ReSCIND Performer HSR Dataset

Cover Sheet – On-Range Cyber CogVuln Study

ReSCIND

Dataset Details

Dataset Title:	On-Range Cyber CogVuln Study (PsyCCDef)		
Dataset Citation:	Gonzalez, C., Aggarwal, P., Rajivan, P., Venkatesan, S., Aggarwal, A., & Ferreira, M. "ReSCIND-PsyCCDef-On-Range Cyber CogVuln Study Dataset". https://osf.io/834at/overview		
Data Format:	Mixed types	Data Size:	60GB
Dates & Duration:	July 2025 – August 2025	Time Zone:	Multiple
How to access dataset:	Email: Dr. Cleotilde Gonzalez (coty@cmu.edu), Dr. Palvi Aggarwal (paggarwal@utep.edu), Dr. Prashanth Rajivan (prajivan@uw.edu)		
Point of Contact for data questions:	Dr. Cleotilde Gonzalez (coty@cmu.edu), Dr. Palvi Aggarwal (paggarwal@utep.edu), Dr. Prashanth Rajivan (prajivan@uw.edu), Dr. Sridhar Venkatesan (svenkatesan@peratonlabs.com)		

Description of Scenario

Experiment Objectives

This experiment was designed to study the effects of cognitive vulnerabilities (CogVulns) and bias triggers within cyber contexts identified in the previous Qualtrics-based studies (i.e., General Cyber CogVuln Study and Situated Cyber CogVuln Study) on a realistic cyber range.

Experiment Description

The experiment was designed as a Capture-the-Flag (CTF) type of exercise where participants collect flags when they complete cyber tasks in a typical enterprise network scenario. The experiment included several features to circumvent the open-endedness of the scenario. First, the participants were presented with a concrete end-goal to guide them with their decision/actions. Second, while different decisions during the exercise may lead to different attack paths, the scenario was designed such that there was only one decision path to the final goal i.e., the set of cyber contexts in which the CogVulns were tested would be the same independent of the exact decisions at any point during the exercise. Third, a public-facing Internet was simulated within the scenario to allow participants to download software packages and exploits while performing the task. Different tools have different capabilities and thus, to avoid any confounds introduced by the choice of the tool, we restricted the attack tools that can be downloaded for the purposes of the exercise. Such a design provided a balance between real-world modus operandi of an adversary and open-endedness of the experiment. Finally, to ensure that the participants can progress and reach the end goal in a reasonable time period, high-level sub-goals/sub-tasks were presented as part of a Qualtrics questionnaire to provide directions towards the next steps. The information provided in the Qualtrics questionnaire was minimal and wasn't sufficient to find the flag to move forward in the exercise. The participants were expected to explore the environment to find the flags thereby, simulating a real-world attack. Additionally, the Qualtrics also included "Help me" and "Answer" buttons to assist participants who may be stuck in a specific sub-task. The "Help me" button provided hints while "Answer" button provided a detailed step-by-step instruction to complete the task. The instructions under the "Answer" buttons were presented such that the participant still needed to make a choice on the parameters of the attack.

Experimental Results

Participant descriptive statistics

Survey	**Number of participants**	**Average Duration (HH:MM:SS)**
Training Scenario	29	3h 48m 19s
On-range Scenario #1 (Trigger)	11	15h 11m 52s
On-range Scenario #1 (Control)	16	4h 2m 1s
On-range Scenario #2 (Trigger)	11	6h 52m 25s
On-range Scenario #2 (Control)	12	5h 52m 37s

Cyber Environment

Two on-range scenarios representing a typical small-scale enterprise network were developed and configured on CyberVAN to support testing of different CogVulns. Each scenario was designed with a different end goal for the participant.

The goal of on-range scenario #1 was to compromise a database server and delete financial data. The network scenario for scenario #1 is shown in the Figure 1. The scenario was composed of 4 subnets namely, the Internet, network monitors, DMZ and the Internal portion of the network. The Internet subnet contains a Kali VM which is the starting point of the attack for the participant. The Internet segment also hosts a software repo with a limited set of attack tools to enable participants to download and install tools on the compromised hosts as they progress through the scenario. The DMZ network segment hosted typical public-facing services such as web, email and database. A log aggregation server was also included in the DMZ region to support testing of an CogVuln. The servers in the DMZ network segment were the main targets of the attack for the on-range scenario #1. The Internal network hosts workstations and internal services. A network monitor segment that hosted Zeek and Suricata IDS for collecting alert logs and other types of network logs (e.g., connection logs) from the network was also included. Finally, similar to a typical enterprise network, we included firewall between the Internet region and the DMZ region, and another firewall between the DMZ region and the internal region with firewall rules to limit connectivity and accessibility of services across the regions.

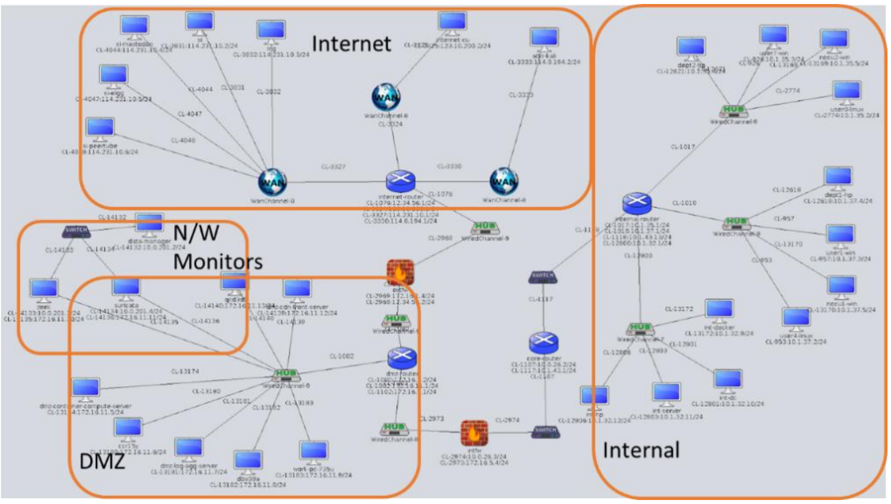

Figure 1: Network scenario for on-range scenario #1

The flow of activities performed by a participant in this scenario is described below. Figure 2 shows the attack paths and the corresponding Cogvulns tested along the path.

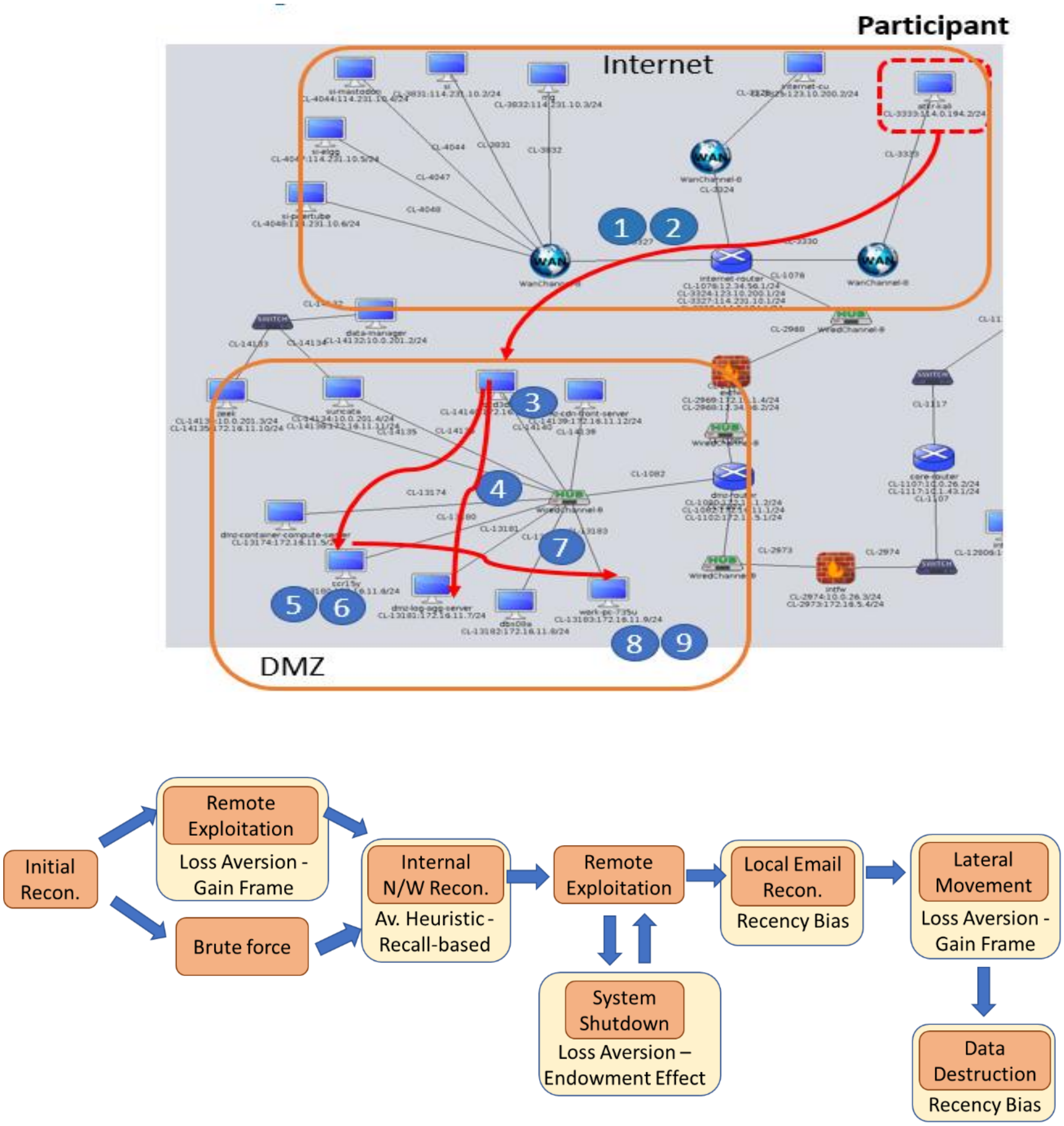

Figure 2: Attack path and CogVulns tested for on-range scenario #1

The goal of the on-range scenario #2 is to compromise internal services and establish foothold. The network scenario for scenario #2 is shown in the Figure 3. In addition to the Internet, DMZ and the Internal regions, network contains an Internal VPN service which acts as a gateway to access hosts in the Secure Lab N/W region.

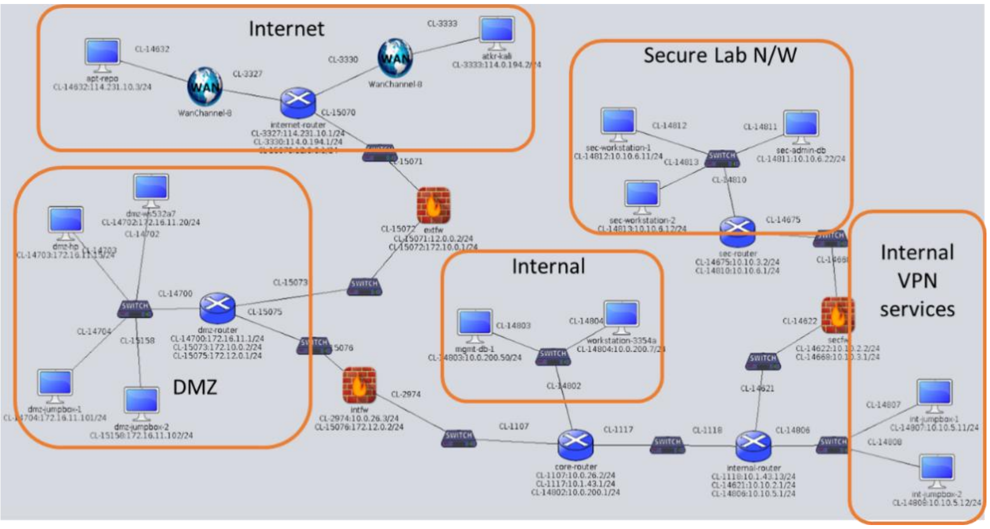

Figure 3: Network scenario for on-range scenario #2

The overall attack scenario is more complex than scenario #1 and the attack path is spread throughout the network. The exact attack path and the tested CogVulns along the path is shown below.

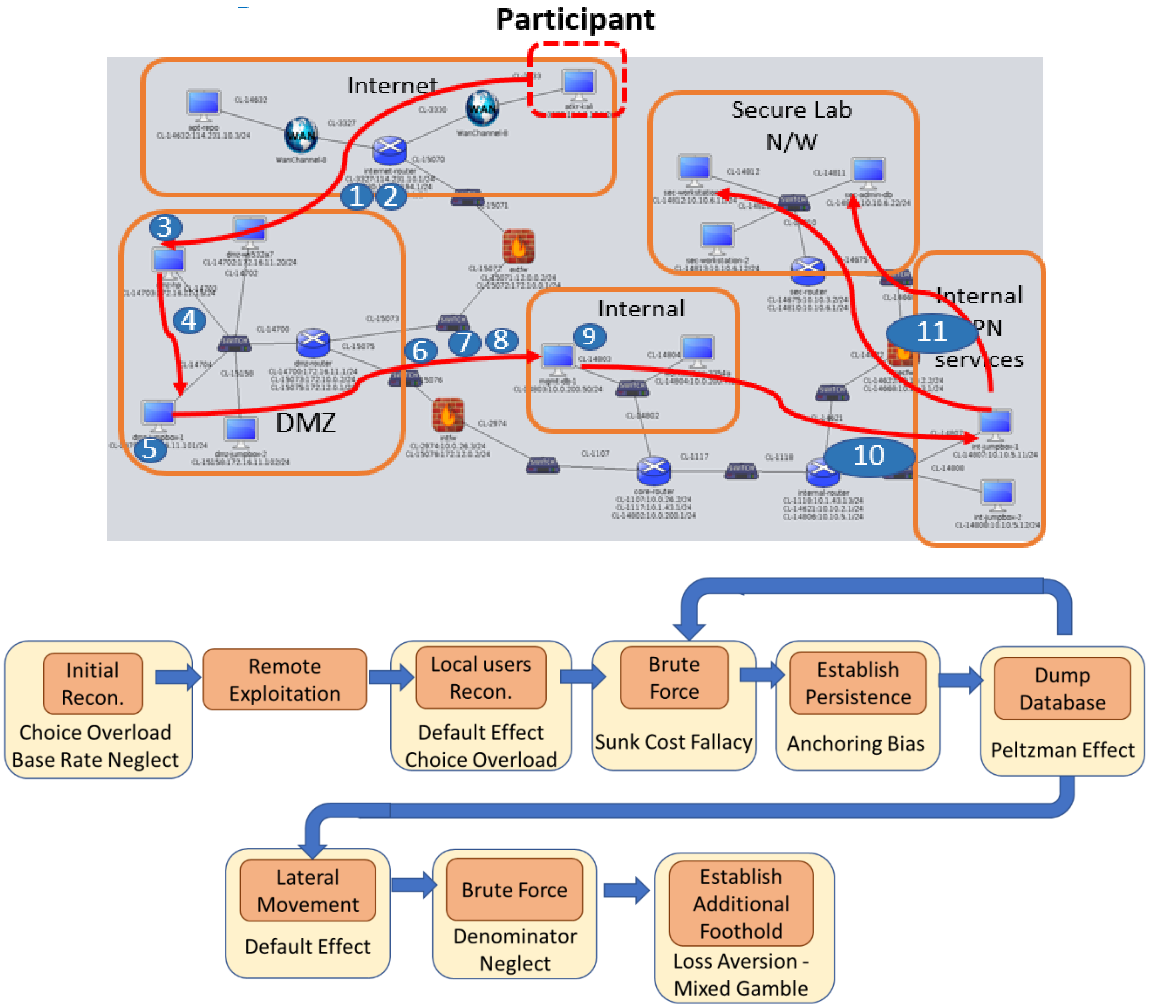

Figure 4: Attack path and CogVulns tested for on-range scenario #1

Data

Data Sources

Primary Data Sources

Collected directly from the cyber range experiment environment.

Category	Data Source	Examples of Select Data Features
Psychometrics Data	Demographics	Age, gender, education level
	Big Five (BFI)	Big Five Indices as defined by [2]
Questionnaires	GRIPS questions	Questions that indicate the individuals risk taking propensity, resistance to sunk cost, resistance to framing effect, and reflectiveness and intuitiveness.[3]
	Experience Questionnaire	Questions relating to practical and theoretical knowledge of networks and cyber security systems [1]
	Knowledge Skills and Experience Questionnaire	Questions on event-based experience, skills and certifications. 14 practical and knowledge questions covering seven stages of the kill chain
Qualtrics	Timing information	Timing of answer submissions throughout the experiment
Raw data logs from network	Host Logs	syslog and auth.log files from each host on the network
	Network Logs	Raw network traffic from all the interfaces on a central pcap collection node (either suricata or pcapsensor hosts)
	IDS Logs	Zeek and Suricata logs from the entire network
	Rootkit Logs	System calls logged by the rootkit on each host where manipulation were performed

Raw data logs from participant host (Kali box)	asciinema logs	Command logs from the participant's terminal sessions in .cast file format
	User Activity track logs	A customized file format logging keystrokes, mouse movement logs, logs on application widget creation, deletion, resize and focus
	Gnome video recordings	A desktop recording session of the participant

** Provide a citation for each psychometric assessment in the References section below.*

Derivative Data Sets

Datasets were created from aggregating, analyzing, curating, and labeling the source data.

Category	**Data Source**	**Examples of Select Data Features**
Self-Reports	Scores for individual participant characteristics	The scores extracted per participant from the GRIPs questions

Research

Hypotheses

The on-range Cyber CogVuln dataset was used to answer the following hypotheses:

[H1] Performer sensors provide similar estimates of CogVuln susceptibility to established sensors

[H2] Performer triggers activate attacker CogVulns, with a medium or high effect size

Publications

Work in progress

References

Include references to psychometric or research-backed methods used to collect data

[1] Ben-Asher, N., & Gonzalez, C. (2015). Effects of cyber security knowledge on attack detection. Computers in Human Behavior, 48, 51–61. https://doi.org/10.1016/j.chb.2015.01.039

***Note**: Additional references will be included as publications are accepted.*