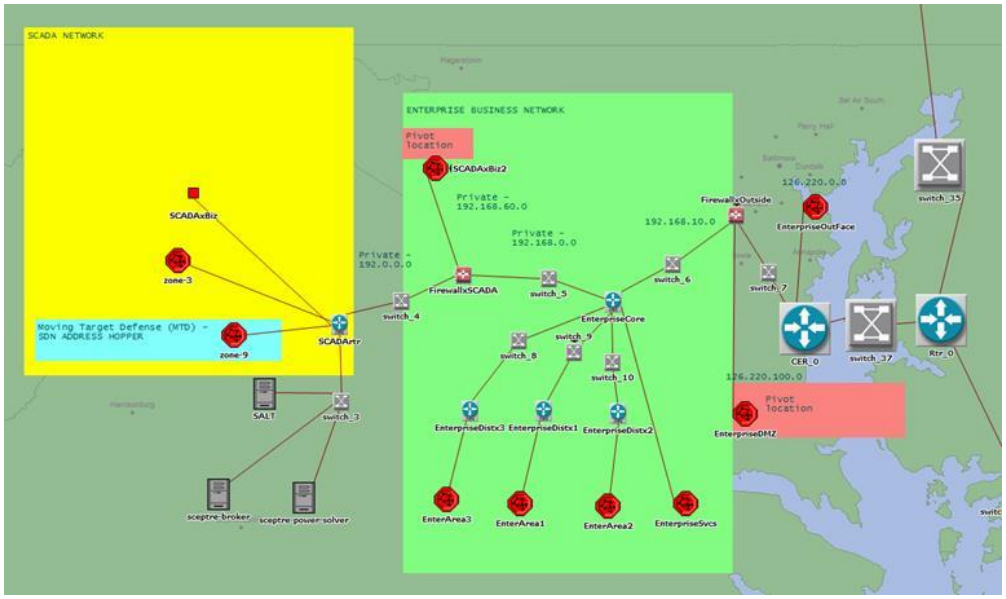

Figure 3. Enterprise Networked Information System

SCADA System

Many critical infrastructure systems rely on complex information systems for control and management. In the case of electric power, the critical infrastructure includes the physical systems; comprised of power generation, transmission and distribution capabilities. The control of the physical systems is accomplished via SCADA systems. Today's SCADA systems employ many of the same information system devices as traditional business or enterprise information systems. SCADA system networks, just as enterprise information system networks, are connected to external networks, including the Internet. In the example system the SCADA network is connected to the enterprise network through a firewall and includes a SCADA/enterprise system interface as shown in Figure 3.

In Figure 3, the devices and subnets in the yellow colored area make up the SCADA system. The SCADA system is segmented into three areas – a SCADA business area and two SCADA zones. The business area includes servers that support operations within the SCADA system area. This includes systems such power trading tools and broader system management tools. The two zones in the SCADA system segment include a group of remote terminal units (RTUs) that interface to physical equipment and report their state. Also in the zones, are front end processor (FEPs) that communicate with the RTUs and other SCADA resources as shown in Figure 4. Also note the various other SCADA system compute platforms represented in each area such as the two human machine interface (HMI) clients and servers, and historian.

An additional feature employed in the SCADA region, area-9 subnet, is a moving target defense (MTD) system being developed at Sandia. The MTD system is based on a software defined networking (SDN) approach that uses an IP address randomization approach (Chavez, et al, 2015). Further details of the MTD approach will be published in an upcoming research paper. The MTD approach is included in a single SCADA area with the objective of observing what a red team is able to discover and exploit in their training exercise. During the training exercise data is collected from the areas with and without the MTD approach. The results are used as additional information in our evaluation of the efficacy of the MTD approach.

To communicate with SCADA system nodes, a pivot point must be established within the neighboring enterprise system. More specifically, unauthorized attempts to gain access to the SCADA subnets can be launched from a node in the SCADA/enterprise subnet connected to the firewall separating the SCADA system from the enterprise system. The SCADA/enterprise subnet is shown in Figure 3. As with the DMZ located at the enterprise/Internet connection the SCADA/enterprise subnet has several vulnerable hosts that can used as pivot points for the red team activity.

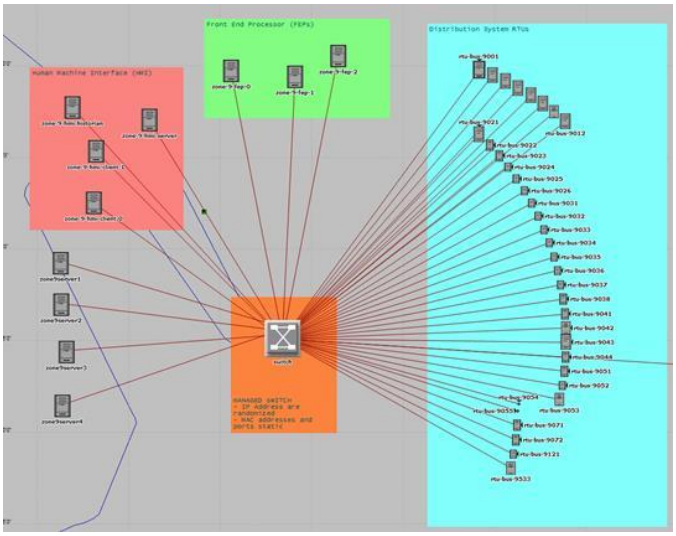

Figure 4. Supervisory Control and Data Acquisition (SCADA) Network Topology

MODELS AND EMULATIONS USED IN EMULYTICS EXPERIMENT

The system models deployed in the Emulytics™ testbed provide the necessary infrastructure to host experimental environments. As mentioned, these environments may consist of any number of subsystems meant to emulate common computer and communication networks. However, the underlying components of these systems are the primitives required to build out virtualizations that promote realism and fidelity. These primitives are primarily steeped in emulated machinery, from networking devices to application servers. The Emulytics™ emulation platform is also versatile enough to incorporate hardware-in-the-loop (HITL) as required. Between emulation and HITL, the end state is to provide environments to address such questions as:

- Do the local and wide area networks respond appropriately? Do routes and paths converge as expected? Are quality-of-service (QoS) parameters and metrics comparable to those in the real world?

- Do devices perform as expected? Are servers, SCADA devices, and security stack devices well integrated? Do the devices offer the same, if not extended capabilities to monitor and introspect upon?

- Do users, such as red or blue teams, feel comfortable in the environment? Do workstations, servers and applications accurately reflect the settings they're accustomed to?

Device Representations

The emulated devices used in the experiments are comprised of a swath of varying technologies, often packaged in the form of virtual machines. The emulation platform itself allows the instantiation of these virtual machines by 'snapshot' (wherein the same virtual machine image may be used for high density experiments), or by 'write-back' as required (where changes made in the virtual machine are written back to the virtual disk image). However, those virtual machines running in snapshot may still be uniquely configured through the use of techniques like virtual disk image file insertion, SNMP, DHCP and device-specific in-band configuration methods. The usual suspects for virtual machines include Windows and Linux operating systems to represent workstations, servers as well as endpoint devices in SCADA networks. To represent network infrastructure, virtual routers, Layer-2 and Layer-3 switches are also instantiated within the experiments. These latter devices not only provide the means to apply QoS and promote network realism – but also the ability to monitor and assess the experiments from the networking perspective.

Application and Traffic Representations

In order to establish a realistic and high fidelity model, applications and traffic generation were added to the primitives of the topology deployed. Through software stubs and scripting, configurations injected at run-time into the virtual machines install, configure, and start applications to provide the look and feel of an enterprise network within the model.

A minimum enterprise network is often comprised of a domain controller and e-mail server. The model is expanded by an array of server-based services such as instant messenger, collaborative wiki, cyber defense tools, and general web servers. To facilitate training environments, exploitable targets are added to the topology to hunt, providing pivot points, as well as provide remediation for multi-day events. These images are typically unpatched versions of Windows server and desktop, but also include some breeds of Linux, with known vulnerabilities, which are often easily identified by tools such as Metasploit.

Presence of the right objects in the network create some realism, but hardly show the fidelity required by the demands of most Emulytics™ use-cases. Traffic generation on the wire between endpoints is added to address this gap in fidelity, simply from the emulation environment. A small cross-platform binary is used to generate HTTP(S), SMTP/TLS, and SSH traffic over both IPv4 and IPv6 links.

DATA COLLECTION AND ANALYTICS IN EMULYTICS EXPERIMENTS

At the crux of any experiment is the ability to extract information about the experiment itself. Even more so, extracting data that is meaningful, concise and actionable. Naturally, the types and quantity of data pulled from the experiment should be based on goals of the experiment. It is often the case that experiment data outputs have not matched the user requirements, and have resulted in lost time and efforts for both sides. In training environments, this chasm is extremely exacerbating when experiment outputs are required for feedback to the trainees, and further development of pertinent training environments.

Thus, the emulation environment must be flexible enough to employ devices that are highly configurable with respect to data output. Furthermore, the emulation environment itself should exhibit this characteristic as well. Data extraction and collection must also pay mind to formatting, to ease the parsing and ingestion requirements for analytic applications. As mentioned, the virtual network devices deployed in the environment provide the ability for network monitoring applications to poll SNMP data (e.g., performance metrics, routes, CAM-table entries). Virtual machine instantiations include agents to query and push host data to collection servers in- and out-of band. The emulation platform itself also includes the capability to: (1) introspect on virtual machines from the hypervisor; (2) capture point-and-click type operations from user VNC sessions; (3) collect summary network traffic and even full-packet capture on the physical host machine virtual switches.

Network monitoring applications are tooled to ingest active and passive network data to generate general and customized reports. This data may also be fed to the analytic engines that receive VM host data via VM agents, hypervisor-based introspection, and in-experiment virtual machine services (e.g., firewalls, IPS, etc.). Collectively, the fusion of the many data sources forms a rich, complex view into the system throughout the course of the experiment. This output may be coarse in nature for high-level discourse or with fine-granularity for detailed analysis.

CYBER TRAINING PLATFORM DEMONSTRATION DEPLOYED IN EMULYTICS

To exercise the Emulytics methodology, emulated systems and training objectives aforementioned, Sandia partnered with an external customer to craft a customized training environment. The customer's primary object was to train teams responsible for network defense and hunt. The teams would have to understand their roles, tasks, and how their actions supported (or impacted) overall mission. However, their current resources lacked the capability to develop diverse scenarios, to stand up experiments quickly and repeatably, introspect into actions and processes, and

to adequately gather data to reason about the experiment itself. Naturally, this latter shortcoming is detrimental to goals of implementing policy, developing TTPs, assessing/analyzing information, and reporting functions.

Three training zones were envisioned to support the training experiment. Two zones (A and B) were located at Sandia, the third zone (C) at the customer's location. The teams responsible for play in the environment, too, were planted at the customer site. Zone A consisted of a DMZ security stack comprised of routers, firewalls, IDS, enterprise services (DNS, mail, web). HITL elements were also included in Zone A, misconfigured with security vulnerabilities to facilitate hunt endeavors, and exercise exploitation discovery and response. Zone B consisted of the virtual systems mentioned above, specifically the Global Internet, SCADA, and Enterprise networks. The three systems in Zone B were stitched together as one seamless virtual environment to exercise the teams' skills at reconnaissance, exploitation and pivoting.

Sandia and the customer site were connected via a VPN. Using Layer2 tunneling, team workstations from the customer site were brought directly into the Global Internet system, occupying virtual workstations in the Tokyo domain. Once presence was established team members were free to roam the environments and put their training to work. Using the data collection techniques deployed in the environment, Sandia was able to capture their movements through the virtual networks. Collection not only revealed their actions from the networking perspective, but also on host – as granular as the points, clicks and keyboard entries on the virtual machines.

CONCLUSION AND FURTHER STUDY

In this paper we have provided a description of how Sandia's Emulytics™ methodology can be used as a cyber-training and cyber analysis environment. The description includes how information systems are deployed using Emulytics™ and how the system models are used for training purposes. Devices that comprise the replicated system can be either real devices, emulated devices, or simulated devices. An example scenario is included that includes multiple levels of network connectivity – from attackers located in distant cities attempting to exploit nodes in an enterprise system that also includes a protected SCADA system. The example illustrated the positioning of potential pivot points in the networks to enable red teams of different skill levels and exercise duration to have opportunity to exploit under the differing vantage points. The example training scenario is instrumented to collect data during the exercises to enable evaluation of red team performance in a quantitative manner.

A major challenge in modeling systems is the validation and verification of the model. Our approach currently employs limited validation approaches in that models that have been validated in other scenarios of similar use. Initial system-level results from multiple device interoperations are compared to those results in operational systems. Any discrepancies in the comparison results are assessed and either identified as acceptable deviation or the deficiency in the model is corrected. Additional research is being performed in this area.

This paper's primary focus is to describe one of the numerous ways Emulytics™ can be applied to challenging problems. Key to cyber training exercises is realism in system and environment. The closer a training platform can replicate an operational system the better the training and evaluation of the subject performance will be. The research team continues to develop additional capability in this area to further cyber training and cyber analysis of systems. In this research, we have developed an important and capable cyber security analysis and experiment environment to help perform analysis of communication networks and networked information systems. In addition to the example scenario described in this research paper, Emulytics™ capability has been applied to answer system level questions pertaining to:

- Evaluate security architectures of systems.
- Provide an immersive environment for red teams to assess different security models and their security risks.
- Used a blue-team training tool for operators to learn to configure components of the system.
- A red team environment to provide targets for training and evaluation of other systems.
- Cyber range environments for emulation of blue and red team activities.
- An effects-based modeling environment to test computer network defense strategies under a variety of conditions.
- A cloud computing testbed to learn and ask questions about open-source cloud solutions and applications.

- Data collection techniques to provide rich views and analysis of experiment outcomes.

REFERENCES

Armstrong, R., and Rinaldi, S. (2010). "Emulytics: concepts for cyber emulation, modeling, and simulation," Sandia National Laboratories Report – SAND2010-1639C.

Chavez, A., Hamlet, J., Lee, E., Martin, M., and Stout, W. (2015) "Network Randomization and Dynamic Defense for Critical Infrastructure Systems," Sandia National Laboratories Report – SAND20XX-XXXX (Pending).

Urias, V., Van Leeuwen, B. and Richardson, B. (2012), "Supervisory Command and Data Acquisition (SCADA) system cyber security analysis using a live, virtual, and constructive (LVC) testbed," MILITARY COMMUNICATIONS CONFERENCE, 2012 - MILCOM 2012, vol., no., pp.1,8, Oct. 29 2012-Nov. 1 2012.

Van Leeuwen, B., Urias, V., Eldridge, J., Villamarin, C., and Olsberg, R. (2010). "Performing cyber security analysis using a live, virtual, and constructive (LVC) testbed," MILITARY COMMUNICATIONS CONFERENCE, 2010 - MILCOM 2010 , vol., no., pp.1806,1811, Oct. 31 2010-Nov. 3 2010.