Defensive Cyber Maneuvers to Disrupt Cyber Attackers

Jennifer A.B. McKneely
Johns Hopkins University Applied Physics Laboratory
Jennifer.mckneely@jhuapl.edu

Tara K. Sell
Johns Hopkins Bloomberg School of Public Health
tksell@jhu.edu

Kathleen Straub
Johns Hopkins University Applied Physics Laboratory
Kathy.Straub@jhuapl.edu

Daniel Thomas
Johns Hopkins University Applied Physics Laboratory
Dan.Thomas@jhuapl.edu

Abstract

Perimeter based defenses are limited in deterring and defeating cyberattacks. Multi-layered defenses with predictable effects are needed to provide robust cybersecurity against Advanced Persistent Threats. Cyber maneuver implements proactive defensive cyber actions to achieve positional or temporal advantages over an adversary in the cognitive, technical, and physical domains. However, understanding cognitive effects on adversaries is nascent. The paper describes application and results of an enhanced cyber maneuver framework designed to predict adversary behavioral and cognitive response to maneuvers. Results from remote pilot testing of cyber maneuvers including "herd/traffic slowing", "deception", and "stimulate a response/reset" are presented. The pilot study builds on prior research of in-lab experimentation of red teamer cognitive and behavioral response to cyber maneuvers and extends the testing method to remote configurations, enabling testing with a broader pool of participants. The framework and test methodology supports design and evaluation of cyber maneuvers' impact on attackers prior to deployment.

Keywords: Cyber Maneuver, cognition disruption, behavior-based cybersecurity.

1. Introduction

Cybersecurity attacks like the SolarWinds Orion attack (FireEye, 2020) continue to demonstrate weaknesses associated with reliance on passive perimeter-based network defenses. While necessary, these defenses are insufficient in keeping out Advanced Persistent Threats (APTs). Attackers have advantage where they can exfiltrate data in minutes (Clay, 2015) while vulnerabilities take, on average, days to mitigate once discovered. However, that mitigation is only possible after discovery which take weeks or months to occur. This reactive cyber defense approach is no match for the continually evolving threat (Clay, 2015; Schoka, 2018). Strategies are needed to more quickly expose adversaries and disrupt their progress prior to exploitation.

Securing a network requires layered defense capabilities that strengthen network edges while also providing active mechanisms to detect breaches and disrupt attackers in the system. An approach to providing layered defense is the use of cyber maneuvers whose aim to eliminate an adversary's advantage. Cyber maneuvers use actions on networks to introduce surprise, confusion, and disorder to distract and disrupt the adversary's cognitive processes (Allen, 2020; Applegate, 2012). Defensive cyber maneuvers implement specific actions intended to cause responses that disrupt adversary's stealth and mission progress. This enables defenders to proactively respond to attacks, shifting the advantage from the adversary back to the defender (C. Huang, 2015; Schoka, 2018).

Cyber defense frameworks typically focus on sophisticated technology solutions (Brantly, 2015; Duan, Al-Shaer, Islam, & Jafarian, 2018; Jafarian, 2017) or are limited to cyber deception (Almeshekah & Spafford, 2016).

The current study was undertaken as an extension of previous research (McKneely, et al, 2021) to mature the enhanced cyber maneuver framework that supports planning and implementation of cyber maneuvers with predictable effects. We aim to evaluate the feasibility of measuring cognitive and behavioral effects of human participants in a controlled distributed experimental environment. This extends the framework by providing additional methods and data to better understand cognitive and behavioral response mechanisms to cyber maneuver actions.

This paper discusses our application of the enhanced cyber maneuver framework to design and evaluate the impact of maneuvers on cyber attackers. A brief review of the cyber maneuver research is followed by description of a human subject pilot study. The study involved red teamers attempting to exfiltrate

URI: https://hdl.handle.net/10125/103448
978-0-9981331-6-4

(exfil) files from an enterprise network that had, unbeknownst to the attackers, implemented a variety of maneuvers. Table 1 illustrates how the framework was used to plan and map out the cyber maneuver to desired effects as mediated by cognitive process impacts and attacker behavior. The study was conducted remotely demonstrating the feasibility of measuring cognitive and behavioral effects of human participants in a distributed experimental environment.

Table 1 Cyber Maneuver to Desired Effect through Cognition and Behavioral Response

Cyber Maneuver	Cyber Action	Cognitive Mechanism	Desired Behavior	Desired Effect
Leverage Deception	Decoy content	Confusion Attention Decision-making Working memory	Access fake systems/ content	Detect attacker Distract from real assets Expose TTP
Stimulate a Response	Reset	Frustration Attention Decision-making Heuristics & biases	Increase activity on network	Detect attacker Deter future operations
Herd	Traffic Slowing	Frustration Confusion Attention Decision-making Working memory Heuristics & biases	Move to fake or low value asset	Thwart data theft Observe TTP

2.0 Background on Enhanced Cyber Maneuver Framework

Military operations (Hart, 1954) have successfully used maneuver to trigger predictable, observable adversary actions. Support for cyber maneuvers is more limited where studies tend to focus on technical rather than behavioral implications. For example, research describes the technical or environmental implications of introducing cyber maneuvers (Chiang et al., 2019; Sengupta et al., 2019) or simulations that model specific responses to technical cyber defense solutions (Carroll, Crouse, Fulp, & Berenhaut, 2014; Carroll & Grosu, 2011; Fugate & Ferguson-Walter, 2019; L. Huang & Zhu, 2019; Wang, Zhou, Li, Guo, & Du, 2019).

Other researchers have applied game-theory to develop a taxonomy for defensive cybersecurity deception (Pawlick, Colbert, & Zhu, 2019). This taxonomy categorizes deception types based on game-theoretic principles which are similar to maneuver aims. However, this approach assumes that actions are based on intelligent rational decision-making and does not incorporate cognitive processing of cyber attackers. Research has shown that adversary decisions are not completely rational and are influenced by cognitive biases (Cranford, et al, 2021; Gutzwiller & Ferguson-Walter, 2019). Empirical support for the cognitive effects of defensive cyber deception and maneuvers is limited but demonstrates promise (Cranford, et al., 2021; Gutzwiller & Ferguson-Walter, 2019; McKneely, et al., 2022)

Operationalization of cyber maneuver is enabled with cyber maneuver frameworks that describe how to use them to attain objectives in cyber domain (Brantly, 2015, MITRE, 2022) primarily through deception (Almeshekah & Spafford, 2016; Duan et al., 2018). However, most current cyber frameworks tend to focus heavily on the technical solutions and anticipated system response. They typically underspecify the effects on human information processing and response and do not address measurement of the impact on the humans-in-the-loop. The Engage framework (MITRE, 2022) does consider attacker perception and response yet the cognitive mechanisms are not specially addressed. The enhanced cyber maneuver framework seeks to address this omission by explicitly representing humans and cognitive processing into the cyber maneuver concept (McKneely et al., 2022). The framework provides a foundation with specific cyber maneuvers and actions mapped through the explanatory influence mechanisms and cognitive effects, shaping the desired behavioral outcomes.

The enhanced cyber maneuver framework (McKneely, et al., 2022) elaborates Allen's (2020) cyber maneuver concept with identification of the desired human behavior and the cognitive mechanism by which that behavior would emerge. Findings from the successful psychological operations (PSYOP) (JP 3-0, 2018; Paul et al., 2018) provide insights in applying deception and other cyber maneuvers with consideration of human limitations in cognitive processing and persuasion. This includes leveraging cognitive biases, working memory and inhibition limitations, intuitive response to messaging, and natural tendencies to minimize mental effort and socially connect to others to achieve predictable desired behaviors. The cognitive processing that shapes desired behavior is specified. This analysis takes into consideration general cognitive limitations (e.g., attention or working memory limits), processing or workload invoked by the maneuver (e.g., increased attention or memory load), emotional response evoked by the maneuver (e.g., increased frustration or decreased confidence), and emergent decision-making

biases resulting from peripheral route thinking and heuristic decision-making (e.g., confirmation bias.)

Using this systematic evaluation, cyber maneuver planners can exclude, select, and prioritize tactics based on a holistic analysis of the cognitive effects that mediate the success of the maneuver. For example, a defender may leverage deception by deploying multi-fidelity decoys, an attacker would interact with the compromised files with an effect that they are kept busy and detected on the network. The decoy design would attract attention of the adversary as they would appear to be worthwhile assets and invoking take-the-best and confirmation bias leading the adversary to believe these are real assets and worth going after.

To get an advantage over adversaries, cyber defenders can leverage cognitive processing theories such as dual process models, heuristic decision-making, and the Elaboration-Likelihood Model (ELM) (Petty & Cacioppo, 1986) when designing maneuvers and related cyber action solutions. Dual process theory suggests that two processing systems are continually monitoring incoming information, reasoning, and making judgements (Evans & Stanovich, 2013; Kahneman & Frederick, 2001; Stanovich & West, 2000). System 1 runs automatically with little effort and System 2 is controlled, effortful processing. Humans are wired to expend the least amount of effort when making decisions and performing tasks. Information load, particularly when combined with pressure and uncertainty, can drive attackers toward System 1 thinking (Kahneman, 2011). This System 1 processing then leads to automated responses and invoking heuristics to minimize cognitive demand and effort. Using heuristics enables faster decisions, but also risks invoking predictable errors known as cognitive biases. For example, confirmation bias occurs when decision-makers readily assimilate information confirming an existing hypothesis while actively ignoring information that counters it.

The ELM (Petty & Cacioppo, 1986) posits that people receive and interpret new information along one of two routes (central or peripheral) based on the motivation and capability of the receiver. Motivated receivers with capacity process information via the central route where it is actively elaborated, assessed, and scrutinized before integrating it into their situational hypotheses and/or current belief model. Alternatively, those receivers who are less interested or have reduced cognitive capacity process information via peripheral route thinking, without active analysis, issue-relevant thinking, or elaboration.

Cyber maneuvers can take advantage of these cognitive processes by executing cyber actions that keep attackers' System 1 engaged, are explicitly designed to leverage biases (and go unnoticed), and create contexts in which attackers are driven toward peripheral route thinking. For example, changing the network connections with system feedback could invoke information overload and help maintain dominance of the attacker's System 1 thinking. These system changes should be consistent with the adversary's goals and expectations to invoke confirmation bias and steer to peripheral route thinking and heuristic decision-making.

Research is beginning to recognize that cognitive biases are related to cyber operations (Johnson, Gutzwiller, Gervais & Ferguson-Walter, 2021). Further, studies are beginning to demonstrate that cyber attacker decisions are susceptible to cognitive biases which can impact their progress in conducting a cyberattack (Cranford, et al, 2021) even when encountering deceptive signals. Their findings reinforce the notion that prior experience and interaction with an environment shape behavior which needs to be understood by defenders to develop effective cybersecurity strategies.

Cyber maneuvers aim to enhance cyber defenses by applying proactive defensive action to establish a positional advantage in the cognitive domain. The cognitive advantage stems from the defender's ability to measurably influence the adversary's cognitive processes and emotional response, including, but not limited to, increasing workload, frustration, reducing confidence, and shaping the adversary's subsequent behavior in a predictable way.

2. Pilot Study

A remote pilot study was conducted to both demonstrate feasibility of measuring cognition and behavior in cyberattack tasks and to evaluate the impacts of maneuvers on cyber attackers. Red teamers attempted to exfil files from an enterprise network that had, unbeknownst to the attackers, implemented a variety of maneuvers. This provides additional testing methods and data to understand the underling cognitive responses to cyber actions. Additionally, the study helps advance the goal of a developing a methodology to enable deployment of maneuvers that result in predicable behavioral effects. This study was conducted in accordance with Johns Hopkins University's Institutional Review Board (IRB) under protocol IRB00214055.

2.1. Study Design

We evaluated three cyber maneuvers in a single session experiment that consisted of three trials where participants were tasked to find and download target files (Figure 1). Three maneuvers were assessed

including 1) ***deception***, implemented using decoy credentials, network assets, and content; 2) ***stimulate a response,*** implemented using network reset requiring a reconnection action resulting in noisy and more detectable adversary; and 3) ***herd,*** implemented by slowing traffic in high value areas while leaving other, low value, places in the network normally responsive. Participants started with the ***deception*** condition, then performed two additional trials; ***herd*** and ***stimulate a response***, which were counterbalanced. ***Deception*** is part of all conditions as well; therefore, the ***deception*** only condition was treated as the baseline.

In-person testing with participation from across the cybersecurity community was planned; however, restrictions on in-person experimentation from the coronavirus pandemic necessitated design of a remote experiment. We used web-based conferencing technology and Virtual Private Network (VPN) log-in to the experimental network for participant access. This connection method to the simulation environment limited research participation to employees of the Johns Hopkins University Applied Physics Laboratory (JHU/APL).

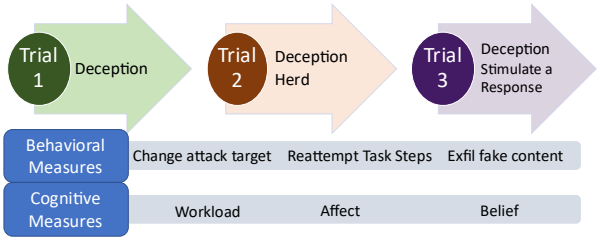

Figure 1 Session Maneuvers, Trials, & Measures

2.1.1. Participants. Individuals were recruited through purposeful sampling from a pool of red teamers (the pool included those who are trained to role play adversaries and test systems for vulnerabilities and access points). Five male staff members of the Johns Hopkins University Applied Physics Laboratory (JHU/APL) who had cyberattack knowledge and skills (cybersecurity, network design, and defensive practices) volunteered to participate in the study. Two participants were under age 35, two between 35 and 49, and one over 50. Due to time constraints and technical issues, two participants (Red02 & Red03) did not finish all trials – they did not complete the deception and herd condition.

2.1.2. Materials. Sessions were run on a simulated virtual enterprise network structure similar to that of our previous study (McKneely et al., 2022). The simulation included Microsoft Exchange servers, configured routers, installed domain controllers, deployed domain-joined Windows 10 enterprise user stations, Kali, Filebeat and Metricbeat sensors to capture host and network-based traffic, content files (including large target files ranging in size 1.5 GB to 3 GB), and applications to complete the experimental task (Cobalt Strike, and Mattermost Chat) (see Figure 2).

Participant were directed to a specific attacker VM configured to attack a specific region (enterprise network). Each network was similar in composition, but had a different selection of IP addresses, users, and named user and IT workstations. Log data was collected both from the corporate network hosts as well as the attacker machine. The corporate networks were reset to a known base configuration between participants. The chat server was used to capture communications between the participant and the experiment administrators posing as the red team supervisor. Corporate servers included a central domain controller maintaining user accounts and computer information, a web server for external web presence, a corporate windows file server, a sharepoint document server and associated database server, and a corporate mail server. The user network includes ten user workstations and five IT administrator workstations. In all cases, the participant was provided an initial beacon on a privileged IT workstation for ingress from their attack machine.

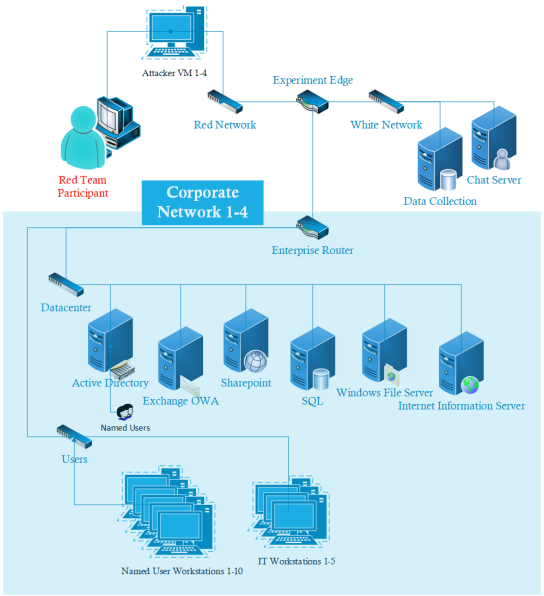

Figure 2 Simulated Network Configuration

In-task cognitive processing and state were assessed using Status Update Reports (STURs) that queried status (open ended response to "provide your current mission status") and confidence of mission completion (rating on 5-point scale from 1-very unconfident to 5-very confident). A post-trial survey was administered via Qualtrics to assess workload, confusion, and surprise. Perceived completion and difficulty were rated on 5-point scales; surprise, and

confusion were measures on a slider scale (anchored 0-Low to 100-High). Workload was measured via the NASA Task Load Index (TLX) workload assessment tool (Hart, 2006) which is a multi-dimensional survey-based measure. Six dimensions (mental demand, physical demand, time demand, subjective performance, frustration, and effort) were rated along a 20-point unanchored scale. Results include the individual factors and an overall workload calculated by a weighted score (derived by paired comparisons of the six dimensions based on which contributed more to workload and adjusting/summing each subscale rating accordingly).

2.1.3. Procedure. Participants individually completed the session accessing the testing environment via VPN connection and interacted with the research proctor via a web-based conferencing tool (Zoom). The participants shared their screen to allow for monitoring of task progress.

Participants role-played a nefarious attacker executing a cyberattack mission based on the following fictitious, disrupt-the-election scenario:

> *In a few days, there will be an election in a neighboring country (Zaltia). It's key to the security of your country that the current leaders of Zaltia stay in power. Your country is conducting cyber missions against the opposition party in support of that outcome. Elections in Zaltia happen at the regional level. There are 4 regions: North, South, East and West. Each region has its own campaign network infrastructure. Your time critical mission is to exfil campaign-related documents from each region.*

Participants were instructed on the primary exfil mission in four areas of a target network and that, if it becomes problematic to pursue, they can pursue secondary target content in each of the four areas. Participants executed attacks using Cobalt Strike running on a Windows desktop machine tunneled into the fictional Zaltian political party network simulation. Participants were provided a network map identifying their established insertion point and key terrain. They were not informed of the potential presence of cyber maneuvers and were allotted 45 minutes to complete each attack.

At predetermined intervals during the trials, participants responded to STURs using a Mattermost Chat window. After each trial, participants completed the NASA TLX and post-trial questions. Upon completing the experimental session, participants took part in a semi-structured interview that explored their general response to the mission task, whether they detected maneuvers, their expectations of deception on networks, and their typical Tactics, Techniques, and Procedures (TTPs) in conducting red team missions.

3. Findings

We collected data from multiple sources to build a converging analysis pointing to the beneficial effects of cyber maneuvers and the cognitive mechanisms that underlie their impacts. Measures included the following:

- Adversary behavior (downloads actions)
- Change in participant cognitive state
- Adversary perception of maneuver and impact on workflow

Analysis investigated cognitive processing and behaviors at the individual level; we use "Red##" to differentiate each participant we refer to. Our first participant, "Red01", tested readiness of networks and data collection apparatus and was therefore excluded from analysis. Each subsequent participant was given the same numbering convention to anonymize data.

Our small sample size limits statistical analysis. However, our analysis integrating the data shows while the cyber actions were limited in their effects, there were impacts on cognition and workflow that suggest use of cyber maneuvers has promise.

3.1. Behavioral Effects

Cyber maneuvers aimed to disrupt adversary exfil tasks by pushing decoy content, steering adversaries away from high value assets, and interrupting workflow. Disruption was evaluated though analysis of download actions (i.e., exfil of files) across the three conditions. The findings show an increase in successful downloads over the trials, which can be attributed to participants learning the tool and target environment.

The cyber actions implemented for maneuvers did not prevent participants from exfilling files. All participants exfilled most, and one participant exflilled all primary target files in the deception condition (Figure 3). Except for one participant (Red02), those who went after secondary target files got most of them and all participants took some decoys. One participant (Red05) was seemingly driven by the primary mission; they did not exfil any secondary target files during this trial. All participants got most and two participants got all primary target files in the herd condition; suggesting that the traffic slowing cyber action pushed attackers to the other mission targets (secondary and decoy). We discovered the stimulate a response maneuver implemented with the reset cyber action was not effective because the network automation reinitiated the session too quickly for network impacts to be perceptible. Therefore, to the participants, the network user experience in this condition was similar

to the first trial experienced. All three participants in this trial (Red04, Red05, Red06) were able to download all the primary and secondary mission files, they took some of the decoy but not all. This is a somewhat higher rate than in the herd condition suggesting that traffic slowing may have had the desired effect in the herd condition.

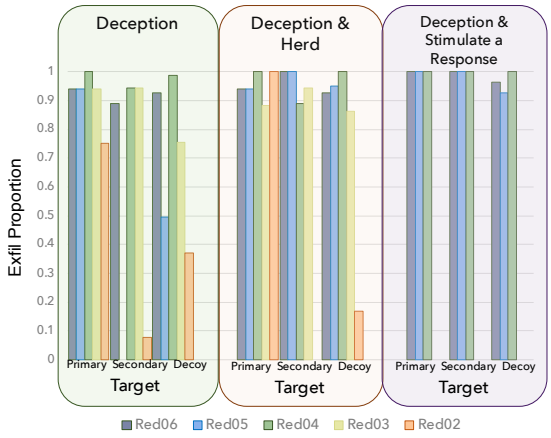

Figure 3 Participant Exfil Target Completion Proporation By Condition & Type

Further investigation into differences between herd and stimulate a response reflects some effect of traffic slowing and deception in completing the exfil. One participant, Red06, canceled downloading files during the exfil task. Investigation of file download attempts across primary, secondary, and decoy files revealed that cancelled file downloads were large, most had deception cues ("fake" or "phony") in their filenames, four were cancelled and restarted, and two of those restarted were completed (Figure 4). The total number of files downloaded could exceed the total primary target file set. There were 116 primary target files in the deception condition, 116 primary target files in the herd, and 124 in the stimulate a response condition.

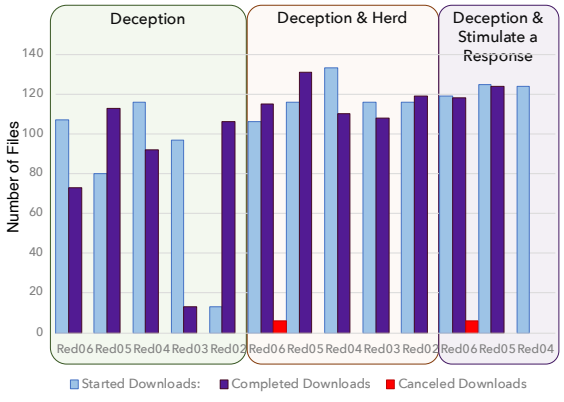

Figure 4 File Download State By Participant and Condition

3.2. Cognitive Effects, Perception, and Workflow

Cognitive measures were consistent with the behavioral data. The findings indicate learning effects where there is higher workload, confusion, and surprise in the first (deception) trial then the following trials. Across trials, workload is shown to be higher in the herd condition, suggesting herd impacts cognitive processing (Figure 5).

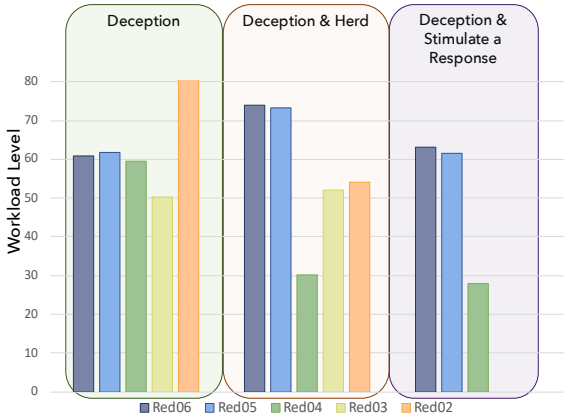

Figure 5 Total Workload (NASA-TLX Score) By Participant and Condition

STURs were submitted every 15 minutes during the trial. A comparison of STUR response to actual progress (exfil percentage) showed all but one participant was accurate in their self-assessment. This suggests that participants had a good sense of how their actions impacted progress towards mission completion.

Confidence in mission completion varied across conditions (Figure 6) and within trials (Figure 7). Some participants (Red04) were quite confident and were able to complete the mission across all trials and others (Red02) had lower confidence and performance throughout.

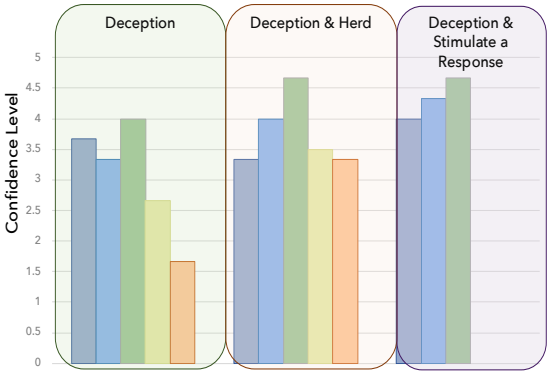

Figure 6 Average In-Task Confidence Rating (5-point scale) By Trial

While we found confidence varied across and within trials, we found confusion and surprise decreased across the conditions (Figures 8 & 9) suggesting that over time participants were figuring out what to do and how to do it. In fact, there is a substantial drop in confusion from the first trial (deception) to the following two for each participant except Red03, whose confusion data is missing for that trial (inspection of interview responses did show consistency with the general finding). This provides additional evidence that it took time for participants to understand the environment and feel comfortable with the task. This also suggests that there were no detectable network actions that surprised individuals.

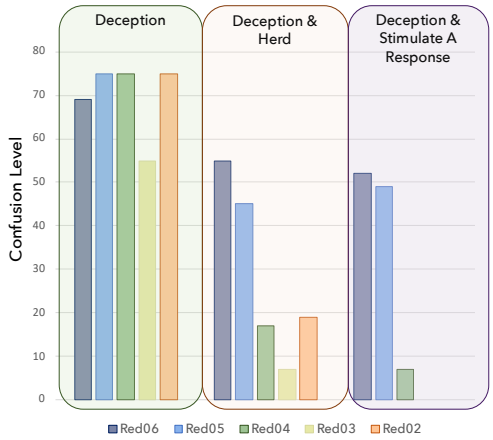

Figure 7 Participant Confusion Rating by Condition

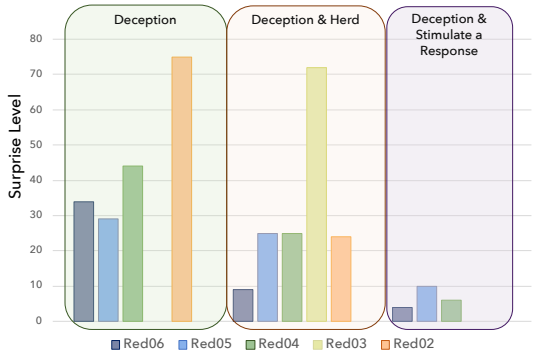

Figure 8 Participant Surprise Rating By Condition

Workflows from participants Red04, Red05, and Red06 (those who completed all three trials) were analyzed from system log files and post-session interviews to gain insight into how individuals conducted the mission and if the maneuvers affected their approach. We found variability of workflow strategy across participants, workflows were refined from repeat exposure to the toolset, and some were impacted by traffic slowing. Interview data showed participants' expectations shaped how they engaged in the task, as exemplified by the comment "I would not know if I encountered a decoy. I was not thinking that, I had my own biases and was therefore focused on the network behavior." Findings from workflow analysis of participants that experienced all three conditions are described below.

Participant Red04's trial order was deception, herd, and stimulate a response. As described in Section 2.1, the reset implementation in the stimulate a response condition was such that the network connection loss was not perceptible. Therefore, this condition was similar in user experience to the deception condition. Task patterns were very similar across the three trials, they went after the secondary files and then attempted to exfil the primary target files for each area. Their interview data was consistent with the observed workflow with comments like "I didn't spend a lot of time exploring around, was set up to go in a specific direction and I took that direction."

Participant Red05's trial order was deception, herd, and stimulate a response. They showed the most shift in strategy over the sessions with interview data indicating this was to ensure something was taken. Representative comments included "changed with each trial as I got more comfortable with the environment." They started by going after the primary mission for each area in their first trial and did not go after the secondary mission. In the second trial (herd) the primary mission was attempted in one area, then they shifted to the secondary mission repeating this across areas; they noticed large files were not downloading quickly and those with "fake" or "phony" in their filenames were cancelled and restarted later in the trial. In the third trial they attempted all targets at once, again cancelling and restarting large files.

Participant Red06's trial order was deception, herd, and stimulate a response. However, they showed little difference across trials which was a different approach from the others. They put forward a parallel processing strategy across all three trials that streamlined exfil and allowed completion of the mission before the trial end. Interview data was consistent with this observed pattern with comments like "as I was more comfortable I was able to parallelize things, can do multiple things while running - getting small files, then looking for 1GB files and looking at what was the same."

4. Discussion

The pilot study addressed three maneuvers, deception, herd (via traffic slowing), and stimulate a response (via reset), from our framework. We found preliminary support that maneuvers can shape attacker behavior through cognitive processing effects. While workload was not significantly impacted, it appears

the attackers were primarily in System 1 thinking. They were executing skilled actions with little need to execute deductive reasoning (Evans & Stanovich, 2013). Essentially, they were focused on the mission at hand and did not dwell on changes in network responsiveness. Network performance anomalies were expected and seeing differences across system assets conformed to this belief. Expectations shaped their perceptions of their environment indicating that heuristics and corresponding biases are ripe for the cyber defender to take advantage (Gutzwiller & Ferguson-Walter, 2019; Kahneman, 2011; Johnson et al., 2021).

Interview results indicate System 1 thinking dominated information processing, suggesting principles of ELM (Petty & Cacioppo, 1986) and cognitive biases (Kahneman, 2011) can be exploited to maintain peripheral processing to reinforce preconceived beliefs about defenses a network may have in place. That is, while attackers are modifying their response to actions on network (cancel a download in traffic slowing) they are likely not actively considering alternative explanations to the network behavior. Therefore, it appears possible to design cyber maneuvers to trigger behavioral responses that leverage low elaboration and low likelihood of System 2 thinking based on take advantage of adversary expectation bias.

Table 2 presents a summary of findings aligned to the cyber maneuvers. Through analysis across data sets, we found evidence that the herd (traffic slowing) maneuver increased workload and behavioral activity, increased confusion, were not detected as a purposeful system action or cyber maneuver, and was effective in interrupting exfil downloads. Disruption to file downloads was not noticed or recognized and could be attributed to a failure to connect diminished progress to defender behavior. That is, the slow progress felt like "business as usual." This could potentially be related to cognitive biases, where the participant is focused on executing the task, not noticing network changes and believing any disruption to their efforts is due to error, not an intentional defensive action. Comments related to task performance highlight this focusing of attention: "*When you have fingers on keyboard, you don't care about other stuff. You are focused on your task*."

While deception and stimulate a response did not show substantial impacts on cognition or behavior, we did find that decoy content was interacted with and downloaded. This is consistent with previous research and shows deceptive assets and content can be useful for trapping and observing cyber attacker TTPs and deploying active cyber defenses in decoy files (Ferguson-Walter, Lafon, & Shade, 2017; Ferguson-Walter et al., 2019; Heckman et al., 2013; Shade et al., 2020; Yahyaoui & Rowe, 2015).

Table 2 Summary Cyber Maneuver Effects

Cyber Maneuver	Cyber Action	Cognitive Finding	Behavioral Finding
Leverage Deception	Decoy content	Confidence in exfil, attention paid to file names - little impact on decision-making. heuristics and bias	Exfilled fake content
Stimulate a Response	Reset	No Effect	No Effect
Herd	Traffic Slowing	Increased workload, reduced confidence	Paused download

Our pilot study results demonstrate the ability to capture data directly related to the measurement of cyber maneuver effects on attacker behaviors and cognition over a remote experiment configuration. Effort to expand this research to understand sensitivity of specific cyber actions (for example, how slow should traffic be without impacting legitimate users), and extend to additional cyber maneuvers/actions is needed. This will provide defenders with a broader variety of cyber maneuvers they can have confidence in. Further, while we found that the maneuvers impacted behavior, they did not prevent mission completion, research into different cyber actions and their predicted response is needed.

There were limitations to this pilot study that should be addressed in future studies. Specifically, experimentation with larger sample sizes who are more representative of the target population (red teamers with more offensive cyber experience) would be beneficial. Our small sample size did not allow for analysis of experience level impacts on cognition or behavior from cyber maneuvers. Larger sample sizes would enable more analysis on patterns of performance and cognitive response. Further, the current study investigated individual attacker performance; however, cyber operations are most often conducted as part of a team. Studies with larger participant pools can be designed to capture individual and team performance considerations measures. Additionally, while our red teamers had work experience role-playing adversaries it was limited, out-reach to organizations and participant pools have move applied work with APT TTPs and cultural insights would allow for more confidence in generalizing cognitive and behavioral insights to real-world cyber attackers.

The experimental task is limited in generalizability; cyberattacks typically occur over extended periods of time. Future experimental task design should investigate protocols that occur over days (or longer). This will challenge cognitive

measurement but result in increased confidence of active cyber defense methods.

5. Conclusions

The enhanced cyber maneuver framework (McKneely et al., 2022) elaborates Allen's (Allen, 2020) concept with specification of desired adversary behavior and explicit identification of cognitive mechanisms underlying that behavior. While Allen's concept embraces behavioral effects it did not specify the cognitive mechanism that led to the desired behaviors and effects. We applied findings from current research in cyber deception and network change effects (Bellekens et al., 2019; Farar, Bahşi, & Blumbergs, 2017; Ferguson-Walter, Lafon, & Shade, 2017; Ferguson-Walter et al., 2019; Heckman et al., 2013; Pal, Lageman, & Soule, 2018; Shade et al., 2020; Yahyaoui & Rowe, 2015) and cognitive/behavioral theory (Petty & Cacioppo, 1986; Stanovich & West, 2000; Kahneman, 2011; Evans & Stanovich, 2013) to define and map cognitive mechanisms to cyber maneuvers/actions, identify measures of cognition and performance and construct a methodology that facilitates the design and implementation of maneuvers. This framework specifies multi-source data collection to assess the efficacy of cyber maneuvers and the cognitive processes mitigating effects enable predictive assessment of defenses prior to deployment.

Defensive cyber maneuvers should be a key component to protection, monitoring, and response approaches in cyber defense. Actions can be taken on a network to expose undetected adversaries by causing them to become visible and actionable and to disrupt adversary missions by reshaping or curtailing adversary progress. Previous study in cyber maneuver explored the technical characteristics and performance of cyber actions and did not address the human attacker. The enhanced cyber framework developed and demonstrated in this research provides a systematic approach to design, develop, and assess defensive cyber actions prior to deployment. The inclusion of measures of (implicit) cognitive effect as well as predicted (explicit) behavioral actions provides explanatory context for maneuver success.

The study contributes to the development of the enhanced cyber maneuver framework by demonstrating the ability to measure behavior and cognitive effects from cyber maneuvers. We found preliminary evidence that cyber maneuvers measurably influence behavior. We demonstrated an effective experimental protocol conducted via remote-based simulated network environment. Application of the enhanced cyber maneuver framework was demonstrated, thus providing a sound foundation for continued research of cyber maneuvers.

12. References

Allen, P. (2020). Cyber Maneuver and Schemes of Maneuver. *Cyber Defense Review*, *November*, 79–96.

Almeshekah, M. H., & Spafford, E. H. (2016). Cyber Security Deception. In *Cyber Deception: Building the Scientific Foundation* (pp. 1–312). https://doi.org/10.1007/978-3-319-32699-3

Applegate, S. (2012). The Principle of Maneuver in Cyber Operations. *2012 4th International Conference on Cyber Conflict (CYCON 2012) IEEE*. https://www.researchgate.net/publication/236020494

Bellekens, X., Jayasekara, G., Hindy, H., Bures, M., Brosset, D., Tachtatzis, C., & Atkinson, R. (2019). From Cyber-Security Deception To Manipulation and Gratification Through Gamification. *HCI for Cybersecurity, Privacy and Trust: First International Conference, HCI-CPT 2019, Held as Part of the 21st HCI International Conference*. http://arxiv.org/abs/1903.08918

Brantly, A. F. (2015). Strategic Cyber Maneuver. *Small Wars Journal*.

Carroll, T. E., Crouse, M., Fulp, E. W., & Berenhaut, K. S. (2014). Analysis of Network Address Shuffling as a Moving Target Defense. *IEEE ICC 2014 - Communication and Information Systems Security Symposium*.

Carroll, T. E., & Grosu, D. (2011). A game theoretic investigation of deception in network security. *Security and Communication Networks*, *4*(December 2010), 1162–1172. https://doi.org/10.1002/sec.242

Chiang, C. Y. J., Venkatesan, S., Sugrim, S., Youzwak, J. A., Chadha, R., Colbert, E. J., Cam, H., & Albanese, M. (2019). On defensive cyber deception: A case study using SDN. *Proceedings - IEEE Military Communications Conference MILCOM*, *2019-Octob*, 110–115. https://doi.org/10.1109/MILCOM.2018.8599755

Cranford, E.A., Gonzalez, C., Aggarwal, P., Tambe, M., Cooney, S., & Lebiere, C. (2021). Towards a Cognitive Theory of Cyber Deception. *Cognitive Science, 45(7),* https://doi.org/10.1111/cogs.13013

Clay, P. (2015). A modern threat response framework. *Network Security*, *2015*(4), 5–10. https://doi.org/10.1016/S1353-4858(15)30026-X

Duan, Q., Al-Shaer, E., Islam, M. M., & Jafarian, H. (2018). CONCEAL: A Strategy Composition for Resilient Cyber Deception– Framework, Metrics and Deployment. *2018 IEEE Conference on Communications and Network Security (CNS).* IEEE.

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*, *8*(3), 223–241. https://doi.org/10.1177/1745691612460685

Farar, A., Bahşi, H., & Blumbergs, B. (2017). A Case Study About the Use and Evaluation of Cyber Deceptive Methods Against Highly Targeted Attacks. *2017 International Conference On Cyber*

Incident Response, Coordination, Containment & Control (Cyber Incident).
Ferguson-Walter, K. J., Lafon, D. S., & Shade, T. B. (2017). Friend or Faux: Deception for Cyber Defense. *Source: Journal of Information Warfare*, *16*(2), 28–42. https://doi.org/10.2307/26502755
Ferguson-Walter, K., Shade, T. B., Rogers, A., Trumbo, M., Nauer, K., Divis, K. M., Jones, A.P., Combs, A., & Abbott, (2019). The Tularosa Study: An Experimental Design and Implementation to Quantify the Effectiveness of Cyber Deception. *Proceedings of the 52nd Hawaii International Conference on System Sciences*. https://Hdl.Handle.Net/10125/60164
FireEye. (2020). *Highly Evasive Attacker Leverages SolarWinds Supply Chain to Compromise Multiple Global Victims With SUNBURST Backdoor.* https://www.fireeye.com/blog/threat-research/2020/12/evasive-attacker-leverages-solarwinds-supply-chain-compromises-with-sunburst-backdoor.html.
Fugate, S., & Ferguson-Walter, K. (2019). Artificial Intelligence and Game Theory Models for Defending Critical Networks with Cyber Deception. *AI Magazine*, *40*(1), 49–62.
Gutzwiller, R. S., & Ferguson-Walter, K. (2019). Are Cyber Attackers Thinking Fast and Slow? Exploratory Analysis Reveals Evidence of Decision-Making Biases in Red Teamers. *Human Factors and Ergonomics Society Annual Meeting Vol. 63, No. 1,* 427–431. https://www.researchgate.net/publication/334376417
Hart, B. H. L. (1954). *Strategy*. New York, New York, USA: Praeger.
Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 50, No. 9,)*, 904–908.
Heckman, K. E., Walsh, M. J., Stech, F. J., O'Boyle, T. A., Dicato, S. R., & Herber, A. F. (2013). Active cyber defense with denial and deception: A cyber-wargame experiment. *Computers and Security*, *37*, 72–77. https://doi.org/10.1016/j.cose.2013.03.015
Huang, C. (2015). *Towards Effective Techniques For Cyber Maneuver Defenses*. Penn State University.
Huang, L., & Zhu, Q. (2019). Adaptive strategic cyber defense for advanced persistent threats in critical infrastructure networks. *Performance Evaluation Review*, *46*(2), 52–56. https://doi.org/10.1145/3305218.3305239
Jafarian, J. H. (2017). *Cyber agility for attack deterrence and deception*. The University of North Carolina at Charlotte.
Johnson, C. K., Gutzwiller, R. S., Gervais, J., & Ferguson-Walter, K. J., (2021). Decision-Making Biases and Cyber Attackers. 2*021 36th IEEE/ACM International Conference on Automated Software Engineering Workshops (ASEW)*, pp. 140-144.
JP 3-0. (2018). *Joint Publication 3-0 Joint Operations*.
Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.
Kahneman, D., & Frederick, S. (2001). Representativeness revisited: Attribute substitution in intuitive judgment 1. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics of Intuitive Judgment: Extensions and Applications*. New York, New York, USA: Cambridge University Press.
McKneely, J. A. B., Sell, T. K., Straub, K. A., Ayenson, M. D., & Thomas, D. (2022). Shaping Attacker Behavior: Evaluation of an Enhanced Cyber Maneuver Framework. In A. El-Saeidi (Ed.), *HCI for Cybersecurity, Privacy and Trust, as part of the 24TH International Conference on Human-Computer Interaction*: *Vol. volume 32*.
MITRE (2022). A Starter Kit in Adversary Engagement. MITRE PR_21-01759-17 2-28-2022 VERSION 1.0; https://engage.mitre.org/wp-content/uploads/2022/04/StarterKit-v1.0-1.pdf
Pal, P. P., Lageman, N. J., & Soule, N. B. (2018). Disrupting Adversary Decision Logic: An Experience Report. *Journal of Information Warfare*, *17*(3), 78.
Paul, C., Clarke, C. P., Schwille, M., Hlávka, J. P., Brown, M. A., Davenport, S. S., Porche, I.R., & Harding, J. (2018). L*essons from Others for Future US Army Operations in and Through the Information Environment*. RAND Corporation www.rand.org/t/RR1925z1
Pawlick, J., Colbert, E., & Zhu, Q. (2019). A Game-theoretic Taxonomy and Survey of Defensive Deception for Cybersecurity and Privacy. *ACM Computing Surveys, 52, 4*.
Petty, R. E., & Cacioppo, J. T. (1986). The Elaboration Likelihood Model of Persuasion. In *Communication and Persuasion*, 1–24. https://doi.org/10.1016/S0065-2601(08)60214-2
Schoka, A. (2018). Training Cyberspace Maneuver | Small Wars Journal. *Small Wars Journal*. https://smallwarsjournal.com/jrnl/art/training-cyberspace-maneuver
Sengupta, S., Chowdhary, A., Sabur, A., Huang, D., Alshamrani, A., & Kambhampati, S. (2019). A Survey of Moving Target Defenses for Network Security. *Submitted to the IEEE*, 1–34. ttp://arxiv.org/abs/1905.00964
Shade, T. B., Rogers, A. V, Ferguson-walter, K. J., Elson, S. B., Fayette, D. K., & Heckman, K. E. (2020). The Moonraker Study : An Experimental Evaluation of Host-Based Deception. In *Hawaiian International Conference on System Science (HICSS)*, (1-10).
Stanovich, K. E., & West, R. F. (2000). Individual Differences in Reasoning. *Behavioral and Brain Sciences*, *23*(5), 645–665.
Wang, S., Zhou, Y., Li, Y., Guo, R., & Du, J. (2019). Quantitative analysis of network address randomization's security effectiveness. *International Conference on Communication Technology Proceedings, ICCT*, *2019-October*, 906–910. https://doi.org/10.1109/ICCT.2018.8600181
Yahyaoui, A., & Rowe, N. C. (2015). Testing simple deceptive honeypot tools. *Cyber Sensing 2015*, *9458*(May 2015), 945803.