SAND2018-11095C

Cyber Range Infrastructure Limitations and Needs of Tomorrow: A Position Paper

Vincent E. Urias, William M.S. Stout, Brian Van Leeuwen, Han Lin
Sandia National Laboratories
Albuquerque, New Mexico, USA

***Abstract*—Cyber networks are extremely non-deterministic, complex systems. To address this, we must develop foundational research protocols to enable reproducible cyber experiments that can systematically uncover deep understanding of a cyber system's security posture. One core tenant of this approach is to have a test environment to enable a space to create and test hypotheses about systems and reason about results. To date, this has generally been done through cyber test beds or cyber ranges. National infrastructure supported by various government agencies have all created multi-million dollar ranges, and support other national infrastructure such as the National Cyber Range (NCR) and the Regional Service Delivery Points (RSDP). The thrust of this paper was based on multi-year studies, the culmination of which uncovered gaps and challenges associated with using various national infrastructures to represent a multitude of complex heterogeneous systems. Ranges, such as the NCR, have experiment life-cycle processes to take a cyber experiment from inception to analysis. However, our position is that processes used are not sufficient to address gaps and challenges. In this position paper, we review current range experiment methodologies and our observations of other considerations that should require inclusion.**

***Index Terms*—cyber range, testing, experiment, cyber security**

I. Introduction

Cyber security cannot be classified under a formal branch of science, such as physics or chemistry. Cyber security is essentially informed by the mathematical constructs of computer science, to include automata, complexity, and mathematical logic. However, unlike physics, cyber security depends on implementation correctness at the hands of developers and users, whose minuscule errors may result in disproportionate impacts on the security of the system itself. These challenges cause significant non-determinism in the system under study. To address this, practitioners must create foundational research protocols to enable reproducible cyber experiments that can systematically uncover deep understanding of a system's security posture. One core tenant of this approach is to have a test environment that enables a space to hypothesize about systems, execute experiments and observe or reason about the results. To date this has been done through numerous cyber test beds or cyber ranges. E.g., national infrastructure supported by the DoD (US Cyber Command, Department of Test and Evaluation, Test Resource Management Center) have all created multi-million dollar ranges and support other national infrastructure such as the National Cyber Range (NCR) and the Regional Service Delivery Points (RSDP).

The impetus of this paper was based on multi-year studies, the culmination of which uncovered gaps, challenges and questions of feasibility for using various national infrastructures to represent a multitude of complex heterogeneous systems. How do we build upon the existing cyber ranges construct to effectively address the significant increase in demand for cyber testing and training and provide the needed operationally relevant and technically representative range environments? And how do we close these gaps to not only support the DoD, but also other organizations in academia and industry who are either creating or beginning to create their own cyber ranges? Although processes have been honed over the past two decades to support cyber range experimentation, we posit that further areas in the experiment life-cycle should be considered.

This position paper is organized as follows: In Section II we discuss the need for cyber ranges and brief background of cyber ranges in general. In Section III, we explore the NCR as a blueprint for a cyber range, examining the goals and capabilities such ranges provide. Finally, in Section IV we support our position through observations with numerous trials, and provide further areas to consider for the experiment life-cycle.

II. The Path to Cyber Experimentation

Networked information systems play a key in role supporting critical government, military and private computer networks. 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,

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

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. Physical testbeds for analysis and training are typically expensive and time-consuming to construct and deploy, are single-purpose, and difficult to maintain. Furthermore, 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 may become extremely challeging with regard to large-scale complex systems [1].

The methodology of developing a cyber analysis platform and training capability includes asking numerous questions, such as:

- 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?
- 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?

The solution to this problem may lie in cyber ranges. According to [2], a cyber ranges are "interactive, simulated representations of an organization's local network, system, tools, and applications that are connected to a simulated Internet level environment." Cyber ranges may be virtual, emulated, or include hardware-in-the-loop (HITL), ranging from stand-alone ranges in a single organization, to multiple ranges emulating the Internet, remotely accessible from anywhere.

In this way, the range should provide testers and analysts a mission-safe and legal environment to do such things as training, mission testing, cyber security effects, etc. The ranges should be comprised of both hardware and software components that may be flexibly allocated and configured to support the range experiments at hand - and be capable of connecting to cyber ranges of other organizations as well. For high density emulations, such as Internet level, realism should be brought to the environment thorugh service and device fidelity (web services, email, application traffic) and emulated traffic as needed.

The range capability should be technologically competent and flexibility enough to support a myriad of customers and missions. Not just cyber security specialists, but it must also be able to specify the context in which such networks would operate (e.g., military applications, law enforcement, industrial control systems, academic institutions, etc.). Tracks should be in place to adequately gather information from Subject Matter Experts (SME) about the domain to build faithful representations.

In the end, the range should be able to provide [2]:

1) Real-time feedback with high-fidelity simulation;
2) An environment where teams can engage to support the range experiment
3) An environment where hypotheses may be tested by various teams.
4) Performance-based assessment metrics and data;

Cyber ranges are in use and provided by organizations across the Government, Private Industry, and Academia [3]. University test ranges are primarily used for education/training and many of which have just been stood up within the past few years [4]. One of the more established university-related cyber ranges is Merit Network in Michigan [5], which is governed by 12 of Michigans public universities. The cyber range has been opened since 2012 and has four physical locations as well as a virtual training environment called "Alphaville" to test cybersecurity skills. The DETERLab at the University of Southern California is another notable testbed due to its size, funding sources, as well as its integration with other test beds [6].

Industry-based cyber ranges are designed for training as well as for companies to find weaknesses in their networks. Some of these are specifically designed for critical infrastructure, while others are open to any type of industry. IBM is breaking ground in this space with a recent $200 million investment in cybersecurity in 2016 which includes what it touts as the first physical cyber range for the commercial sector [7]. It is a huge, 153,000 square foot facility in Massachusetts where participants can take control of a fake Fortune 100 company.

Given the trend and need for specialize testing and training environments, we can anticipate more cyber-related ranges being established in near future, as well as expansion among the existing facilities.

III. A Blueprint for Cyber Ranges

To further our discussion on cyber ranges, we use one particular well-documented cyber range to illustrate a pragmatic model for the experiment life-cycle. The National Cyber Range (NCR) is an innovative Department of Defense (DoD) resource originally established by the Defense Advanced Research Projects Agency (DARPA). NCR is now owned by the DoD Test Resource Management Center (TRMC) that is

currently managed by Lockheed Martin. It provides a unique environment for cybersecurity testing throughout the program development life cycle using new methods to assess resiliency to advanced cyberspace security threats. The NCR capability, when applied, allows the DoD to incorporate cybersecurity early to avoid high-cost integration at the end of the development life cycle [8]. The NCR not only falls under the TRMC, but is also aligned with Corporate Operations, the T&E Range Oversight, Test Capabilities Development, Interoperability, and Technologies Development. It is accredited by the Defense Intelligence Agency (DIA) and can operate at levels up to Top Secret/Sensitive Compartmented Information. The graphic below highlights the four key components of the NCR [9]:

- Secure facility
- Unique security architecture
- Integrated tools for cyber testing
- Multi-disciplinary staff

The graph shown in Figure 1 provides a snapshot view of how the NCR has been used since its creation. A proliferation in usage and number of applications has been realized since 2013. In FY11 when the range originally opened, it was utilized only one time for cyberspace capability development, testing, and evaluation (DT&E). Since then, the range has been increasingly utilized and, in FY16, was used 58 times for a variety of projects, compared to only 8 times in FY13. The range is most frequently used for cyber training/exercises which is closely followed in number of uses by Major Defense Acquisition Program (MDAP) Cybersecurity DT&E [10] [11] [12] [13] [14]. This increasing for cyber security speaks to the need for effective deployment of range environments to support a broad, far-reaching need for many organizations.

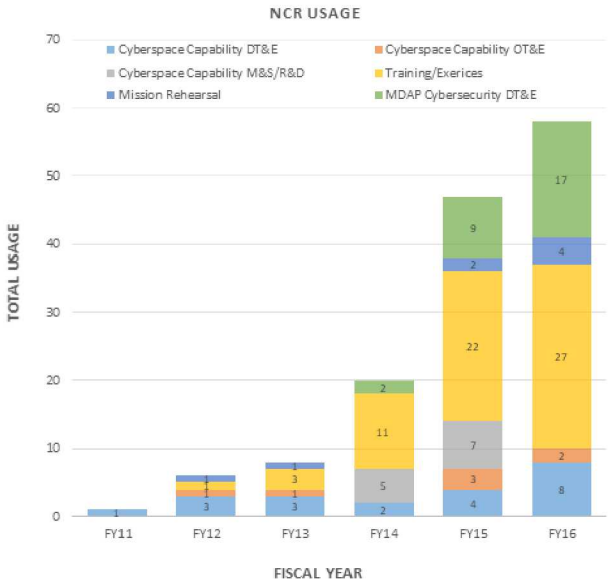

Fig. 1. NCR Usage

The NCR, as well as other government-use ranges, are highly focused on experiment support in the cyber domain. To do so, NCR topologies are comprised of complex networks that include high-fidelity representations of public and private infrastructures supporting various network services and architectures, to include sensitive DoD network enclaves. NCR customers may access the its cyber capabilities through a development and operational life-cycle. The series of events that transpire in an experiment may be executed securely using test-rooms at the NCR facility, or remotely from authorized sites. The process for deploying a test environment in the NCR is comprise of several steps that form an test life-cycle. Those steps are shown in Figure 2 [15].

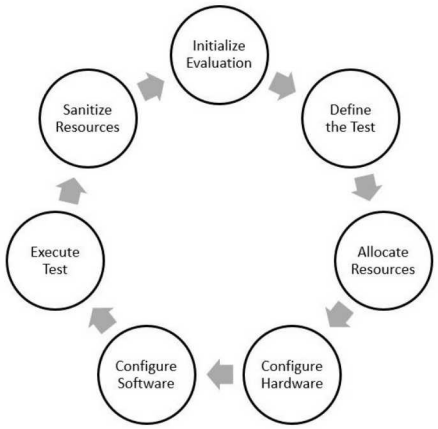

Fig. 2. NCR Life-cycle

The process starts with a common pool of hardware and software resources, and the Cyber Tool Set. Next, a *Test Specification Tool* is used to to define end-to-end aspects of the test. *Resource Allocation* consists of determining what resources from the pool are needed, which are then allocated to the Event. *Range Provisioning Tools* automatically wire the hardware assets to the appropriate configuration(s). *Range Configuration* tools then automatically configure the software needed to run the event. Following configuration, *Test Execution Tools* are used by the event team along with event-specific systems for execution and data collection/analysis. Finally, following completion of the experiment, *Sanitization Tool* sanitize hardware and "virtually" puts resources back into the pool.

The NCR may represent complex network topologies with sufficient realism to portray a variety of attack strategies [16]; NCR's sanitization capability enables assets to be sanitized at the conclusion of an event and reused in future events that may be at different classification levels. As a result, users can conduct advanced developmental and operational tests and evaluations, and provide realistic operational training in environments that emulate specific computing, networking, and information systems environments.

The NCR experiment life-cycle provides a coarse, stepped process to execute a cyber experiment event. The life-cycle is sufficient to capture at a high-level a generalized approach to

see an event through. In our observations, we feel there are additional requirements that may be included in the life-cycle, either at the same tier or as sub-requirements.

IV. Limitations and the Needs of Tomorrow

As cyber ranges become the "go to" place to conduct cyber tests, we find more research and development into the betterment of the preparation, deployment, and assessment methodologies they use [17] [18] [19] [20]. It is our opinion that cyber ranges are quite valuable to not only understand the effects of cyber attacks, but also to provide fertile ground to train future cyber defenders. What follows are our observations of several shortcomings in operational testing, and what areas are needed for future T&E in cyber ranges.

Integration Conferences: When in the initial planning phases of an experiment, conference calls may be struck up to ensure all stakeholders are working together toward the same goal. However, in experience, as time goes on such conferences begin to wain, and integral sub-teams are left to progress in their own lanes under their own assumptions. We have seen this result in great investments in misaligned efforts that must be discarded. In another instance, significant delays occurred in distributed provisioning information from many agencies. The delays impacted the execution of task items on the critical path and did not allow sufficient time for the integration of dependent environments into a functioning whole environment. The need for regular conference time should be adhered to, for the system designers, configurers and deployment teams to discuss possible issues and walk through the deployment in a step-wise fashion as progress is made.

Licensing: To promote fidelity in environments, sometimes it is the case that vendor-specific devices and software be used. In one such case, the required license for a device could only be activate online; however, the environment was not connected to the Internet. In another instance, licensed firmware for a device could not support required features. An extremely slow, out-of-band connection was required to download the new licensed firmware - which set back deployment by several hours. Through the environment specification stage, attention should be given to unique devices to the network so as to anticipate such issues regarding licensing.

Remote Access Management: With large-scale deployments of interconnected systems, various parties may require access to various subsystems (users, blue/red teams, developers, deployment team). The number of individuals requiring remote access to the systems should be tallied as a predeployment measure to ensure adequate remote access resources are available. In one scenario, a single Windows machine was setup as a remote log-in server. The user had to routinely logoff disconnected users in order to have enough free memory for the desktop (20-some people should not be logged into the same host at the same time).

Credential Management: As multiple portions of the environment come together, it may be the case that the assets of others may overlap or become integrated into another's area of responsibility in the network. Thus, when requiring access to another's assets, the appropriate credentials and roles should be known to the accessor. During one particular build out of the event environment, multiple different organizations provided sections of the network infrastructure. Frequently, the necessary credentials to access and troubleshoot system components in the environment were not handed off or made available (e.g., router and user-workstation passwords).

Configuration Management: Multiple organizations may contribute to the event environment. As their portions are integrated, reconfiguration of assets may be required to properly connect the networks. During one build out of an event environment, multiple organizations provided sections of the network to a single party, who would work with the subject matter experts from each of the organizations and then package and send captures of the environment to the range provider for integration. The environment would be reconfigured without logging changes, leaving differences between the actual environment and what the organizations sent to the range provider; as a consequence, images needed to be reloaded multiple times to fix misconfigurations. A single configuration management mechanism should exist to keep configuration versions, both for current and historical purposes.

Documentation: Network Maps and Instructions: When several players are responsible for different parts of the environment, it is essential that communication between the parties be established. Specification and deployment details should be appropriately documented and conveyed through these channels. We have seen the results of poor communication and documentation from parties, resulting in uncertainty in responsibilities, when things needed to be done, and who could provide help when needed. Or delays caused by naming issues of devices on network maps vs. device names on VMs. Often, responses are veryslow, as email is not a good means to convey an issue. Shared storage, improved documentation and detailed network maps should be provided by the design team, along with efficient communication mediums such as IRC channels.

Automation and Pre-configuration Templates: Configuring devices and servers should take advantage of automation mechanisms, or out-of-band management applications to deploy service/device templates or actual configurations. We have seen instances where networking devices were not configured ahead of time, resulting in field engineers entering configurations line-by-line, where a copy and paste operation in a terminal would have save considerable amounts of time.

Build in Debugging Processes: Establish a critical path for debugging systems when initial tests do not return expected behavior. Not unlike a contingency plan, the process should include written points of contact for specific systems and a tiered support chain, as needed for the experiment. This should be done with system designers and deployment team. This should attempt to prevent disruption of previous known configurations that then become suspect.

Periodic Hardware Testing and Refresh: Between deployments, nodes and servers should undergo period performance testing; however, it should be noted that ancillary equipment should also undergo testing. Range equipment is often "air-gapped" (physically or logically) from operational networks to uphold accreditation boundaries and prevent bleed over. As such, getting information in and out of the networks may be cumbersome. One observation noted the simple process of burning a disc in a room and trying to get a file onto the higher network turned out to be disastrous due to DVD burners and readers failing, resulting in wasting many man hours.

Architectural Integration Testing: When multiple ranges are connected, integration of the logical networks should not occur until a proper test plan has been developed and executed to test the physical connections between ranges. During a deployment of virtual machines (VMs) incorrect operation of system was identified; numerous devices were isolated and not showing expected connectivity. After moving the virtual devices on to a single compute blade, the deployed VMs operated as expected. It was determined that the infrastructure switching was not completely functional or reliable. Deliberate analysis and testing of transport between all sites must be conducted prior to execution.

Experiment Integration: Although a necessary evil, the use of a single integrator for an event can be problematic. One exercise concept had a single integrator for the various architectures and technical requirements. This created a bottleneck in the dissemination of technical information to others for interfacing requirements. In another exercise, multiple organizations providing systems for use in the environment had different levels of knowledge of the proper functioning of their systems, yet no one organization had cross-system knowledge. making it difficult to compile the data necessary to perform proper function checks on the environment. The integrator should not be a single person, but a team consisting of members from the relevant subsystems.

V. Conclusion

In this paper we discussed the need for cyber ranges, explored the NCR as a blueprint for a cyber range, and examined the capabilities and benefits such ranges provide. Every experiment carried out in a cyber range will have various nuances due to the goals of the experiment; nuances should be captured and addressed in the life-cycle of the experiment. It is our position that the basic structure of experiment life-cycles is not enough, flexibility should address the many challenges and shortcomings that arise out of cyber range testing.

References

[1] R. Fujimoto, C. Bock, W. Chen, E. Page, & J. H. Panchal (Eds.), 2017, Research Challenges in Modeling and Simulation for Engineering Complex Systems, Simulation Foundations, Methods and Applications Series, Springer International Publishing.

[2] NIST, Cyber Ranges, National Initiative for Cyber Education, https://www.nist.gov/sites/default/files/documents/2018/02/13/cyber_ranges.pdf

[3] D. Lohrmann, "Cyber Range: Who, What, When, Where, How and Why?," 2018, http://www.govtech.com/blogs/lohrmann-on-cybersecurity/cyber-range-who-what-when-where-how-and-why.html

[4] C. Franklin Jr., "7 University-Connected Cyber Ranges to Know Now," 2018, https://www.darkreading.com/cloud/7-university-connected-cyber-ranges-to-know-now/d/d-id/1331224?

[5] Merit Michigan Cyber Range, 2018, https://www.merit.edu/cyberrange/

[6] The DETER Project, 2018, http://deter-project.org/

[7] IBM, "IBM Invests $200M to Help Clients Respond to Cybersecurity Incidents," 2016, https://www-03.ibm.com/press/us/en/pressrelease/51066.wss

[8] B. Ferguson, A. Tall and D. Olsen, "National Cyber Range Overview," 2014 IEEE Military Communications Conference, Baltimore, MD, 2014, pp. 123-128.

[9] National Cyber Range, 2018, https://www.acq.osd.mil/dte-trmc/ncr.html

[10] J. B. Hall, Department of Defense Development Test and Evaluation FY 2016 Annual Report, https://www.acq.osd.mil/dte-trmc/docs/FY2016_DTE_AnnualReport.pdf

[11] C. Brown, Department of Defense Development Test and Evaluation FY 2015 Annual Report, https://www.acq.osd.mil/dte-trmc/docs/FY2015_DTE_AnnualReport.pdf

[12] C. Brown, Department of Defense Development Test and Evaluation FY 2014 Annual Report, https://www.acq.osd.mil/dte-trmc/docs/FY2014_DTE_AnnualReport.pdf

[13] C. Brown, Department of Defense Development Test and Evaluation FY 2013 Annual Report, https://www.acq.osd.mil/dte-trmc/docs/FY2013_DTE_AnnualReport.pdf

[14] S. Hutchinson, Department of Defense Development Test and Evaluation FY 2012 Annual Report, https://www.acq.osd.mil/dte-trmc/docs/FY2012_DTE_AnnualReport.pdf

[15] National Cyber Range Overview, 2015, https://www.acq.osd.mil/dte-trmc/docs/20150224_NCR%20Overview_DistA.pdf

[16] J. Keller, "Lockheed Martin continues work on National Cyber Range training for exceptionally virulent code," Military & Aerospace Electronics, https://www.militaryaerospace.com/articles/2018/01/cyber-training-virulent-code.html

[17] Department of Defense Cybersecurity Test and Evaluation Guidebook, 2018, https://www.acq.osd.mil/dte-trmc/docs/CSTE%20Guidebook%202.0_FINAL%20(25APR2018).pdf

[18] J. Vykopal, M. Vizvary, R. Oslejsek, P. Celeda and D. Tovarnak, "Lessons learned from complex hands-on defence exercises in a cyber range," 2017 IEEE Frontiers in Education Conference (FIE), Indianapolis, IN, 2017, pp. 1-8.

[19] J. Vykopal, R. Ošlejšek, K. Burská, and K. Zákopčanová. "Timely Feedback in Unstructured Cybersecurity Exercises," In Proceedings of the 49th ACM Technical Symposium on Computer Science Education (SIGCSE '18). ACM, New York, NY, USA, 173-178 Pages 173-178

[20] M. Frank, M. Leitner and T. Pahi, "Design Considerations for Cyber Security Testbeds: A Case Study on a Cyber Security Testbed for Education," 2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech), Orlando, FL, 2017, pp. 38-46.