Australian Government
Department of Defence
Defence Science and
Technology Organisation

A Survey of Cyber Ranges and Testbeds

Jon Davis and Shane Magrath

Cyber and Electronic Warfare Division
Defence Science and Technology Organisation

DSTO-GD-0771

ABSTRACT

This document reviews the state-of-the-art in Cyber Range implementations and related computer network operations (CNO) testbeds. We summarise recently published examples and describe their purpose and functionality. The compiled information should assist organisations to make an informed decision when considering a Cyber Range capability.

RELEASE LIMITATION

Approved for public release

Published by

Cyber Electronic Warfare Division
DSTO Defence Science and Technology Organisation
PO Box 1500
Edinburgh South Australia 5111 Australia

Telephone: 1300 333 362
Fax: (08) 7389 6567

AR-015-770
October 2013

APPROVED FOR PUBLIC RELEASE

A Survey of Cyber Ranges and Testbeds

Executive Summary

This paper reviews recent publications about Cyber Ranges (CR) and computer network operations (CNO) testbeds. The analysis provided in this review aims to assist organisations in making an informed decision when considering a CR capability.

A CR provides an environment to practice CNO skills such as penetration testing, defending networks, hardening critical infrastructure and responding to attacks. It should represent real-world scenarios such as emulating large-scale, complex networks. It should offer isolation from other networks to contain malicious activity. The same environment should also support experimentation and testing with cyber security products.

We are interested in exploring the approaches used to build existing CRs, the merits of each approach and their functionality. Of particular interest are cost-effective ways to obtain a CR capability. Hence this review looks for the availability of mature software for constructing and managing a CR, software for monitoring and analysis, training scenarios, communities for collaboration and commercial offerings. By exploring published approaches we aim to outline the diversity of options available and hence assist decision-makers to find an approach which best meets their requirements.

The review first categorises CRs by their type, and second by their supporting sector: academic, military or commercial. The types of CR are identified as simulation, overlay or emulation. CRs are considered simulations if they use software models of real world objects to explore behaviour. They are labelled as overlays if they operate on live production hardware with experiments sharing their production resources, rather than using a dedicated CR laboratory. Lastly, CRs are labelled emulations if they run real software applications on dedicated hardware. Emulation refers to the software layer which allows fixed CR hardware to be reconfigured to different topologies for each experiment.

Most CRs in the review can be categorised as either a simulation or emulation. Simulations have high scalability and generally operate on either a single or a small number of servers. Therefore they are easy to deploy and relatively cheap to install and maintain. However several academic papers question whether test results from a

simulation reflect reality. This issue can be mitigated by only simulating the high-level impacts of computer network attacks. This is the approach taken by several of the simulation CRs in the review.

Emulation CRs support high fidelity testing. Since emulation uses real computers, operating systems and applications with limited resources, the experiments represent a realistic environment. This increases the likelihood that test results will apply to a deployed situation. However these testbeds are generally high cost, since a significant amount of hardware is required to emulate large networks. This cost can be reduced (and the scalability of the CR increased) through virtualisation. For example, some CRs in the review support multiple types of virtualisation to allow the experimenter to select high fidelity components of interest and low fidelity, but high scalability, for the rest.

While this review distinguishes simulations from emulations, CRs can use a mixture of both. By supporting both simulation and emulation in a CR, the best aspects of each can be utilised.

The review found CRs are predominantly used for training. The training varies in complexity from computer security fundamentals to advanced tactics, techniques and procedures for computer network operations (CNO) staff. Another popular form of training is custom exercises, where staff practice CNO in a competitive environment.

This does not diminish the importance of other CR roles such as cyber security testing and research and development. CRs with these roles were also found. For researchers, the review identified a trend towards using federated CRs and testbeds. This conclusion was based on two large initiatives called FIRE in Europe and GENI in the US. These projects currently link testbed resources to assist sharing and reuse among researchers.

The review also identified several CRs which make their software publicly available and which are supported by an active development community. These potentially provide a cost-effective way to build and maintain a private CR.

The US military has already invested significantly in CRs. Several mature US military CRs are described in this review. These could provide an effective capability if access is available.

Contents

Glossary

API	Application programming interface
CNCI	Comprehensive national cyber security initiative (US)
CNO	Computer network operations
CONOPS	Concept of operations
CR	Cyber range
CTF	Capture the flag
DARPA	Defense Advanced Research Projects Agency (US)
DDoS	Distributed denial of service
DHS	Department for homeland security (US)
DNS	Domain name system
DoD	Department of Defense
DPI	Deep packet inspection
EW	Electronic warfare
FIRE	Future internet research and experimentation (EU)
GENI	Global environment for network innovations (US)
GiG	Global information grid
GUI	Graphical user interface
ICT	Information and communications technology
IA	Information assurance
IDS/IPS	Intrusion detection system / intrusion prevention system
IO	Information operations
IP	Internet protocol
IPSEC	Internet protocol security suite
JIOR	Joint information operations range
LAN	Local area network
LVC	Live, virtual and constructive
MFD	Multi function device
MILCOM	Military communications
MILDEC	Military deception
NCR	National cyber range (US)
NSF	National science foundation (US)

NSTB	National SCADA testbed (US)
OPSEC	Operations security
PSYOP	Psychological operations
SAM	Surface-to-air missile
SCADA	Supervisory control and data acquisition
SOE	Standard operating environment
TTPs	Tactics, techniques and procedures
USAF	US Air Force
USCYBERCOM	US cyber command
USSTRATCOM	US strategic command
VM	Virtual machine
VoIP	Voice over internet protocol
WAN	Wide area network

1. Introduction

In this paper we review Cyber Range (CR) implementations in an attempt to find state-of-the-art implementations.

1.1 Scope

This review is limited to public-domain (unclassified) information about CRs. Since CR work has been extensively funded by the US military, it is likely interesting classified projects exist which are not covered here. On the other hand, the US military realises that defending military networks is necessary, but not sufficient, to ensure their long-term mission success. Their nation's critical infrastructure, government networks and large private companies must also be protected from cyber attack [1]. This is both to maintain economic vitality and to ensure continued supply of goods from the private sector. Hence there is an intended flow of computer and network security capability from the military to other sectors as evidenced by the work-in-progress National Cyber Range (NCR). The NCR is a large military-sponsored initiative and is intended for use by the military, commercial, academic and government sectors.

While this review focuses on CRs, we also include facilities which support a substantial subset of CR capabilities (for a comprehensive list of capabilities see the NCR requirements [2]). We therefore include a number of testbeds which replicate large, heterogeneous networks, enable multiple simultaneous experiments, assess information assurance or support rigorous scientific testing. In the literature, these testbeds are given various labels such as an attack lab, a testbed for CNO, or a testbed for network warfare or cyber war.

A facility with broader goals than a CR is an Information Operations (IO) Range. IO is defined as the integrated use of electronic warfare (EW), CNO, psychological operations (PSYOP), military deception (MILDEC) and operations security (OPSEC). IO Ranges are therefore not specifically covered in this review as they are a significant extension to a CR.

1.2 Related Work

While similar reviews are available in the open literature, they were not considered sufficient for our purposes. An informative review from the Royal Military College, Canada, in 2011 covers the similar topic of state-of-the-art CNO simulation and modelling [3]. It discusses approximately 13 simulation-type CRs categorizing them into private, academic or public sector research. In comparison, not only does this review cover modelling and simulation-type CRs, but also ad hoc and emulation types. Hence we cover a wider range of CRs and testbeds (approximately 30).

Another review from 2010 analyses the software tools underpinning network testbeds [4]. Their analysis is limited to publicly available software used in those testbeds. Other researchers can use this collated information to decrease the cost of developing their own testbed. Again, our review is broader in scope since it discusses CRs and testbeds irrespective of whether their software is publicly available.

1.3 Contributions

The main contribution of this paper is to review current and previous work on CRs from military, academic and private sectors. The review discusses the general approach used to construct the CRs and what role each serves. The paper then analyses this information to identify trends.

The remainder of this review is structured as follows. Background information about CR in provided in Section 2. This includes the motivation for building CRs and their relation to testbeds. Section 3 begins the main part of the review, categorising each CR as one of three types: simulation, overlay or emulation. These types are described in detail in Sections 4, 5 and 6, which also describe a number of reference CR implementations. Section 7 lists some capture the flag (CTF) competitions, which are used to develop and assess CNO skills in a similar way to CRs. A discussion of the findings is then presented in Section 8.

2. Background

The word "range" implies an environment for offensive target practice, much like a shooting range for soldiers. A *cyber* range would therefore be an environment where CNO staff can practice skills such as penetration testing, defending networks, hardening critical infrastructure and responding to attacks. The environment should emulate large-scale complex networks to reflect real-world scenarios and offer isolation to contain malicious activity. A CR therefore provides a realistic environment suitable for training CNO staff. The same environment should also support experimentation and testing with cyber security products.

The concept of a CR is relatively new compared with other military ranges, for example shooting ranges. As such, CR capabilities are still developing. CRs are built to support CNO which has only recently received high recognition. In 2009 the US created USCYBERCOM which centralises command of CNO across the US Army, Navy, Air Force and Marine Corps. This organisation requires a realistic training and proving ground to prepare and conduct military CNO. That is, they need a CR.

Operational networks are not suitable for testing CNO capability. Firstly, there is a high risk of adverse impacts on that network and its services. Secondly, operational networks cannot be fully controlled, so tests are not generally repeatable as required for rigorous experimentation. To solve these problems, a standalone, controlled environment such as a CR or testbed is usually built. However, for cost reasons, standalone networks have historically been small scale and overly simplified versions of operational networks. The applicability of results to real world scenarios is therefore questionable. Hence, new CRs aim to be built at-scale to provide a realistic environment. Currently, the most ambitious of these is the National Cyber Range (NCR) under construction in the US. It aims to be sized appropriately to run Global Information Grid (GIG) or Internet scale testing. While DARPA is the lead agency, the NCR is not intended to be used solely by the military. Instead it is part of the Comprehensive National Cyber security Initiative (CNCI) to increase the nation's defences against electronic attack. The NCR will be open to industry, government, military and academia, with resources allocated dynamically for both classified and unclassified work. One of the NCRs key requirements is the automation of both the test processes and range management. This should accelerate the testing and validation of cyber tools. A requirements document for the NCR is provided in [2].

Introductory presentations about the NCR compare the testing environment of US adversaries versus current CRs. Adversaries obtain access to real US networks with real users. They can therefore test the effectiveness of their tools directly on large scale heterogeneous networks. Conversely, as discussed above, US CNO staff have historically only tested on small standalone ranges. This motivates building a large scale NCR which is representative of real military and commercial networks.

The US Air Force (USAF) is a leader in CRs. It has used CRs as far back as 2002 when SIMTEX underpinned an exercise called Black Demon. In 2011 it won a US National Cyber security innovation award from SANS [5]. This was awarded to the 39th Information Operations Squadron for its use of a CR to train staff in IO tactics, techniques and procedures (TTPs).

An interesting recent trend in network testbeds has been to reduce costs by sharing infrastructure. This has been done by creating federated testbeds shared by researchers across the globe. An example of this in the US is the current Global Environment for Network Innovations (GENI) project sponsored by National Science Foundation [6]. This open project creates the opportunity for researchers to access a large virtual laboratory for network experimentation without having to build their own. In Europe, a similar large initiative is called Future Internet Research and Experimentation (FIRE) [7]. This project started in 2008 and is connecting existing and upcoming testbeds for internet technology research and testing. Given that federation has potential benefits of lower cost and greater capability, it is likely that CRs will increasingly adopt a similar approach.

2.1 Functionality

CRs create a realistic cyber environment for a number of different roles including:

- project support
    - product evaluation
    - testing fitness for purpose

    - specific capability development
    - benchmarking testing
- operational support
    - skill development and training for Cyber Operations
    - war-gaming or capture-the-flag exercises
    - doctrine development
    - counter-cyber warfare
- research and development
    - research advanced cyber threats
    - experimentation
    - ideas development
    - tools development.

Given these diverse roles, it is important that before any CR is implemented that a clear concept of operations (CONOPS) is developed. The CONOPS will make clear the purpose and types of usage the capability will provide and this will strongly guide development.

The breadth of information and communications technology (ICT) hardware used in organisations is large and growing. The literature indicated that cyber test labs (not explicitly called CRs) have been developed for a number of specialised areas including:

- LAN/WAN enterprise equipment
- telecommunications carriers
- voice networks, for example PABX, VoIP infrastructure
- SCADA systems
- data-centers, for example servers, SANs, clusters
- MILCOMs networks
- mission critical control systems, for example avionics
- end-user environments, for example desktop SOEs, smart phones, printers, MFDs.

What is clear is that as more devices become networked, the cyber attack surface grows. It is therefore important that any proposed CR facility has a clearly defined technology profile that limits the scope of its usage else they risk being very expensive or of diminished value through diluted focus. It is possible that federated CRs could assist with this task. Individual CRs could specialise in one technology area, for example SCADA. Other CRs could then access these specialised facilities to run experiments. The National SCADA testbed (NSTB) already does this. It consists of multiple linked laboratories including SCADA, cyber security and wireless laboratories at Idaho, and cryptographic and red team facilities at Sandia.

3. Cyber Range Review

We now present the reviewed CRs and categorise them by their method of experimentation. We considered the methods defined in [8]:

1. Simulations: can be used where models of each component exist. Using these models, networks can be simulated and their performance monitored.
2. Ad hoc or overlay testing: involves running tests on production network hardware with some level of test isolation provided by a software overlay.
3. Emulation: is the process of mapping a desired experimental network topology and software configuration onto the physical infrastructure of the CR. While the actual infrastructure may consist of a cluster of machines, the emulation component can configure the cluster to behave as per the desired experiment topology, for example multiple linked USAF bases.

An additional method is called analytics. This is usually restricted to low complexity problems where simplifying assumptions are made to keep the problem tractable. Since no CRs using the analytics approach were found in the review, it is therefore omitted from this paper.

Other papers use an alternative taxonomy called live virtual constructive (LVC). This describes a continuum of approaches from real users interacting with real systems (live), through to simulated users operating simulated systems (constructive). Live is generally high cost, hard to repeat, but very realistic. Constructive generally has the opposite characteristics. While LVC terminology is used in several CR papers, we prefer to use the four categories outlined above since they provide a good basis for comparing the different CR implementations.

4. Modelling and Simulation

In simulations, models of real world components (such as ICT infrastructure) are created, and these models are then made to interact. Instrumentation is used during the simulation to measure performance. Simulation has the advantage of being highly scalable, since a large number of host and device models can be run on a single physical machine (assuming simple models). However, a disadvantage is that because these models are often high level abstractions of the real world objects, the test results can lack fidelity and may not reflect reality. Developing models which capture the complex, dynamic and stochastic nature of networks and computers requires a significant effort.

The use of pure simulation for cyber-security research and development was initially very popular. However now other CR types have gained in popularity. This appears to be for a few main reasons:

- Simulators never really demonstrated the fidelity and complexity of actual cyber-attacks.
- Hardware has become more affordable and virtual machine technology more available. Hence setting up a large scale testbed is more achievable.
- Open source and commercial penetration testing tools and cyber-attack databases simplify the use of real attacks in a testbed. The availability of real attacks reduces the need to create simulated attacks.

A recent paper [3] discussed how simulation has been an effective tool for war gaming. The US Army, Navy and Air Force use war gaming for training and experimentation in their core domains. It has proven to be a safe and cost-effective way to assess new technologies before deployment into the battlefield. A natural progression is therefore to also use simulation for CNO war gaming. We now discuss the use of these simulators in the military, academic research and by commercial companies.

4.1 Military and Government

The number of CR papers published by military organisations indicates the extent to which they are used. Probably due to the classified nature of the work, most technical details are omitted from these reports and generally the software is not made available.

4.1.1 SIMTEX

The “Cyber Flag” concept of operations was proposed in 2008 [9]. It aimed to provide a training environment for CNO in the USAF in much the same way as “Red Flag” has been a realistic training environment for aircrew over decades. A future goal in the proposal was to integrate Cyber Flag and Red Flag to have a realistic environment to explore the role and impact of cyber tools in a simulated military exercise (i.e. it would eventually become an IO Range). The proposal listed some training gaps including: showing the relevance of CNO to military objectives; the lack of training on non-TCP/IP networks; defining a set of cyber targets to achieve an operation; choosing from a suite of cyber attacks; and obtaining battle damage assessments.

The USAF chose the Simulator Training Exercise Network (SIMTEX) for Cyber Flag [10] (although LARIAT was also considered, see Section 4.2). SIMTEX was developed for the USAF based on their training requirements [11]. The primary requirement was initially to create a secure network environment for network operations staff to practice network and server troubleshooting. SIMTEX was used in Cyber Flag 2011 where the 200 participants engaged in red team versus blue team exercises.

SIMTEX has a long history in the USAF, being used in the Bulwark Defender Exercise since 2006, and in Black Demon in 2002 and 2003 [12]. These two exercises required participants to defend simulated computer networks under attack from the NSA and other USAF squadrons. SIMTEX simulates the mechanics of computer network attacks on a network architecture mimicking the USAF network, and also has an "internet simulator" for added realism. It can provide automated attack scenarios that can be rebaselined in less than ten minutes. It also supports multiple locations remotely connecting to the SIMTEX network which is useful for training or exercises. It is not totally clear from the literature how SIMTEX is implemented. It appears to be a complete simulation, but it is also possible it can include real hosts and attacks on an isolated training network.

SIMTEX's use in exercises specifically addresses cyber activities. Addressing the broader IO impact of cyber attacks including their kinetic effect is the purpose of another military simulator called Cyber and Joint Effects Demonstration (CAAJED) [13].

4.1.2 CAAJED

CAAJED is a USAF project which is the integration of a commercial war game simulator called Modern Air Power (MAP) with a cyber/kinetic inference model called Simulated Enterprise for Cyber Operations Training (SECOT). SECOT also contains a scoring system. CAAJED therefore focuses on the high-level effects of cyber attacks in a war scenario. It was used in the Cyber Defence Exercise in 2007. Before the exercise, students had weeks to prepare their physical networks. They were given a set of simulated defensive assets and offensive targets. During the exercise, students aimed to achieve their mission while suffering loss of assets due to kinetic attacks outside their control.

CAAJED allows the interaction between the kinetic and cyber domains to be investigated. Physical assets exist in the kinetic domain, while processes which control them exist in the cyber domain. A capability is the synergy of the two, for example the ability to control the launch of a surface-to-air missile (SAM) depends on the physical SAM site and the communications to it. CAAJED is set up so capabilities are open to attack through cyber vectors. The impact of these attacks is then shown in the simulation, such as disabling a radar site. As such, CAAJED is aimed to enhance training and develop TTPs.

The longer term aim of many simulations is to be a "full spectrum virtual environment". This environment would replicate the real world with high fidelity and include instructor support, automated performance measurement and replay functionality. A set of requirements for a full spectrum virtual environment for cyber training is given in [12] and includes well-specified learning objectives, realistic context, scenario developments, reproducibility and performance measurements.

4.1.3 SAST

The Security Assessment Simulation Toolkit (SAST) was chosen for specialised training of USAF CNO staff [11]. SAST was developed by the Pacific Northwest National Laboratory (PNNL) to simulate the network environment of many DoD organisations. Its main capabilities are:

- Secure Environment for Accelerated Learning: an isolated network which simulates a large network under attack
- Multi-User Training Tool (MUTT): realistically simulates millions of users performing normal work activities to provide realistic background traffic
- Coordinated Attack Tool (CAT): based on existing hacker tools it enables threats to be incorporated into simulations.

Although LARIAT and SIMTEX were also considered, SAST was chosen because it is primarily a specialised training platform and is easy to install and use.

4.1.4 StealthNet

A CR funded by the US Army called StealthNet is a Live-Virtual-Constructive (LVC) framework for cyber operations test, evaluation and training. It is a three year program started in 2010 [14]. Given the army's extensive use of wireless communications for tactical operations, StealthNet has first focussed on representing the impact of jamming and DDoS. It plans then to simulate computer hosts which have OS and browser vulnerabilities. StealthNet has many goals including the ability to assess the impact of cyber threats on tactical networks and net-centric systems under test. Real equipment can be connected to the virtual network and real sensor feeds can be sent through it. The impact of attacks can then be observed. The testbed therefore aims to assess the impact of cyber attacks on operational systems and missions. StealthNet also aims to interface with the Army's other LVC simulations. It is unclear from the literature what technologies are being used to implement StealthNet.

An example of the US Army's other LVC simulators is the US Joint War Fighting Center's (JWFC) Joint Training and Experimentation Network (JTEN) which simulates military operations. It includes sophisticated computer models to aid the simulation. This network has been connected from the US to allied networks including Australia's Defence Training and Experimentation Network (DTEN) [15]. Adding a Cyber component to such an LVC framework would enhance the simulation.

4.1.5 High-level wargaming

Simulation has also been used to test preparedness of organisations. For example, Cyber Storm exercises 1, 2 and 3 were held in the US in 2006, 2008 and 2010 respectively. They involved industry and military participants with the aim of examining the "preparedness, response, coordination, and recovery mechanisms to a simulated cyber event". Attacks were simulated over 4 days, requiring organizations to develop strategies to respond. This did not test their technical ability, but rather the policy response. The simulated attacks were against infrastructure in Energy, IT, Transport and Telecommunications sectors. Similar exercises involving fictitious scenarios in other countries include France's Piranet and India's Divine Matrix. The USAF also participates in high-level war gaming to explore strategies. Past war games have included "Unified Engagement" and "Future Capabilities Game". These high level simulations do not attempt to integrate the technical aspects of cyber attacks with their effects. Instead they concentrate on co-operation between stakeholders and developing appropriate responses. Adding the technical integration would lead to an increased understanding of the threat cyber attacks pose, but could also add a large amount of complexity to the exercise.

4.2 Academic

There have been several efforts from universities to simulate the effects of computer network attacks. These approaches are primarily used for training.

4.2.1 SECUSIM

A 1999 seminal paper in the field described a constructive simulation of cyber attacks, defences and consequences [16]. The simulation setup is similar to a game and runs on a single computer. An important innovation in their approach is the classification system which determines the success of attacks based on threats, attack types and protection measures. While it is an interesting simulation, the classification system is not validated and is hard to compare to real-world scenarios. Building upon this work another constructive simulator called SECUSIM was developed in 2001. The main advantages of SECUSIM were its GUI and being highly customizable.

4.2.2 RINSE

Another project used for training is the Real Time Immersive Network Simulation Environment (RINSE) from the University of Illinois in 2006 [17]. It aims to simulate a WAN consisting of hundreds of LANs. Attacks are carried out against the WAN by a game coordinator, and each user is tasked with protecting their LAN. The simulator mainly targets DDoS, worms and high-traffic situations; hence the simulator is very limited in its attack vectors. Students can be trained in using defence mechanisms such as adding packet filters to block or mitigate attacks and keep network services operating. The software is based on a simulation framework called SSFNet which does not appear to have been updated since 2004.

4.2.3 NetENGINE

A cyber attack virtual simulation tool called NetENGINE aims to provide IT staff with training to combat cyber attacks in large IP networks [53]. It provides a web-based GUI where each user can view network topology maps, router loads and network status. It is not intended to model the technical details of the attacks, but rather the outcomes and effects.

4.2.4 ARENA

Simulations have also been used to test the effectiveness of security tools and the resilience of networks. An example was developed at the Rochester Institute of Technology in 2007 using ARENA simulation software [18]. It models computer networks and intrusion detection systems (IDSs) and then applies simulated attacks. The resultant IDS alerts are then used to test downstream analysis tools. It is therefore aimed at testing situational awareness tools. Only high level models of attacks are used. The models include the attack type and target, and associated features such as efficiency, skill and stealth. That is no technical aspects of the attack are included. The user can construct a computer network using the simulator and define hosts on it with particular operating systems and installed IDSs. A simulated attack can then be launched against that network architecture. The focus of the simulation is to produce realistic IDS alerts, including simulated false positives (noise), based on the network topology and location of the IDSs. Users are then tasked with analysing the IDS alerts with situational awareness or other tools. The simulation also produces a ground-truth file to evaluate effectiveness. No low-level information such as packet-level data is produced.

4.2.5 OPNET-based

Other testbeds have used commercial simulation software as their basis. OPNET was used to generate probe and DoS attacks in an evaluation of a frequency-based IDS [54]. It has also been used to examine network performance under DoS. Both implementations are very limited subsets of a CR.

4.2.6 LARIAT

A well developed and useful CR is the Lincoln Adaptable Real-time Information Assurance Testbed (LARIAT) which has its origins back in 1997 [19]. It was originally built as an extension to the DARPA 1998 and 1999 intrusion detection data generation testbed. It was designed as a deployable testbed for information assurance by generating background traffic, real attacks, and verifying success or failure. LARIAT's differentiating feature is its ability to generate realistic user traffic through user simulation. Testbed staff are still required to build the test network, install operating systems on hosts, install applications and deploy defensive host and network tools. LARIAT then deploys virtual hosts and virtual users on top. The virtual users are driven by Markov models, each with a different user role interacting with applications, content and other users. Some internet traffic is also simulated. In this way, LARIAT is a mixture of simulation and real hardware. LARIAT is suitable for IO testing, as well as for security research. It runs applications and services natively, so vulnerabilities and flaws can be found and investigated. To simplify the process of setting up a LARIAT testbed, a GUI called Director was created. This improves test specification and control such as software deployment, troubleshooting, control and monitoring. LARIAT is one of a few simulation tools used within the USAF for training [11]. It has also been used for real time, automated testing of IDSs.

4.2.7 VCSTC

Virtual Cyber-Security Testing Capability (VCSTC) was another DoD-funded academic project for an automated testing capability to assess the security impact of a new device before deployment [20]. Their approach was to simulate a larger network and then automatically generate test cases to cover the security properties under test. The test harness ran on a single physical server and consisted of a hybrid honeynet with 4 VMs as "thick" nodes and 1 VM running Honeyd to emulate up to 1024 thin nodes. The setup is used to test security devices on the market. At the time of their published paper the automated test generation was not fully implemented, and consequently a Java developer required 2 days to write 50 test cases for a specific device under test. This reduces the usefulness of the capability.

4.3 Commercial

A number of commercial cyber security simulation products exist. A few are now described.

4.3.1 Breaking Point

Commercial appliances from Breaking Point [21] are advertised as providing CR capabilities. Their products provide traffic generation and a Strike Pack of network security and malware attacks in a single rack-mountable appliance. Traffic generation is highly configurable up to and including layer 7 of the OSI model. The appliance supports traffic generation for more than 150 popular layer 7 web applications such as Facebook and Google maps. It can simulate millions of users simultaneously to provide realistic background traffic. The strike pack includes over 4,500 live security attacks and 28,000 live malware attacks. The attacks can be setup with a large number of evasion techniques and sent, on the network, to devices under test.

Breaking Point uses simulation to achieve its high scalability. Large network topologies involving hundreds of thousands of hosts can be simulated in a single appliance. This has the major advantage to the purchaser of drastically reducing the hardware infrastructure setup and maintenance costs compared to an emulation-type CR. On the other hand, since hosts are simulated in this appliance, operators cannot interact with them in the same way as a real host. It could therefore be argued that the Cyber Range lacks “reality”. To overcome this, the appliance could be used in conjunction with other hardware to provide some interaction. The main uses of the Breaking Point appliance within a CR are to simulate traffic from a realistic network, test and harden DPI devices, research advanced cyber threats, and train staff.

Breaking Point customers who have used their appliances within a CR include [55, 56]:

- DISA: to generate large amounts of realistic user traffic, play canned scenarios, script simulated data flows and emulate MPLS and IPv6
- Northrop Grumman: use it in their Federated CR
- Cisco: to test their Firewall-IPS devices.

4.3.2 EXata

Scalable Network Technologies [22] offer simulation software products for planning, testing and training. One of these is called EXata Cyber which can form part of a Virtual Cyber Range. The product appears to focus on wireless communications. In particular it can simulate how communications would behave in battlefield conditions including eavesdropping, radio jamming and DDoS. The same company has a product called QualNet which has a large library of network protocol libraries for simulation. This includes Link 11, Link 16, satellite and wireless networks. According to their website, they have a contract with US STRATCOM to adapt the EXata Cyber simulation platform to meet specific training requirements of the military.

4.3.3 Building blocks

Many other commercial vendors offer tools which can be used as building blocks for a Cyber range. Examples include:

- OMNeT++: an open source simulation framework which provides APIs for designing networks and traffic modelling [57]. Commercial use requires a license. Similar tools include NS-3 and OPNET.
- Candela Technologies: background network traffic generator [58].

4.4 Modelling and Simulation Discussion

A variety of simulators have been used as the basis of a CR. Early implementations aimed to train staff about the effects of cyber attacks and about appropriate defensive measures. These tended to abstract details of the attacks and hence were hard to validate and justify against real networks.

From this review of simulation-style CRs, the most extensively used products in the US military were SIMTEX and LARIAT. These were used for training and exercises. From the commercial sector, a widely used product is the Breaking Point appliance which can simulate large networks and provides a library of cyber attacks for testing in a CR.

5. Ad hoc or Overlay

ad hoc (or overlay) test networks operate on top of production networks. The overlay network can dynamically change the topology of the network for applications under test, while the underlying network remains static. It therefore provides flexibility in configuring testbeds for different experiments. Overlays are a very common method for adding functionality to an existing service, for example the internet overlays the telephone network. Other networks such as peer-to-peer or cloud services now overlay the internet.

Using overlays for test networks provides size, cost and fidelity advantages compared to other approaches. This is because testing can be performed at the same scale as the production network it overlays, real hardware is used for high fidelity, and they avoid the additional cost of building a separate test network or laboratory. However, a disadvantage is the difficulty in performing formal testing including repeatability and experiment control (since the underlying network cannot always be controlled). Another disadvantage is the potential adverse impact of the experiment on the production network.

5.1 PlanetLab

PlanetLab [23] is a large global research overlay network. At the time of writing, PlanetLab consists of 1143 nodes at 545 sites connected via the internet. It is used to develop new technologies for many applications including distributed storage, network mapping, peer-to-peer systems and network security. This overlay test network is made possible by participating institutions running one or more PlanetLab VMs at their site. Each node is Linux-based and uses virtualisation techniques to divide resources into slices. Experiments can then be run across PlanetLab nodes by reserving resource slices. PlanetLab's software is publicly available. This allows institutions to create their own private PlanetLab as an alternative to joining the worldwide test network. An option then exists to connect these private labs to the wider PlanetLab at a later date. The software includes Plush which is a framework for deploying applications in a distributed environment. A derivative of PlanetLab is Everlab. This testbed was made by installing PlanetLab software on clusters of hardware across Europe. OneLab was also created as an extension of PlanetLab to support wireless research. The large European FIRE initiative has ongoing projects (such as OpenLab) which are based on PlanetLab. FIRE links federated testlabs via the internet to allow better sharing and reuse of resources. While PlanetLab is a general platform for deploying and testing large-scale services, it can be used for CR activities. The website lists the projects which have been carried out on PlanetLab. Many of these projects concern cyber security.

5.2 X-Bone

X-Bone [24] software is another way to create and configure overlay testbeds. It installs routes, configures interfaces, updates DNS entries and installs IPSEC keys. X-bone works by using IP within IP packets to create the overlay, and it uses multicast packets for resource discovery. It supports multiple concurrent overlays and provides data security through the use of IPSEC. An extension of X-Bone is DynaBone which is a system of overlays to resist DDoS attacks.

5.3 Overlay Discussion

A review [4] of network testbed software concluded that PlanetLab provides the most sophisticated software and documentation in its class.

Testing cyber attacks on an overlay testbed may be considered risky given the minimal separation between the test network and the underlying live network. Hence, this may not be the ideal environment for some CR activities.

6. Emulation

The above overlay network testbeds use a live network as their basis. In contrast, this section discusses emulation testbeds which use a standalone physical testbed to emulate any number of different experiment configurations. Emulation is the process of mapping a desired experimental network topology and software configuration onto the physical infrastructure of the CR. While the actual infrastructure may consist of a cluster of machines, the emulation component can configure the cluster to behave as per the desired experiment topology including routing and WAN links.

Therefore, emulation allows the same CR to be used for many different purposes and experiments. It supports experimentation with scientific rigour. Fidelity is high due to the use of real hardware and having total control of background traffic, workloads and general events. This contrasts with simulations where it is not always easy to know which parts of the model can be abstracted without influencing the test results. Fidelity is also higher than with overlay networks where experiments are not always repeatable due to the shared infrastructure. Emulations can also be used to verify the accuracy of simulators and if there are inaccuracies, to investigate why. Virtualisation is commonly used as a tool for flexible emulation. However, the more virtual machines placed on a single physical host, the lower the fidelity of the experiment due to contention for resources. A reasonable physical to virtual ratio is 1:10 to balance scalability of the testbed with fidelity. A disadvantage of an emulation-style CR is the requirement to setup and maintain dedicated computer and network resources. This problem is partially mitigated through virtualisation which allows a smaller physical network to emulate a much larger network. Emulation-style CRs also require software to configure the CR infrastructure to a different emulated topology for each experiment.

Emulation appears to be the most widely used approach in modern CRs. Prominent examples found in the literature and discussed below are Emulab, DETER, the NCR and the JIOR.

6.1 Military and Government

The US military and government is investing predominantly in emulation type CRs. The National Cyber Range alone (which emulates the public internet and other infrastructure) was estimated to cost US$130 Million to build [59].

6.1.1 NCR

Perhaps the most ambitious CR is the DARPA National Cyber Range (NCR). This project was announced in 2008, and is still being built and assessed by contractors. It aims to simulate cyber attacks on computer networks to help develop defensive strategies. The NCR is planned to be built on a very large scale to emulate the complexity of defence and commercial networks. This should allow new cyber technologies to be tested and validated in a representative environment. However, since it is a large project taking many years to implement, other government and military agencies built their own CRs in the meantime, for example the USAF's Cyber-safari range [25, 26]. The ambitious requirements of the NCR include: GiG scale, simple experiment design tools, automated range build to match experiment, real-time data visualization tools, sanitisation and simultaneous testing at multiple security levels.

One of the main purposes of CRs is to test security appliances in a representative environment (i.e. the CR emulates the projected operational environment). It therefore fills the testing gap between black-box testing and in-situ operational testing. Lockheed Martin is developing software for the NCR called a Flexible, Automated Cyber Technology Range (FACTR) [27]. The software consists of automated tools to construct and validate tests quickly and ensures the scientific method is used. Currently the prototype NCR is in a one-year beta operation phase.

6.1.2 JIOR

The US JFCOM Joint Information Operations Range (JIOR) is described [28, 29] as a realistic environment to practice tactics, techniques and procedures for IO. The JIOR infrastructure is capable of supporting CNO exercises, cyber testing and training. Multiple sites can participate via encrypted links. The JIOR can create a number of realistic environments by combining traffic generators, CNO labs, computing infrastructure, telecommunications equipment, EW platforms, threat systems and red teams, communications systems, SCADA systems and other models and simulations. While the JIOR has broader scope than a CR it is still an important facility for training CNO staff. The main purpose for an IO range is presented in [28] as testing combat effectiveness. This includes testing survivability under IO attack. To discover issues as early as possible, testing is performed at all stages of the development cycle. Initial tests can be paper-based, followed by lab testing and then building up to a distributed IO range which offers the most realistic test environment. A 2006 paper [30] about the JIOR discusses a planned path from JIOR inception to full spectrum IO support. The author provides a long list of requirements for the JIOR including experimentation, training, development of tactics, techniques and procedures, battle damage assessment and also mission rehearsal. Many of these requirements are equally applicable to a CR for CNO staff.

6.1.3 INL

Idaho National Laboratory (INL) performs assessments of vendor Industrial Control Systems to enhance critical infrastructure protection. They now support two main US industrial control cyber security programs, with one including a large CR [31]:

- DHS Control Systems Security Program (CSS): raises awareness in the international community about threats facing critical infrastructure.
- DOE National SCADA Test Bed (NSTB): was established to help equipment vendors assess and verify the security of their devices in a full-scale test environment. The test bed includes "17 testing and research facilities, encompassing field-scale control systems, 61 miles of 138 kV transmission lines, seven substations, and state-of-the-art visualization and modelling tools"[32]. This

is a testbed of real devices. While this has the advantage of being a representative environment, it lacks flexibility for setting up different types of experiments.

Both programs offer training in the form of a week long red team versus blue team exercise.

6.1.4 Military Academy CRs

The remaining military CRs in this review focus on training staff. The Information Warfare Analysis and Research laboratory (USMA IWAR) is a US military academy isolated network used for training cadets in information operations. The network requires extensive effort to setup and maintain. It can simulate defences such as cryptography, encryption and access control methods, and simulate attacks including trojans, vulnerability scanners, viruses, worms, DoS and password cracking. Its configuration must be customised to support the aims of each exercise.

The Royal Military College of Canada Computer Security Laboratory (RMC CSL) also operates an isolated network for CNO education. It uses virtualisation to allow multiple guest operating systems to run on each physical host. However it is still human-resource intensive: it requires 1 full-time staff to maintain seven physical hosts, has about 10 trainees defending approximately 20 guest operating systems, about 5 attack operators and 3 controller staff.

These standalone private network setups have been widely used despite their lack of automated simulation software and high maintenance costs, for example they have been used for many years in the Cyber Defence Exercise (CDX) [3]. This is an annual four day exercise run by the NSA to train students in CNO. It has been running since 2001 with each participating organisation (e.g. US military academies) supplying a blue team. The blue teams were required to defend their infrastructure while under attack from the red team, and being observed by the white team. Generally the white team develops scenarios and anomalies, establishes scoring criteria and referees exercises. The exercise involved the blue teams receiving misconfigured hardware and software, and then trying to fix vulnerabilities and add layers of security to mitigate any threats. The blue teams also had to maintain services such as a mailserver, webserver, database, instant messenger and a domain controller. This hands-on exercise has proven very effective for teaching computer security. In 2012 and other years, Lockheed Martin hosted the CDX headquarters using a private network at its Hanover facility, with each college connecting via VPN for the exercise.

Smaller training exercises, which do not require a large dedicated CR but instead simply use VMs supplied to each student, include CANVAS and DefEX. CANVAS is an exercise organised by the Academy Centre for Cyberspace Research (ACCR) for US cadets [33]. This exercise has been running since 2006 and is a one day training course in penetration testing. Students get to "pentest" a simulated system such as an e-voting machine or a social networking site. The students then prepare a report on the "pentest" for grading. ACCR publically publishes old versions of this training covering network traffic forensics, password usage and cracking, encryption and even the VMs used in the CANVAS exercise. This public repository [34] includes software in the form of applets and standalone applications. Similarly, DefEX was a cyber security exercise designed by AFRL for undergraduate students across academic institutions in USA [35]. It consisted of hands-on exercises for code and system hardening, reverse engineering of malware, intrusion detection and prevention, digital forensics and a wireless treasure hunt. The exercises required virtual machines and access to computer security software. The exercise focussed on education, so included both introductory lectures and tasks which were led by an instructor.

6.2 Academic

Two well-known emulation facilities and software developed in academia are Emulab and DETER.

6.2.1 Emulab

Emulab [36], originally from the University of Utah, refers to both the facility at the university and to the open source emulation software for testbeds. The software is used in over twenty other emulation testbeds worldwide. It is mainly used by researchers in the fields of networking and distributed systems. The facility is available free of charge for experiments by researchers worldwide. Researchers can also download the software to create their own private Emulab.

Emulab can be used to conduct scientifically rigorous experiments with high fidelity, repeatability and measurement accuracy. Emulab makes use of a number of tools including Tmix-ns2 for generating realistic background network traffic, tcpreplay for replaying packet captures and DummyNet for simulating network links with different characteristics. It also automates experiment setup and teardown including installation of operating systems, reservation of resources and creating network topologies using programmable switches. A 2010 review [4] of testbed software found Emulab to be the most sophisticated publicly available software in its class. A number of projects listed on the Emulab website involve security research, testing or training showing that Emulab has been used as a basis for CR functionality. DETER (a derivative of Emulab) seems even more suitable as a CR.

6.2.2 DETER

The DETER testbed is a public facility for medium scale national experimentation in cyber security. It is therefore more focussed on CR capabilities than the general purpose Emulab. The DETER project [37] operates DeterLab which is open to researchers worldwide. It is funded by the Department of Homeland Security, the National Science Foundation and the Department of Defense. It is run by a research team at USC ISI and U.C. Berkeley and is also supported by more than 20 collaborators. Since 2003, the project community has grown to over 2000 people.

This community has developed a library of tools which are publicly available from its website. The tools include models and software for traffic and network topology generation and result analysis. The DETER testbed supports remote access for experimenters while also providing experiment isolation and containment. The environment enables experiments in cyber defence technology and aims to accelerate the field of cyber security research through sharing, re-use and repeatability. The software is based on Emulab with the main difference being that DETER is customised for cyber security. Their website contains a list of projects using DETER. The list includes a number of experiments, university modules and training in the field of computer network defence.

The project website explains some benefits of running experiments in their lab: “In practice in DeterLab, DETER enables experimenters to share data, lab set-up, software and tools, experimental procedures, results, and other information that should enable new experimenters to stand on their predecessors’ shoulders, rather than start at ground level for every new project.” [37]

DETER started in 2004 when the motivation for building it was to create a large scale security testbed to overcome deficiencies in evaluating cyber defences. These deficiencies included a lack of scientific rigour, the lack of representative network traffic, and inadequate model of networks, attacks and defences [38], and as a result innovative tools failed in real deployments due to the huge jump from small scale testing to the real world. The planned functionality of DETER to overcome these deficiencies were attack scenarios, attack simulators, background traffic generators, datasets derived from live traffic and tools to monitor and summarise the tests.

A 2011 report [39] on DETER after seven years of operation listed a number of lessons learnt. These included:

- Experiment construction was considered difficult: setting up and configuring a test required a lot of system administration. Generally there was a lack of experiment abstraction and reuse. Hence effort was put into building an experimenter’s workbench. The first generation was called SEER, and the second ELM.
- Federation was useful: to extend the functionality of DeterLab, the project enabled controlled remote access to other specialised testbeds and facilities such as PlanetLab, GENI, SCADA systems and supercomputers.
- Experiment isolation was considered limiting: the original requirement for DETER was rigid segregation from the internet. However this was later relaxed to “controlled internet access” for experiments where it would be a huge effort to replicate online services in a standalone network with sufficient fidelity. This includes experiments on botnets or with a privacy network such as Tor.
- Requirement for flexibility in experiment scale and fidelity: the approach was to allow the experimenter to select high fidelity for components of interest, and low fidelity for the rest of the test. Low fidelity generally means high scalability since the component is only simulated. This allows very large networks to be modelled on limited hardware.

- A requirement for improved tools and methods to mine experimental data for results.

DeterLab uses the Experiment Lifecycle Manager (ELM) as the experimenter's workbench. It aims to offer more abstraction and reuse, such as saving the configuration of a host including its network settings as an object which can be reused. It also supports the abstract test definitions to be automatically deployed to the testbed. The software is built on the eclipse IDE. DETER now falls under the umbrella of GENI. However it is not yet integrated to accept single-sign-on GENI credentials.

6.2.3 Virtualised CR

Another academic project which built an emulation-based CR is described in [40]. The project was built on Sandia National Laboratories' Thunderbird cluster. Its aim was to understand complexity in cyber systems due to the large scale of networks (e.g. a million nodes). Hence the researchers modelled a million virtualised hosts on a computing cluster. The particular scenario they investigated was botnet behaviour, where they were trying to model and predict current and future behaviours. To ensure accurate results, they ran botnet software directly rather than creating a simple simulation. The main contribution of their paper was to demonstrate their ability run a million stripped linux nodes on a cluster of computers. While this CR was setup for a single experiment it could be adapted to other purposes.

6.2.4 Reassure

A testbed which makes use of virtualisation is Reassure from Purdue University [41]. The university hosts a testbed with its own hardware, but also make its software available publicly so anyone can setup their own private Reassure testbed. The software includes an experiment manager, a network GUI tool for creating a topology and virtual machine image creation. One of the project goals is to provide a software archive for vulnerability research and analysis.

Different types of virtualisation can be used to make a flexible and cost effective testbed [42]. Building a large realistic lab is costly, requires employee time to maintain and needs space allocated in a data centre. A lot of these issues can be reduced by using virtualisation. It allows easier management and repeatability by saving state with snapshots. Different types of virtualisation can also be used to balance between high scalability with OS virtualisation (e.g. OpenVZ) and more accurate emulation with paravirtualisation (e.g. Xen or VMWare ESXi). The authors propose taking a hybrid approach for flexibility – OpenVZ for up to one hundred nodes per GB of memory, and Qemu for full emulation. The virtualisation manager "libvert" supports both.

6.3 Commercial

It does not appear that private companies sell emulation CRs as a standard product. Instead they either build it as a project for each customer, or host the CR themselves and provide remote access.

6.3.1 Northrop Grumman (including in Australia)

A small number of commercial emulation-style CRs were found during this survey. One example is from Northrop Grumman [43] which provides a dedicated cyber test range with capabilities they have developed since 1999. Recently they have used Breaking Point Storm CTM appliances in their Federated CR to generate high-volume benign and malicious computer network traffic. In 2012 they were also selected to build a cyber test range for the University of New South Wales, Canberra Campus at the Australian Defence Force Academy (ADFA) in Australia [64]. This cyber range will be used by the Australian military to develop, test and evaluate cyber technologies.

6.3.2 Counter Hack Challenges

Another commercial offering is Counter Hack Challenges [44]. The company offers capture-the-flag and quiz-oriented exercises for the US Cyber Challenge, the SANS Institute NetWars [45] program and others. Interactive training is accessible from a web browser and the CR appears to be hosted by the company. The exercises can be done intensively in a 2 or 3 day tournament or stretched out over an extended period remotely. Participants capture flags to indicate understanding of topics. Hints are provided to assist in learning, although the number of hints used is recorded. Individual scores are then added to a scoreboard. Areas covered include vulnerability assessments, system hardening, malware analysis, digital forensics, incident response, packet analysis and penetration testing.

6.3.3 Detica

A commercial cyber security training range course is also offered by Detica [46], a subsidiary of BAE Systems. The Cyber Academy training offerings include 1 or 2 days courses in web app security, pentesting, wireless security and public key encryption. The academy is underpinned by a purpose built cyber training range.

6.3.4 ATC

Architecture Technology Corporation (ATC) offers a commercial, virtualised training platform for network defence and computer forensics called Cyber Defence Trainer (CYDEST) [47]. It provides automated training and assessment with live scenarios. Rather than using simulation, CYDEST makes use of real software on virtualised hosts. For training, three networks are setup:

- Target network: this contains the hosts for which the trainee is responsible

- Attack network: simulates internet traffic and contains both benign and malicious data
- Control network: for permanent infrastructure independent of the training scenario.

Students are assessed using both active monitoring, such as whether they have accessed a key file, and passive monitoring where the student records actions in an online notebook visible to the assessor.

7. Capture the Flag (CTF) Competitions

Many cyber CTF competitions are held around the world. They range from puzzle-like challenges to team-based offensive and defensive hacking competitions running over a day or two. Teams or individuals are scored on their performance to create a leader board during the competition. An extensive list of these competitions is currently maintained by forgottensec [48]. This paper does not review CTF competitions in detail, but since they are a good tool for training and practicing CNO, they have some overlap with CRs. Several websites make past CTF scenarios (including VMs) publicly available. These are useful for training. Examples of current CTF competitions include:

- DefCon [60]: is a yearly hacker convention. It runs a CTF competition which is in the form of team on team, with qualification round to enter. This year 20 teams qualified. It has been running since 1993.
- International Capture the Flag (iCTF) [61]: is organised by the University of California, Santa Barbara. It is a very large academic CTF running since 2004. The 2011 competition involved 89 teams from across the world. Historically it has been a team on team competition, but now each team runs an identical parallel version of the game.
- Collegiate Cyber Defence Competition (CCDC) [62]: is an annual college-level competition. It differs from other competitions in that it focuses on defending your own network from outside attack.
- Cyber Defence Exercise (CDX) [3]: is an annual competition limited to military academy participants. See Section 6.1.
- MIT/Lincoln Labs CTF [49]. While all CTFs test the skills of participants, this competition has a strong education component. Participants attend five preparatory lectures covering necessary skills, for example fundamentals of web application security, linux OS server security and web application common exploits. In 2011 scoring was based on the availability of each team’s web services, and on the number of flags captured. The infrastructure was VMWare ESX servers, with each team getting a pre-configured VM for them to defend. This allowed teams to revert to known good snapshots when they were aware of being compromised.

8. CR Review Discussion

8.1 Simulation, Overlay and Emulation

This review has primarily found CRs to use either simulation or emulation. Pure simulation and full hardware emulation appear to be two extremes on a continuum of approaches. Hardware emulation excels in realism and fidelity of results since they run the real software on real hardware, while simulations have flexibility and scalability advantages. However, the middle ground is increasingly being explored. CRs which support simulation and emulation, and include support for multiple types of virtualisation could combine the best aspects of each approach [50]. An example is the emulation-type CR called Emulab, which can run on a large laboratory of hardware, but which uses simulation software to create network links with different characteristics and to create background network traffic. DETER is a cyber-security focussed derivative of Emulab and so also uses the mixed approach. LARIAT was categorised as a simulation because it simulates user behaviour, however it can run on and interact with various hardware to support information assurance testing. It is a combination of simulated and real networks.

One of the goals of a CR is to provide repeatable testing of new cyber security products before deployment into the field. For accurate testing, the CR should provide an environment representative of where it would be used. However, this goal is not easily achievable due to the huge diversity of hardware products used in real networks. Installing and maintaining each type of hardware device in a CR would be extremely expensive. Another option is to model each device in a simulation. However, as discussed in Section 4.4, simulation results do not always transfer to the real world due to simplifying assumptions made in modelling. This problem is most pronounced for load-critical attacks such as DoS [51]. To explore how different testbeds perform, the authors performed low-rate TCP DoS attack on a network simulated with ns-2 [63] and repeated the tests on networks emulated with DETER and Emulab. They found key differences in the results. The differences were due to emulation using real software and hardware with finite capabilities resulting in bottlenecks which the simulators do not model. However, the results between DETER and Emulab were also different. The conclusion was that attacks which overload the system can have very different outcomes depending on the exact hardware, software and configurations used. Other researchers came to a similar conclusion [52]. They recommended using emulation when studying DoS attacks due to emulation generally achieving higher fidelity results than simulation. Since emulation CRs use real hardware and software, they can recreate realistic and complex network behaviour necessary for security and resilience testing. Also, since emulation generally underperforms real router hardware, the damage from DoS attacks is overestimated. They consider reporting worst-case scenario results to be a fair solution. In summary, due to hardware and other implementation differences, the quantitative test results from an emulation CR may not always transfer exactly to the real world, but qualitative results should.

While some papers in this review indicated that simulations are hard to validate against real attacks, this has not stopped interactive simulations being successfully used for training. The simulations are often aimed at determining large-scale effects and so the details of the attacks do not necessarily matter. Advantages of using simulation in training include simple pause and replay actions within a training session, re-use of the scenario and low infrastructure costs.

Overlay test networks were also briefly discussed. These appear to be used for experiments involving global-scale applications where the network links are integral to it. These WAN experiments are difficult to simulate accurately in a standalone testbed. PlanetLab was the best example of an overlay testbed, with experiments including network storage and peer-to-peer communications. The large European project called GENI uses ideas from Planetlab, and consists of many federated sites (overlayed on the internet) to achieve its wide-area shared platform for experimentation.

In our view, the most promising type of CR is emulation. The review covered several emulation-type CRs including Emulab and DETER from academia, and the NCR and JIOR from the military. When built at scale, these CRs can provide a representative environment critical for accurate testing of cyber products, for training CNO staff and for performing cyber security research. The use of dedicated hardware in these CRs enables high fidelity experiments to be performed. Recent advances in virtualisation enables easier configuration and also enables improved scalability (by enabling multiple virtual hosts to run on a single computer).

While beyond the scope of this review, some specialised ranges were also mentioned. These included the NSTB for SCADA networks, the JIOR for full spectrum IO testing and OneLab for wireless research. Each of these testbeds requires significant expertise and resources to run. For efficiency, some of these resources are shared between collaborators, rather than each site investing in their own hardware. This sharing is achieved by federating testbeds, generally using VPN links over the internet.

8.2 CR Roles

Three main roles of a CR were identified. The first role is testing. CRs are useful for project support such as evaluating the effectiveness, reliability and fitness for purpose of a new device. The second role is training. CRs assist operations by providing a training platform for CNO staff and allowing tactics techniques and procedures to be developed plus war gaming. The third main role is research and development since CRs provide a safe environment for scientific experiments to be performed.

CRs more focussed on testing and research roles include DETER, Emulab, the NCR, the NSTB, StealthNet, LARIAT and Breaking Point. These CRs support repeatable experimentation and testing.

The most popular role for the reviewed CRs was training. That is, they aimed to assist operations by providing a platform for training staff. Examples included the USMA IWAR and RMC CSL laboratories which provided a training platform for use in exercises such as the CDX. Training is also a major objective in the JIOR and in simulations such as SIMTEX, SAST and CAAJED. Additional exercises which provide training opportunities include DefEX, CANVAS, Cyber Flag, the Counter Hack Challenge and numerous capture-the-flag competitions.

8.3 CR Availability

One of the aims of the review was to identify cost-effective ways to obtain a CR capability. The review found the public CRs DeterLab and Emulab to be worth investigating. DeterLab offers cyber security researchers a shared CR which they can login to remotely and run experiments. The CR software is publicly available including the SEER management software. Emulab also offers a CR facility, plus makes their entire CR software available publicly so researchers can setup their own private Emulab instance. Since they are supported by large communities of researchers, the software is maintained and will likely expand in functionality. The software already supports a number of important tasks such as deploying experiments to the CR, monitoring and analysis, and user access controls. Also included is a repository of experiment setups. These can be used to create training scenarios. However, prebuilt training scenarios may be more easily sourced from exercises such as CANVAS.

The availability of commercial CRs allows much of the setup and maintenance of the CR to be outsourced. This may be a cost-effective approach.

It was found from the literature that the US military has already invested significantly in CRs. They have been involved in the development of the NCR and JIOR and have been running cyber exercises for at least 10 years. Collaborating with the US military is another possible approach to obtaining a CR capability.

Researchers may prefer to join collaborative initiatives such as FIRE in Europe and GENI in the US. These provide a pool of testbed resources to assist in sharing and reuse among researchers. Federated testbeds such as these are likely to generate a larger community of support than standalone testbeds and hence are more likely to succeed. However, Defence organisations may still require their own trusted CR to ensure privacy when developing and practicing CNO techniques.

9. Conclusion

This review paper discussed a number of published Cyber Ranges (CRs). It found the majority of CRs use either simulation or emulation. Emulation CRs use testbeds with real hardware and software. These can be configured using emulation software to different network topologies for each experiment. The main advantages of emulation CRs are the ability to create a representative environment for training and testing, and the ability to perform high fidelity and repeatable experiments. However a disadvantage is their high cost due to the large infrastructure requirements (e.g. NCR). This cost can be reduced through virtualisation or resource sharing.

On the other hand, simulation CRs make use of software models of real world objects. This allows large simulations to be run on relatively modest hardware. That is, simulation CRs are highly scalable, flexible and low cost. A criticism of some simulations is it is hard to verify they accurately reflect reality. Despite this, simulations have been successfully used in training of CNO staff.

It was also found that CRs are increasingly using emulation, perhaps due to the decreasing cost of hardware and the rise of virtualisation which mitigates cost, scalability and reconfiguration issues.

10. References

[1] W. F. Lynn III, "Defending a new domain: the Pentagon's Cyberstrategy," 2010 DTIC Document.

[2] "Broad Agency Announcement (BAA) National Cyber Range, https://www.fbo.gov/download/c33/c330660f00c9820d05c9f4c54422024b/080505_BAA_National_Cyber_Range_Final.doc," DARPA, 2008.

[3] S. P. Leblanc, A. Partington, I. Chapman and l. Bernier, "An overview of cyber attack and computer network operations simulation," in *Proceedings of the 2011 Military Modeling & Simulation Symposium* Boston, Massachusetts: Society for Computer Simulation International.

[4] C. Siaterlis and M. Masera, "A survey of software tools for the creation of networked testbeds," *International Journal On Advances in Security,* vol. 3, 2010.

[5] "SANS Announces the US Air Force Wins U.S. National Cybersecurity Innovation Award," 2011, http://www.sans.org/press/air-force-cyber-security-award.php, viewed 21/09/2012.

[6] "GENI: Exploring Networks of the Future," 2012, http://www.geni.net, viewed 21/09/2012.

[7] "Future Internet Research and Experimentation (FIRE)," http://www.ict-fire.eu/home.html, viewed 21/09/2012.

[8] S. W. Neville and K. F. Li, "The Rational for Developing Larger-scale 1000+ Machine Emulation-Based Research Test Beds," *WAINA 2009*, pp. 1092-1099.

[9] A. P. Hansen, "Cyber Flag: A Realistic Cyberspace Training Construct," DTIC Document 2008.

[10] L. Stoeckmann, "AFNIC supports U.S. Cyber Command's first major tactical cyber exercise," 2011, http://www.afnic.af.mil/news/story.asp?id=123283198, viewed 21/09/2012.

[11] J. Michael G. Wabiszewski, T. R. Andel, B. E. Mullins, and R. W. Thomas, "Enhancing realistic hands-on network training in a virtual environment," in *Proceedings of the 2009 Spring Simulation Multiconference* San Diego, California: Society for Computer Simulation International, 2009.

[12] J. L. Winner, L. S. Holt, J. Duran, and E. Watz, "Cyber Operations Virtual Environment," DTIC Document 2010.

[13] R. S. Mudge and S. Lingley, "Cyber and air joint effects demonstration (caajed)," DTIC Document 2008.

[14] M. Varshney, K. Pickett, and R. Bagrodia, "A Live-Virtual-Constructive (LVC) framework for cyber operations test, evaluation and training," *MILCOM 2011*, pp. 1387-1392.

[15] C. Hoffpauir, "USJFCOM Gets Approval to Connect U.S., Australian Networks," 2006, http://www.army.mil/article/1065/usjfcom-gets-approval-to-connect-us-australian-networks/, viewed 21/09/2012.

[16] F. Cohen, "Simulating cyber attacks, defences, and consequences," *Computers & Security,* vol. 18, pp. 479-518, 1999.

[17] M. Liljenstam, J. Liu, D. M. Nicol, Y. Yuan, G. Yan, and C. Grier, "Rinse: The real-time immersive network simulation environment for network security exercises (extended version)," *Simulation,* vol. 82, pp. 43-59, 2006.

[18] M. E. Kuhl, J. Kistner, K. Costantini, and M. Sudit, "Cyber attack modelling and simulation for network security analysis”, *WSC 2007*, pp 1180-1188.

[19] L. M. Rossey, R. K. Cunningham, D. J. Fried, J. C. Rabek, R. P. Lippmann, J. W. Haines, and M. A. Zissman, "Lariat: Lincoln adaptable real-time information assurance testbed," *Aerospace Conference Proceedings* 2002, pp. 6-2671-2676, 6-2678-6-2682 vol. 6.

[20] P. Pederson, D. Lee, G. Shu, D. Chen, Z. Liu, N. Li, and L. Sang, "Virtual Cyber-Security Testing Capability for Large Scale Distributed Information Infrastructure Protection," *Technologies for Homeland Security*, 2008, pp. 372-377.

[21] "Breaking Point Cyber Range Deployment," http://www.breakingpointsystems.com/solutions/cyber-range-deployment/, viewed 21/09/2012.

[22] "Scalable Network Technologies," http://www.scalable-networks.com/content/, viewed 21/09/2012.

[23] "PlanetLab," 2012, https://www.planet-lab.org/, viewed 21/09/2012.

[24] "X-Bone," http://www.isi.edu/xbone/, viewed 21/09/2012.

[25] D. A. Fulgham, "Cyber-Warriors say DARPA is Too Slow," Aviation Week, 2010, http://www.aviationweek.com/Blogs.aspx?plckBlogId=Blog%3a27ec4a53-dcc8-42d0-bd3a-01329aef79a7&plckPostId=Blog%3a27ec4a53-dcc8-42d0-bd3a-01329aef79a7Post%3ae06794a1-1a22-468b-bd8a-00149537b0f3, viewed 21/09/2012.

[26] N. Shachtman, "DARPA taking fire for its Cyberwar Range," Wired, 2010, http://www.wired.com/dangerroom/2010/06/darpa-taking-fire-for-its-cyberwar-range/, viewed 21/09/2012.

[27] L. Pridmore, P. Lardieri, and R. Hollister, "National Cyber Range (NCR) automated test tools: Implications and application to network-centric support tools," *Autotestcon*, IEEE, pp. 1-4.

[28] J. Payne, "Combat Effectiveness Testing of System of Systems in the Face of Cyber Threat",2010, http://www.dtic.mil/ndia/2010test/WednesdaySessionLJonathonPayne.pdf.

[29] E. Hernandez, "Information Operations Range Test Week 2010 Mission Review," 2010, https://acc.dau.mil/adl/en-US/386254/file/52142/01.20100521%20Test%20Week%2010%20IOR%20Review.pdf, viewed 21/09/2012 .

[30] R. P. Sabo, "Standing Up the Information Operations Range," 2006, http://www.au.af.mil/info-ops/iosphere/06fall/iosphere_fall06_sabo.pdf .

[31] R. S. Anderson, "Cyber Security and Resilient Systems," Idaho National Laboratory (INL) 2009.

[32] "National SCADA Test Bed (NSTB)," 2009, http://energy.gov/sites/prod/files/oeprod/DocumentsandMedia/NSTB_Fact_Sheet_FINAL_09-16-09.pdf.

[33] "Academy Centre for Cyberspace Research (ACCR)", http://www.usafa.edu/df/dfe/dfer/centers/accr/, viewed 21/09/2012.

[34] "Public Repository for Information Security Material (PRISM)," http://www.prismhome.org/category/resource-type/activities-and-labs?page=1, viewed 21/09/2012.

[35] S. M. Glumich and B. A. Kropa, "DefEX: Hands-On Cyber Defense Exercise for Undergraduate Students," DTIC Document, 2011.

[36] C. Siaterlis, A. Garcia, and B. Genge, "On the Use of Emulab Testbeds for Scientifically Rigorous Experiments", *Communications Surveys and Tutorials*, IEEE, issue 99, pp 1-14.

[37] "The DETER Project." 2012, http://deter-project.org/, viewed 21/09/2012.

[38] R. Bajcsy, T. Benzel, M. Bishop, B. Braden, C. Brodley, S. Fahmy, S. Floyd, W. Hardaker, A. Joseph, and G. Kesidis, "Cyber defense technology networking and evaluation," *Communications of the ACM,* vol. 47, pp. 58-61, 2004.

[39] T. Benzel, "The science of cyber security experimentation: the DETER project," ACSAC 2011, pp. 137-148.

[40] J. R. Mayo, R. G. Minnich, R. C. Armstrong, and D. W. Rudish, "Approaches for scalable modeling and emulation of cyber systems : LDRD final report," SAND2009-6068, 2009.

[41] "Reassure: A safe virtual imaging instrument for logically destructive experiments," http://projects.cerias.purdue.edu/reassure/, viewed 21/09/2012.

[42] K. E. Stewart, J. W. Humphries, and T. R. Andel, "Developing a virtualisation platform for courses in networking, systems administration and cyber security education," *Spring Simulation Multiconference* 2009, p. 65.

[43] "Northrop Grumman Cyber Test Range," http://www.is.northropgrumman.com/products/cyber_test_range/index.html, viewed 21/09/2012.

[44] "Counter Hack Challenges," https://www.counterhackchallenges.com/ viewed 21/09/2012.

[45] "SANS NetWars Cyber Range," http://www.sans.org/cyber-ranges/netwars/, viewed 21/09/2012.

[46] "The Stratsec Cyber Academy," http://www.stratsec.net/Services/Other-Services/The-stratsec-Cyber-Academy, viewed 21/09/2012.

[47] S. Brueckner, D. Guaspari, F. Adelstein, and J. Weeks, "Automated computer forensics training in a virtualised environment," *digital investigation,* vol. 5, pp. S105-S111, 2008.

[48] "Forgottensec CTF and Information Security competition information," http://ctf.forgottensec.com/wiki/index.php?title=Main_Page, viewed 21/09/2012

[49] J. Werther, M. Zhivich, T. Leek, and N. Zeldovich, "Experiences in cyber security education: The mit lincoln laboratory capture-the-flag exercise," *Cyber Security Experimentation And Test,* vol. 8.

[50] W. A. Fagen, J. W. Cangussu, and R. Dantu, "A virtual environment for network testing," *Journal of Network and Computer Applications,* vol. 32, pp. 184-214, 2009.

[51] R. Chertov, S. Fahmy, and N. B. Shroff, "Emulation versus simulation: A case study of TCP-targeted denial of service attacks," *TRIDENTCOM* 2006, pp. 10

[52] J. Mirkovic, S. Fahmy, P. Reiher, and R. K. Thomas, "How to test DoS defenses," *Conference for Homeland Security* CATCH 2009, pp. 103-117.

[53] Brown, Bill, et al. "Simulation of cyber attacks with applications in homeland defense training." AeroSense 2003. International Society for Optics and Photonics, 2003.

[54] Zhou, Mian, and Sheau-Dong Lang. "A Frequency-based approach to intrusion detection." Proc. of the Workshop on Network Security Threats and Countermeasures. 2003.

[55] "Northrop Grumman Brings Cyber Range to Government and Enterprise Customers", http://blogs.ixiacom.com/ixia-blog/northrop-grumman-brings-cyber-range-to-government-and-enterprise-customers/, viewed 24/10/2013

[56] "Testing Next Generation Data Center Security", http://blogs.ixiacom.com/ixia-blog/testing-next-generation-data-center-security/, viewed 24/10/2013

[57] "OMNeT++ Network Simulation Framework", http://www.omnetpp.org/, viewed 24/10/2013

[58] "Candela Technologies Network Testing and Emulation Solutions", http://www.candelatech.com/, viewed 24/10/2013

[59] "Cyberwar heats up with Pentagon's virtual firing range", http://www.theguardian.com/technology/2011/jun/17/pentagon-virtual-firing-range, viewed 24/10/2013

[60] "DEF CON Hacking Conference", http://www.defcon.org/, viewed 24/10/2013

[61] "The UCSB iCTF", http://ictf.cs.ucsb.edu/, viewed 24/10/2013

[62] "National Collegiate Cyber Defence Competition" http://www.nationalccdc.org/, viewed 24/10/2013

[63] "The Network Simulator ns-2" http://www.isi.edu/nsnam/ns/, viewed 24/10/2013

[64] "Northrop Grumman to build Cyber Test Range in Australia", http://www.securityweek.com/northrop-grumman-build-cyber-test-range-australia, viewed 24/10/2013

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION DOCUMENT CONTROL DATA

1. DLM/CAVEAT (OF DOCUMENT)

2. TITLE A Survey of Cyber Ranges and Testbeds	3. SECURITY CLASSIFICATION (FOR UNCLASSIFIED REPORTS THAT ARE LIMITED RELEASE USE (L) NEXT TO DOCUMENT CLASSIFICATION) Document (U) Title (U) Abstract (U)
4. AUTHOR(S) Jon Davis and Shane Magrath	5. CORPORATE AUTHOR DSTO Defence Science and Technology Organisation PO Box 1500 Edinburgh South Australia 5111 Australia

6a. DSTO NUMBER DSTO-GD-0771	6b. AR NUMBER AR-015-770	6c. TYPE OF REPORT General Document	7. DOCUMENT DATE October 2013

8. FILE NUMBER 2012-1220994/1	9. TASK NUMBER 07/348	10. TASK SPONSOR CIOG	11. NO. OF PAGES 29	12. NO. OF REFERENCES 64

13. DSTO Publications Repository http://dspace.dsto.defence.gov.au/dspace/	14. RELEASE AUTHORITY Chief, Cyber and Electronic Warfare Division

15. SECONDARY RELEASE STATEMENT OF THIS DOCUMENT

Approved for public release

OVERSEAS ENQUIRIES OUTSIDE STATED LIMITATIONS SHOULD BE REFERRED THROUGH DOCUMENT EXCHANGE, PO BOX 1500, EDINBURGH, SA 5111

16. DELIBERATE ANNOUNCEMENT

No Limitations

17. CITATION IN OTHER DOCUMENTS Yes

18. DSTO RESEARCH LIBRARY THESAURUS

Cyber range, testbed, simulation, overlay, emulation

19. ABSTRACT

This document reviews the state-of-the-art in Cyber Range implementations and related computer network operations (CNO) testbeds. We summarise recently published examples and describe their purpose and functionality. The compiled information should assist organisations to make an informed decision when considering a Cyber Range capability.