Proceedings of the 2024 Winter Simulation Conference
H. Lam, E. Azar, D. Batur, S. Gao, W. Xie, S. R. Hunter, and M. D. Rossetti, eds.

DESIGN, MODELING AND SIMULATION OF CYBERCRIMINAL PERSONALITY-BASED CYBERATTACK CAMPAIGNS

Jeongkeun Shin[1], Geoffrey B. Dobson[2], L. Richard Carley[1], and Kathleen M. Carley[1,2]

[1]Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA
[2]Software and Societal Systems Department, Carnegie Mellon University, Pittsburgh, PA, USA

ABSTRACT

Cybersecurity challenges are inherently complex, characterized by both advanced technical elements and complex aspects of human cognition. Although extensive research has explored how victims' human factors affect their susceptibility to cyberattacks, the influence of cybercriminals' personalities on the pattern of cyberattack campaigns and the resulting damage to organizations has not received equivalent attention. To bridge this research gap through computer simulation, we introduce a cyberattack campaign designer that enables modelers to construct cyberattack campaigns. This tool allows modelers to define how the process or pattern of the cyberattack campaign can change based on the cybercriminal's personality and simulates how these personality differences influence the magnitude of cyberattack damage on the target organization. In this paper, through computer simulation, we demonstrate how cautious and reckless personalities result in variations in the cyberattack pattern and, consequently, affect the magnitude of cyberattack damage, despite identical cyberattack objectives and techniques used at each step.

1 INTRODUCTION

Effective cybersecurity strategies must address both technological and human factors. The technological aspect involves creating secure systems and networks to minimize the potential system vulnerabilities and protect against cyberattacks. On the other hand, understanding the human dimension is essential because the human behavior often plays a crucial role in cybersecurity breaches. This includes the tactics of cybercriminals, who exploit human psychology through methods like spearphishing and social engineering, as well as the behaviors of end users, whose actions can either mitigate or increase the risk of a security breach. To address the human aspects of cybersecurity, numerous researchers have investigated which human factors make individuals more vulnerable to cyberattacks. For example, Eftimie et al. carried out empirical phishing simulations to analyze how age, gender, and the Big Five personality traits influence an individual's susceptibility to spearphishing attack (Eftimie, Moinescu, and Răcuciu 2022). In another study, Ribeiro et al. undertook an empirical study to investigate how human factors such as age, gender, technological competencies, education level, income level, routine internet activities, and knowledge of phishing influence to an individual's phishing susceptibility (Ribeiro, Guedes, and Cardoso 2024). Compared to the analysis of victims, there has been significantly less exploration into how the personalities and other human factors of cybercriminals influence their patterns of attack behavior and decision-making processes. Kranenbarg et al. compared the personality traits of cyber-dependent crime suspects, offline crime suspects, and a community sample not involved in crime, using the HEXACO personality domain (Lee and Ashton 2004) to identify distinct personality profiles among these groups (Kranenbarg et al. 2023). Bada and Nurse explored the psychological traits, motivations, backgrounds, and behavioral patterns of cybercriminals through detailed case studies, highlighting the complex interplay of socio-technical factors in cybercrime activities (Bada and Nurse 2023). Although these studies effectively analyze the personalities of cybercriminals, they do not extend their analysis to explore how these personalities influence the patterns or the temporal progression of cyberattacks. Further investigation is needed to understand how these traits impact the decision-making

processes and flow of cyberattack campaigns. However, the primary challenge in empirically studying the relationship between cybercriminals' personalities and their cyberattack patterns stems from the inherent difficulties in accessing cybercriminals for research purposes. Additionally, conducting a scenario in which cybercriminals are encouraged to carry out an attack to observe their cyberattack patterns poses significant ethical concerns. In light of these challenges, computer simulation can be a viable alternative solution. Computational models can serve as testbeds, enabling researchers to explore "what-if" scenarios and predict the potential impact of new technologies or policies (Carley 2002). However, while current cybersecurity models are adept at replicating real-world cyberattack cases to measure the potential magnitude of damage and evaluate various defense strategies within simulation environments (Dobson and Carley 2017; Shin, Carley, and Carley 2024), they do not venture into generating and testing various "what-if" scenarios that consider the diverse personalities of cybercriminals and corresponding cyberattack patterns.

In this paper, to address this gap, we introduce the Cyberattack Campaign Designer, an environment where modelers can design, simulate and analyze the personalities of cybercriminals and the corresponding cyberattack patterns that modelers can imagine. By leveraging this tool, in this paper, we take the first step of how allowing cybercriminal personality to be cautious or reckless creates subtle differences in the pattern of the phishing campaign for data exfiltration, and consequently, make a difference in the magnitude of cyberattack damage in the target organization.

2 RELATED WORKS

There have been numerous efforts to incorporate the dynamics of various cyberattack campaigns and cyber warfare scenarios into simulation models (Kavak et al. 2021). Initially, Vernon-Bido et al. explored various social science theories to analyze human factors that may influence individuals to become cyber attackers, using an agent-based model (Vernon-Bido et al. 2016). Next, Dobson and Carley developed the Cyber-FIT framework (Dobson and Carley 2017), which focused on simulating Denial of Service (DoS), Routing Protocol Attacks (RPA), and Phishing Attacks within a military context. Building on this foundation, Dobson et al. later advanced their modeling to align with the Cyber Kill Chain (Yadav and Rao 2015), which maps the stages of cyberattacks from reconnaissance to action (Dobson, Rege, and Carley 2018). Ultimately, they refined their approach by modeling each step of the cyberattack campaigns according to the tactics and techniques outlined in the MITRE ATT&CK framework (Strom et al. 2018), providing a more detailed and comprehensive simulation of cyber threats (Dobson and Carley 2021; Dobson 2022).

Following the development of cyberattack campaign models, researchers have also explored their application in assessing potential organizational damage and evaluation of the effectiveness of defense strategies against various cyber threats. Initially, Carley and Svoboda constructed a virtual organization model that can possibly serve as a potential target in simulations (Carley and Svoboda 1996). However, this model did not include direct interactions with cyberattack campaigns. Next, Dobson and Carley developed a modeled the cyber response team agent tasked with mitigating the damage from cyberattacks (Dobson and Carley 2020). This model also includes components to simulate the team's cyber situational awareness to represent their ability to understand and respond to cyberattacks within the simulation environment (Dobson and Carley 2018). Subsequently, Shin et al. introduced the OSIRIS framework (Shin et al. 2022), an agent-based modeling and simulation (Macal and North 2009) framework that simulates end-user agents, incorporating their work schedules, behavior patterns, and social networks. The framework also integrates human factors into these agents to more accurately reflect their susceptibility to phishing attacks (Shin, Carley, and Carley 2023). The OSIRIS framework has been applied in a range of scenarios, including estimating potential damage from ransomware attacks (Shin et al. 2022), evaluating the human firewall strategies against phishing campaigns (Shin et al. 2023b), assessing the effectiveness of intrusion detection systems (IDS) against various types of Denial of Service (DoS) attack campaigns (Shin et al. 2023a), and exploring the adverse impacts of high IDS false alarm rates during concurrent DoS and phishing attacks (Shin, Carley, and Carley 2024).

3 CYBERATTACK CAMPAIGN DESIGNER

In this section, we present our Cyberattack Campaign Designer, developed to help modelers in designing cyberattack patterns for red-cyber attacker agents within the Cyber-FIT framework (Dobson and Carley 2021; Dobson 2022). While the Cyber-FIT framework introduced a method for modeling cyberattack campaigns through sequences of MITRE ATT&CK (Strom et al. 2018) tactics and techniques, and using these sequences for cyberattack simulation, it lacked an intuitive way for modelers to modify these campaigns by adding or removing specific tactics or techniques. Our Cyberattack Campaign Designer addresses this limitation, enabling modelers to create tailored cyberattack campaigns with ease.

The design process unfolds in two primary steps. Initially, modelers select the MITRE ATT&CK (Strom et al. 2018) tactics and techniques they want to include in their campaign, as depicted in Figure 1. In the subsequent step, the modelers can construct a Markov Chain (Norris 1998) using the chosen tactics and techniques. This chain outlines the progression of the attack, allowing modelers to specify both the starting point (for example, technique T1592 in Figure 2) and the intended targets (for example, techniques T1485 and T1486 in Figure 2). Additionally, our Cyberattack Campaign Designer permits the formulation of specific rules for exceptions. For instance, if the red-cyber attacker agent encounter a particular MITRE ATT&CK (Strom et al. 2018) technique more than a predetermined number of times, it can be configured to automatically proceed to a designated target technique, bypassing the Markov Chain's usual flow. This feature enhances the flexibility and customization of the cyberattack campaign design process, empowering modelers to accurately represent complex attack scenarios.

Upon completing their cyberattack campaign designs, modelers can export their work, which can then serve as input for the Cyber-FIT framework (Dobson and Carley 2021). During simulation, the red-cyber attacker agents within the Cyber-FIT framework adhere to the crafted scenarios to execute the virtual cyberattack campaign.

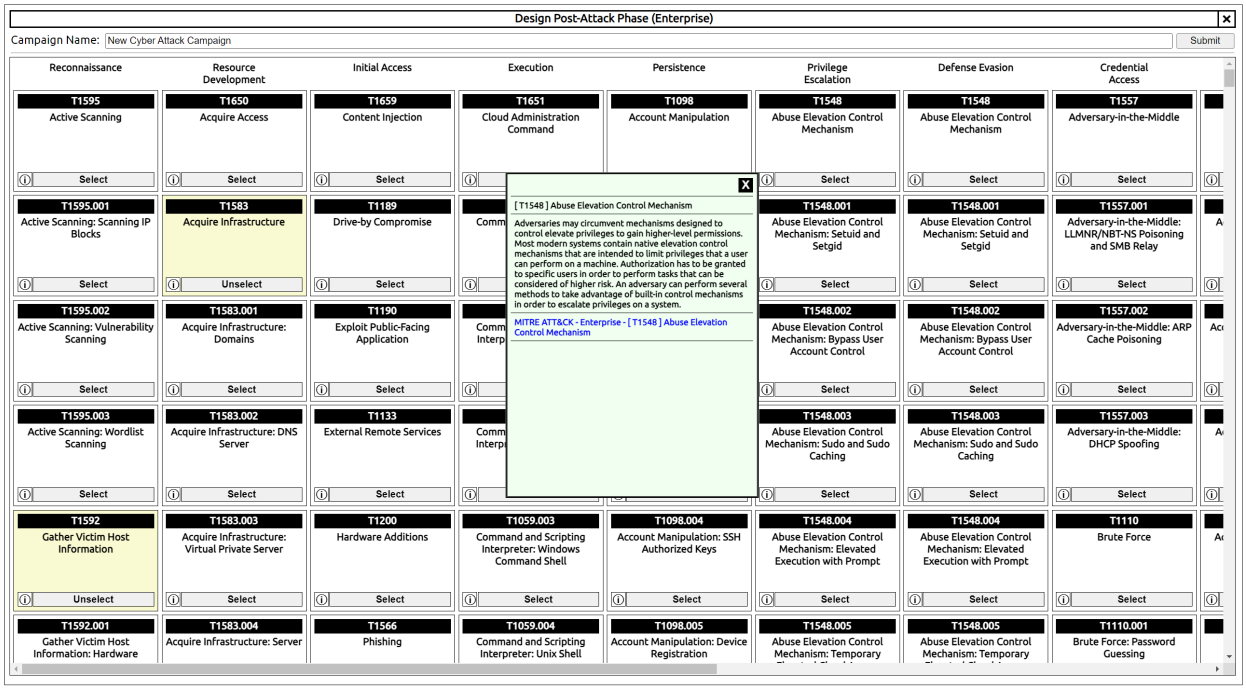

Figure 1: Cyberattack campaign designer step 1 - Selecting MITRE ATT&CK tactics and techniques.

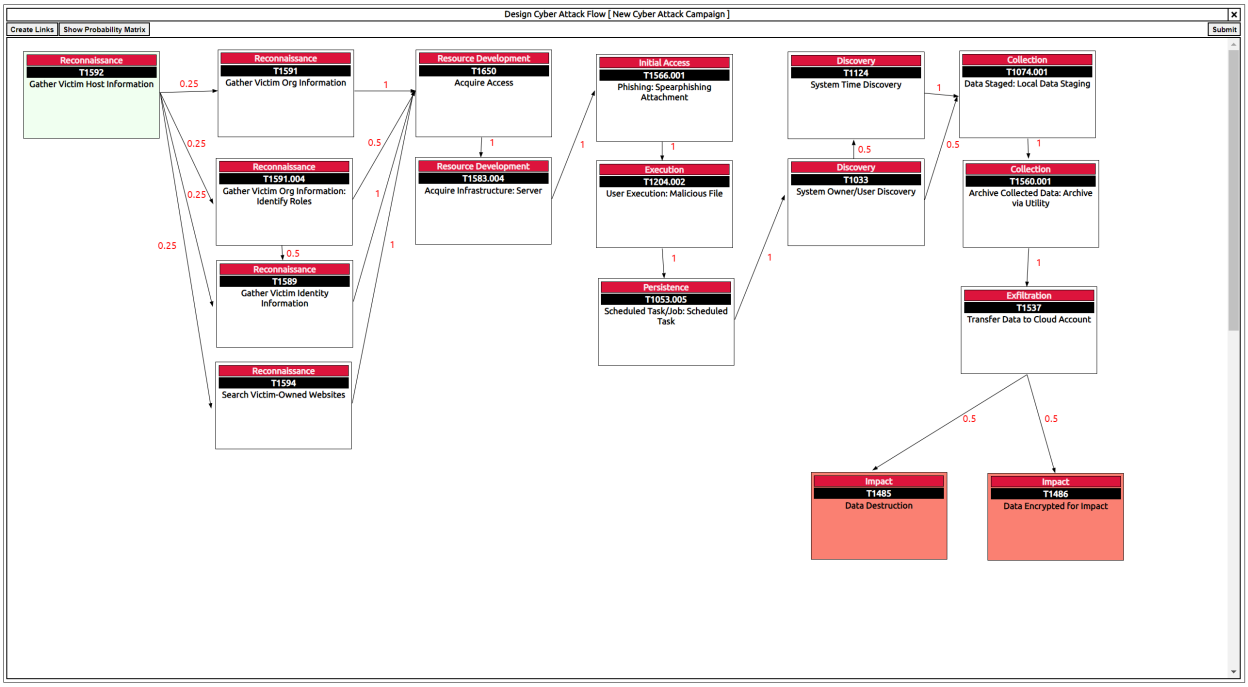

Figure 2: Cyberattack campaign designer step 2 - Building the Markov Chain for cyberattack progression.

4 PERSONALITY-BASED CYBERATTACK CAMPAIGNS

In this section, we detail two distinct cyberattack campaigns. The goal of both campaigns is to successfully exfiltrate critical data from the target organization's computing devices via spearphishing. We assume that every computing device agent within the virtual organization holds 100GB of critical information. The primary objective of the red team attacker agent is to exfiltrate as much data as possible from each compromised device. We adapt, simplify, and modify the cyberattack campaign model from Shin et al.'s study (Shin, Carley, and Carley 2024), which modeled an 11-day phishing campaign for data exfiltration and ransomware attacks, drawing upon actual cyberattack incident reports (The DFIR Report 2021; The DFIR Report 2023). Both of our cyberattack campaigns are structured using the same MITRE ATT&CK techniques (Strom et al. 2018), as shown in Figures 3 and Figure 4. The primary distinction between these campaigns lies in the personality of the red team attacker agent: in the first campaign, we assume that the agent exhibits a cautious behavior, while in the second, we assume that agent exhibits reckless behavior. In the rest of this section, we will specify the details of each technique employed and explore how the differing personalities of the red team attacker agent introduce the subtle variations to the attack patterns. Detailed information on each MITRE ATT&CK tactic and technique can be found at the official MITRE ATT&CK website (MITRE 2013). Additionally, how each technique was used in real cyberattack incidents can be found at reports available on The DFIR Report (The DFIR Report 2021; The DFIR Report 2023).

4.1 Phase 1: Pre-Attack

In the pre-attack phase, prior to initiating a phishing campaign, red team attacker agents meticulously gather information to refine their social engineering strategies, making them more convincing. They engage in comprehensive reconnaissance of the target, which involves collecting data on individual end users **(T1589)** from a variety of sources. Detailed information about the target organization **(T1591)**, as well as information on the websites it operates **(T1594)**, are also acquired. Following this intelligence gathering, the attackers proceed to prepare the necessary malware **(T1588.001)** and tools **(T1588.002)** for the phishing

campaign. This preparation includes the spear-phishing emails for each end user agent within the target organization, the Word documents embedded with malicious macros, MEGA cloud storage, and RClone.

4.2 Phase 2: Infiltration & Establishment

During the Infiltration & Establishment phase, the red team attacker launches the attack by sending spear-phishing emails to end user agents within the virtual organization **(T1566.001)**. Upon receipt of the spear-phishing email, a number of end user agents are deceived into downloading and opening the attached Word document **(T1204.002)**. The likelihood of each end user agent falling for the spear-phishing attempt is influenced by their phishing susceptibility, which is calculated based on individual human factors as outlined in Section 5.1. After an end user agent executes attached malicious file, its computing device becomes compromised, enabling the red team attacker to gain access. Immediately upon accessing a device, the attacker employs PowerShell **(T1059.001)** to establish persistence by setting up a scheduled task **(T1053)**. Subsequently, the attacker proceeds to gather system information of the compromised computing device **(T1082)** and information about the current user **(T1033)**. The attacker then scans various directories to identify the locations of specific files and folders it targets to exfiltrate.

4.3 Phase 3: Exploitation

In the Exploitation phase, the red team attacker agent endeavors to fulfill its cyberattack objectives. This phase marks a pivotal moment where the attack's behavior pattern diverges, influenced by the attacker agent's unique personality (cautious vs. reckless). This phase is composed of four different steps. Initially, the attacker agent aggregates all targeted data into a centralized location **(T1074.001)**. Considering that transferring substantial volumes of data might trigger alerts due to its anomalous nature, cybercriminals typically impose a speed restriction on the data aggregation process. For this experiment, the data collection speed limit is set at 10MB/s. Subsequently, the attacker compresses all collected data into a zip file **(T1560.001)**. The third step involves the exfiltration of the compressed zip file to the MEGA cloud storage service **(T1567.002)**. Finally, the attacker eliminates all traces of the collected and compressed data **(T1485)**.

Two distinct red team cyber attacker agents (Cautious vs. Reckless) employ four techniques in different manners. Trimpop et al. note that individuals scoring higher in cautiousness on the Tension Risk Adventure Inventory (TRAI) (Keinan, Meir, and Gome-Nemirovsky 1984) tend to exhibit lower risk-taking behaviors (Trimpop, Kerr, and Kirkcaldy 1998). Leveraging this empirical evidence, we configured the attack pattern of the cautious-type red team cyber attacker agent in the Cyberattack Campaign Designer to partition the goal of exfiltrating 100GB of data into ten separate sessions. In each session, the agent gathers 10GB of data, compresses it, exfiltrates the compressed data, and then deletes both the collected and compressed data. As illustrated in Figure 3, this process is repeated across ten sessions to fully accomplish its cyberattack objective. The primary rationale behind partitioning the goal is to mitigate the risk of detection by the organization's IT security or cybersecurity software during the red team attacker's infiltration and exploitation efforts. Without segmenting the objective, there's a possibility that the whole target data could be fully collected and compressed but fail to be exfiltrated due to detection prior to the exfiltration process. To circumvent this scenario and maximize goal achievement, cautious-type cyberattackers opt for a partitioned strategy, aiming to reduce the likelihood of full mission compromise by distributing the attack into manageable segments. Conversely, the reckless-type red team cyber attacker agent adopts a more direct and riskier approach, opting to collect the entire 100GB of data, compress it, exfiltrate, and delete all the data in a single operation. According to Trimpop et al.'s study, individuals who score higher in recklessness on the TRAI (Keinan, Meir, and Gome-Nemirovsky 1984) are more inclined towards thrill and adventure-seeking and demonstrate greater risk-taking behaviors (Trimpop, Kerr, and Kirkcaldy 1998).

If the red team attacker agent successfully exfiltrates the entire 100GB of data from the compromised computing device, it concludes the phishing campaign and terminates the connection with that device.

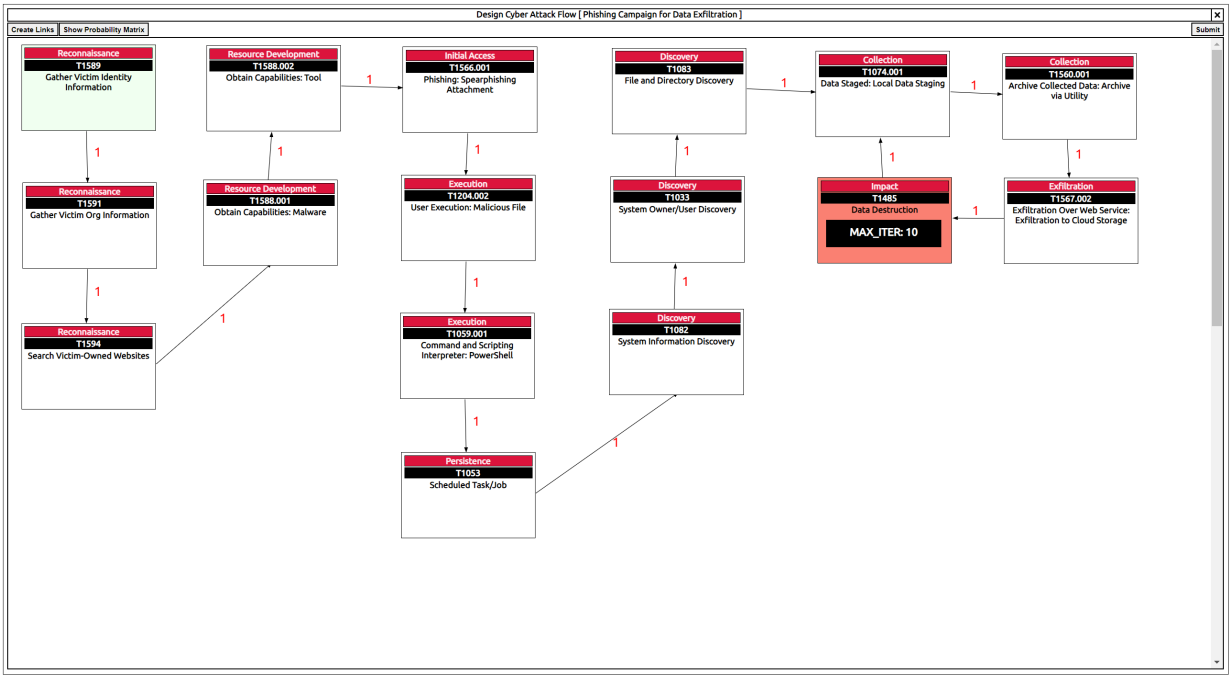

Figure 3: Phishing campaign for data exfiltration by cautious type red team attacker agent.

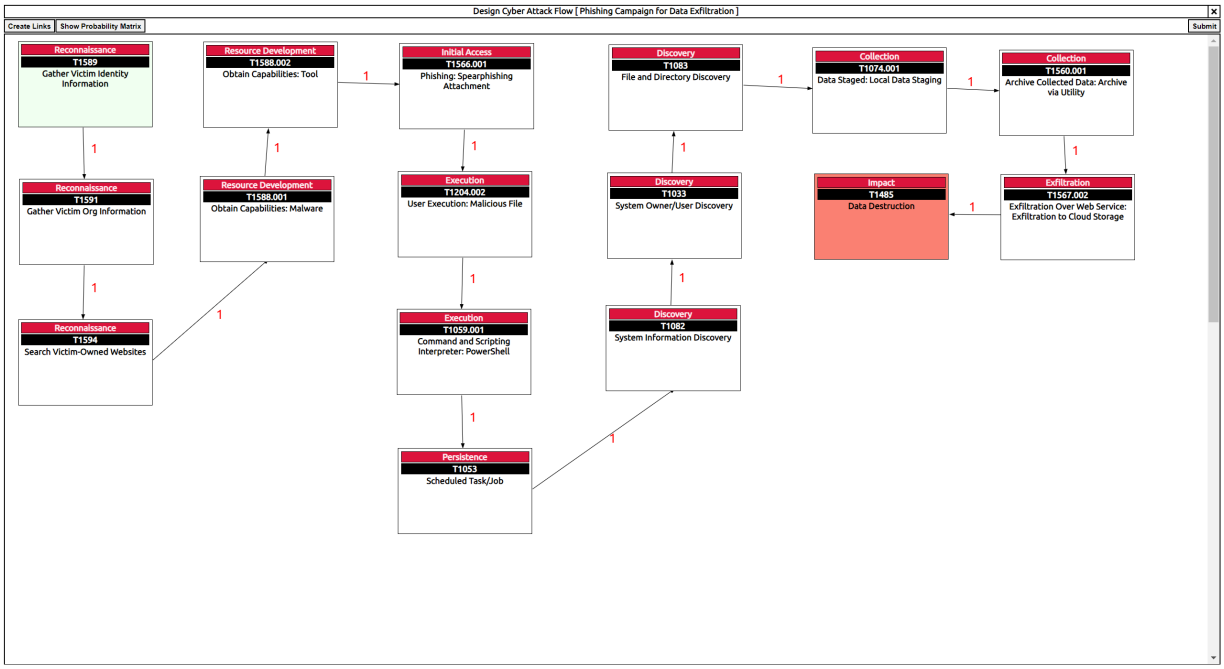

Figure 4: Phishing campaign for data exfiltration by reckless type red team attacker agent.

5 ORGANIZATION MODEL

We employed the OSIRIS framework (Shin et al. 2023a; Shin et al. 2023b) to construct a virtual target organization. OSIRIS offers an environment for creating organizations that feature end-user agents characterized by distinct work and behavior patterns, social networks, and assigned devices (Shin et al. 2022). Moreover, it accurately captures each end-user agent's susceptibility to phishing by incorporating specific human factors assigned to each agent and applying a dedicated phishing susceptibility equation (Shin et al. 2023).

5.1 End User Agent

Similar to prior work using the OSIRIS framework (Shin, Carley, and Carley 2024), we replicated a software company comprising 235 employees that were subjects in Eftimie et al.'s empirical study (Eftimie, Moinescu, and Răcuciu 2022). Each end-user agent is assigned specific attributes including age, gender, and values across the Big Five Personality traits (Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism) (John and Srivastava 1999). These attributes serve as inputs to determine the unique phishing susceptibility of each end-user agent, utilizing the logistic regression equation provided by Eftimie et al (Eftimie, Moinescu, and Răcuciu 2022).

As highlighted in the previous OSIRIS publication (Shin, Carley, and Carley 2024), since Eftimie et al. did not disclose the demographic and personality details of each individual employee, this information for each simulated agent is generated randomly, utilizing the mean and standard deviation for each human factor derived from Eftimie et al.'s study. Naturally, this approach introduces a discrepancy in the overall number of employees tricked to phishing in the simulation compared to the actual findings, 14.3%, in Eftimie et al.'s work (Eftimie, Moinescu, and Răcuciu 2022). However, through the application of a calibration technique (Carley 1996), the overall phishing susceptibility rate was adjusted to 14.22%, aligning closely with the empirical outcome (Shin, Carley, and Carley 2024).

5.2 IT Security Agent

Cyber-FIT incorporates defender agents tasked with regularly scrutinizing device vulnerabilities within the organization and repairing compromised devices (Dobson and Carley 2017). In this study, we deploy these defender agents as IT security agents within the virtual organization in OSIRIS (Shin, Carley, and Carley 2024). Prior to deploying the IT security agents, it is required to assign their inspection frequency and the success rate of these inspections. Once deployed with these parameters, the IT security team regularly inspects each end user agent's computing device agent. If a device is discovered to have been compromised by a cybercriminal agent, the IT security agent fixes the issue and disconnects the cybercriminal agent's access to the compromised device.

6 MODEL VALIDATION

In this section, we validate the cyberattack process design, particularly focusing on techniques in the Exploitation Phase (Section 4.3) that significantly affect the severity of cyberattack damage. The evaluation of each cyberattack technique's speed was conducted on an HP Laptop equipped with an Intel Core i9 13900H processor, an NVIDIA GeForce RTX 4060 graphics card, and 64GB of RAM. Initially, we conducted a pilot study involving 10 trials to calculate the initial mean and standard deviation. We then determined the appropriate sample size for each validation case using a 95% confidence level and a 10% margin of error. When the required sample size exceeded 10, additional trials were conducted, and the average time to complete each attack technique was recalculated.

T1074.001 - Data Staged: Local Data Staging: To accurately measure the duration required to transfer data to a central directory with 10MB/s speed limit, we executed ten trials for moving both 10GB and 100GB files. On average, transferring 10GB files at this speed limit took approximately 1042 seconds

(9.83MB/s), with a standard deviation of 1.16 seconds. Similarly, for 100GB files under the same speed constraint, the average time was approximately 10458 seconds (9.79MB/s), accompanied by a standard deviation of 86.46 seconds.

T1560.001 - Archive Collected Data: Archive via Utility: To accurately determine the time required for this attack technique, we conducted ten trials measuring the duration required to compress files sizes of 10GB and 100GB. On average, compressing 10GB files resulted in a size of 9.49GB and took approximately 276 seconds, with a standard deviation of 43.65 seconds. Compression of 100GB files typically reduced them to 94.9GB and required roughly 4020 seconds on average, with a standard deviation of 166.94 seconds.

T1567.002 - Exfiltration Over Web Service: Exfiltration to Cloud Storage: To precisely measure the time required to exfiltrate compressed data, we conducted ten trials to measure the time it takes to upload a 9.49GB compressed file to the MEGA cloud. On average, uploading this 9.49GB file to MEGA cloud took 3952 seconds, with a standard deviation of 252.74 seconds. Given the stable upload speed observed and considering the 20GB storage limit of MEGA's free version, we assumed that uploading a 94.9GB compressed file would take approximately ten times longer than the 9.49GB file, amounting to 39520 seconds in total.

T1485 - Data Destruction: We conducted seventy trials to accurately measure the time required to destroy both original and compressed data together. For the case of eliminating a 10GB file plus its compressed version of 9.49GB, the average time taken was 2 seconds, with a standard deviation of 0.337 seconds. Similarly, for the case of a 100GB file plus its compressed version of 94.9GB, the average destruction time was a 37 seconds, with a standard deviation of 16.372 seconds.

The time required to execute and complete each MITRE ATT&CK (Strom et al. 2018) technique, measured during the validation process, is accurately reflected in the red team attacker agent's completion time for each technique. Given that the OSIRIS framework's minimum time unit is one minute (Shin, Carley, and Carley 2024), time measured in seconds are converted to minutes and rounded up to the nearest whole minute.

7 VIRTUAL EXPERIMENTS

In this section, we present our plan for the virtual experiment. We will leverage the virtual organization established in Section 5 as the target for our cyberattack simulation. The experiment will feature two distinct scenarios, each defined by the personality of the attacker agents: a cautious-type red team attacker agent (Figure 3) or a reckless-type red team attacker agent (Figure 4). During the simulation, each agent will execute its own spearphishing campaign for data exfiltration described in Section 4. Additionally, to assess how the number of IT security agents protecting the target organization affects the level of mitigation against each cyberattack campaign, we will vary the number of IT security agents in each simulation, ranging from 0 to 10. We assumed that IT security agents randomly select and inspect a computing device within the virtual organization at 30 minutes intervals. It is also assumed that all deployed IT security agents have the capability to identify a compromised computing device, disconnect the attacker's access, and repair it. These variations result in a total of 22 unique simulation cases (Two types of attacker personalities and eleven levels of number of IT security agents). In every simulation case, we measured the average amount of data, in gigabytes, that was exfiltrated and destroyed for each compromised computing device. Given that data destruction occurred immediately following data exfiltration and was completed within one minute, the volumes of data exfiltrated and destroyed were equivalent across all 22 simulation cases. We initially conducted 100 simulations for each scenario as a pilot study, resulting in a total of 2200 simulations. To determine the appropriate number of simulations for our experiment, we applied the statistical guidelines introduced by Ritter et al (Ritter et al. 2011). Under a 95% power level and an alpha of 0.05, our analysis showed that no scenario required more than 100 simulations to achieve statistical significance and power. Therefore, based on 100 simulations for each scenario, we calculated the mean and standard deviations of the average amount of data exfiltrated and destroyed per compromised computing device, measured in gigabytes. These results are presented in Figure 5.

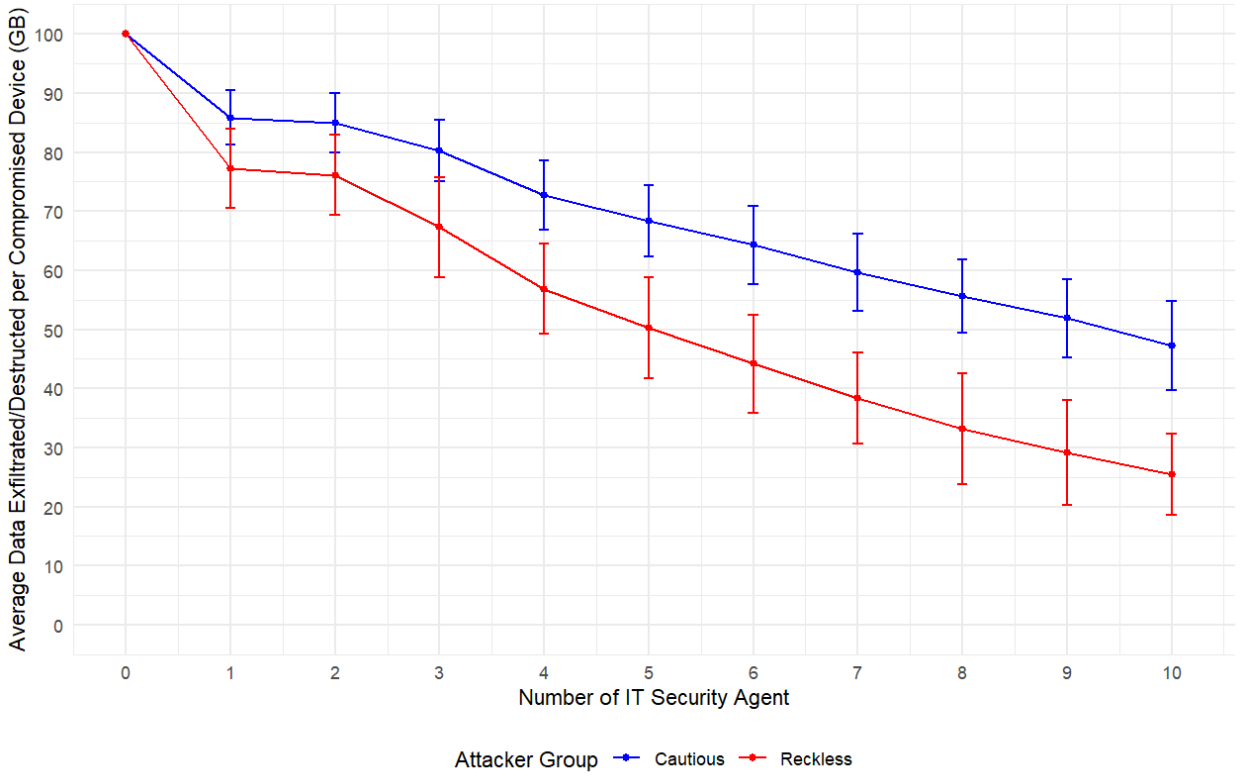

Figure 5: Virtual experiment result.

7.1 Results and Analysis

Initially, we computed the average number of end-user agents deceived by spear-phishing emails distributed during Phase 2 of the cyberattack campaign (Section 4.2) across 2200 simulation cases. The results yielded a mean of 34 employees with a standard deviation of 4.6. Given that the virtual organization is composed of 235 employees, the overall rate of employees tricked by phishing is approximately 14.5%, closely aligning with empirical findings (14.3%) (Eftimie, Moinescu, and Răcuciu 2022).

Given our assumption that the red cyber team's attacker agent aims to exfiltrate and destroy 100GB of data from each compromised computing device (Section 4), the maximum potential average of exfiltrated and destroyed data is capped at 100GB. This scenario occurs in the simulation case where no IT security agents are deployed within the virtual organization. Besides this baseline scenario, from Figure 5, we can observe that simulations involving cautious attacker agents consistently result in higher volumes of data exfiltration and destruction compared to those involving reckless attacker agents, given an equal number of IT security agents. We conducted t-tests to compare the simulation outcomes between the cautious and reckless attacker scenarios across different levels of IT security agents, at a 99% confidence level. Except for the scenario with zero IT security agent, all other comparisons yielded extremely small p-values, close to zero. This statistically significant difference confirms that the personality type of the attacker (cautious vs. reckless) is a crucial explanatory variable for estimating the amount of data exfiltrated. Furthermore, the disparity in the magnitude of damage (measured by the amount of data exfiltrated and destroyed) between scenarios with cautious and reckless attacker agents widens as the number of IT security agents increases. This suggests that as the security environment becomes more robust, the likelihood of a reckless attacker agent successfully completing their mission decreases significantly relative to a cautious attacker agent. Conversely, although simulations with cautious-type attacker agents also display a decreasing trend in the average magnitude of attack damage as the number of IT security agents increases, these attacker agents

still partially achieve their objectives. They do so by repeatedly exfiltrating and destroying small quantities of data, rather than attempting to exfiltrate and destroy all collected data in a single, large-scale action at the end. These experiments demonstrate that even when two cyber attackers have identical objectives and employ same techniques in their cyberattack campaigns, differences in their personalities can result in varied patterns of cyberattack campaigns. Consequently, this leads to differences in the magnitude of cyberattack damage.

8 CONCLUSION AND FUTURE WORKS

In this study, we present the Cyberattack Campaign Designer, a tool that assists modelers in designing and simulating various cyberattack campaign scenarios they envision by selecting and linking MITRE ATT&CK techniques (Strom et al. 2018) through a Markov chain (Norris 1998). In this study, we leveraged this tool to design two separate phishing campaigns aimed at data exfiltration from the target organization. Each campaign was designed to reflect two distinct cybercriminal personality type: one cautious and the other reckless. Although the objectives of both campaigns and the MITRE ATT&CK techniques (Strom et al. 2018) employed were identical, the personality of the cybercriminal introduced subtle differences in the attack patterns. Our virtual experiments demonstrated that these subtle differences significantly influenced the magnitude of damage resulting from each campaign. The findings highlight the importance of considering the cybercriminal's personality in evaluating the potential impact of a cyberattack.

This study opens several avenues for future research. In our current work, we focused on how a cybercriminal's personality affects the pattern of attack specifically during the stages of data collection, exfiltration, and destruction on target computing devices. However, the impact of a cybercriminal's personality extends beyond these phases. It can significantly influence the amount of time spent on gathering information about the target user and organization prior to launching an attack, the development of backup strategies to maintain access to the target system, and the formulation of tactics to evade detection. Moreover, personality traits can shape the cybercriminal's decision-making process when faced with obstacles, such as the failure of a particular cyberattack technique. Depending on their personality, some cybercriminals may persist with the same techniques, while others might quickly switch to alternative methods to achieve their objectives. In the future, we aim to incorporate diverse human dynamics from the fields of psychology and decision science into our models. This will enable us to simulate a more authentic representation of personality-driven cyberattack patterns in future studies.

We also plan to refine other components of our simulation models. A cyberattack campaign exploits a range of system and human vulnerabilities. To enhance the realism of our simulations, we plan to incorporate vulnerabilities list from the CASOS technical report (Shin et al. 2023). This approach will allow us to detail which vulnerabilities are targeted at each stage of a cyberattack campaign and how different defense strategies can mitigate these vulnerabilities, thereby reducing the potential damage caused by cyberattacks during the simulation. Furthermore, we plan to make IT security agents' behavior in our simulations more realistic by basing their defense strategies on the MITRE D3FEND framework (Kaloroumakis and Smith 2021). This approach will allow for a more detailed and authentic representation of how cybersecurity defenses are implemented and operated within organizations in response to threats.

ACKNOWLEDGMENTS

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported in part by the Minerva Research Initiative under Grant #N00014-21-1-4012 and by the Center for Computational Analysis of Social and Organizational Systems (CASOS) at Carnegie Mellon University. The views and conclusions are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Office of Naval Research or the US Government.

REFERENCES

Bada, M. and J. R. Nurse. 2023. "Exploring Cybercriminal Activities, Behaviors, and Profiles". In *Applied Cognitive Science and Technology: Implications of Interactions Between Human Cognition and Technology*, 109–120. Springer.

Carley, K. M. 1996. "Validating computational models". *CASOS technical report (2017)*.

Carley, K. M. 2002. "Computational organizational science and organizational engineering". *Simulation Modelling Practice and Theory* 10(5-7):253–269.

Carley, K. M. and D. M. Svoboda. 1996. "Modeling organizational adaptation as a simulated annealing process". *Sociological methods & research* 25(1):138–168.

Dobson, G. 2022. *Cyber-Forces, Interactions, Terrain: An agent-based framework for simulating cyber team performance*. Ph. D. thesis, Carnegie Mellon University.

Dobson, G., A. Rege, and K. Carley. 2018. "Informing active cyber defence with realistic adversarial behaviour". *Journal of Information Warfare* 17(2):16–31.

Dobson, G. B. and K. M. Carley. 2017. "Cyber-FIT: an agent-based modelling approach to simulating cyber warfare". In *Social, Cultural, and Behavioral Modeling: 10th International Conference, SBP-BRiMS 2017, Washington, DC, USA, July 5-8, 2017, Proceedings 10*, 139–148. Springer.

Dobson, G. B. and K. M. Carley. 2018. "A computational model of cyber situational awareness". In *Social, Cultural, and Behavioral Modeling: 11th International Conference, SBP-BRiMS 2018, Washington, DC, USA, July 10-13, 2018, Proceedings 11*, 395–400. Springer.

Dobson, G. B. and K. M. Carley. 2020. "Towards Agent Validation of a Military Cyber Team Performance Simulation". In *Social, Cultural, and Behavioral Modeling: 13th International Conference, SBP-BRiMS 2020, Washington, DC, USA, October 18–21, 2020, Proceedings 13*, 182–191. Springer.

Dobson, G. B. and K. M. Carley. 2021. "Cyber-FIT Agent-Based Simulation Framework Version 4". *Center for the Computational Analysis of Social and Organizational Systems*.

Eftimie, S., R. Moinescu, and C. Răcuciu. 2022. "Spear-phishing susceptibility stemming from personality traits". *IEEE Access* 10:73548–73561.

John, O. P. and S. Srivastava. 1999. "The Big-Five trait taxonomy: History, measurement, and theoretical perspectives".

Kaloroumakis, P. E. and M. J. Smith. 2021. "Toward a knowledge graph of cybersecurity countermeasures". *The MITRE Corporation* 11.

Kavak, H., J. J. Padilla, D. Vernon-Bido, S. Y. Diallo, R. Gore and S. Shetty. 2021. "Simulation for cybersecurity: state of the art and future directions". *Journal of Cybersecurity* 7(1):tyab005.

Keinan, G., E. Meir, and T. Gome-Nemirovsky. 1984. "Measurement of risk takers' personality". *Psychological Reports* 55(1):163–167.

Kranenbarg, M. W., J.-L. Van Gelder, A. J. Barends, and R. E. de Vries. 2023. "Is there a cybercriminal personality? Comparing cyber offenders and offline offenders on HEXACO personality domains and their underlying facets". *Computers in human behavior* 140:107576.

Lee, K. and M. C. Ashton. 2004. "Psychometric properties of the HEXACO personality inventory". *Multivariate behavioral research* 39(2):329–358.

Macal, C. M. and M. J. North. 2009. "Agent-based modeling and simulation". In *Proceedings of the 2009 winter simulation conference (WSC)*, 86–98. IEEE.

MITRE 2013. "MITRE ATTCK®". https://attack.mitre.org/.

Norris, J. R. 1998. *Markov chains*. Number 2. Cambridge university press.

Ribeiro, L., I. S. Guedes, and C. S. Cardoso. 2024. "Which factors predict susceptibility to phishing? An empirical study". *Computers & Security* 136:103558.

Ritter, F. E., M. J. Schoelles, K. S. Quigley, and L. C. Klein. 2011. "Determining the number of simulation runs: Treating simulations as theories by not sampling their behavior". *Human-in-the-loop simulations: Methods and practice*:97–116.

Shin, J., K. M. Carley, and L. R. Carley. 2023. "Integrating Human Factors into Agent-Based Simulation for Dynamic Phishing Susceptibility". In *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*, 169–178. Springer.

Shin, J., L. R. Carley, and K. M. Carley. 2024. "Simulation-Based Study on False Alarms in Intrusion Detection Systems for Organizations Facing Dual Phishing and DoS Attacks". In *2024 Annual Modeling and Simulation Conference (ANNSIM)*. Forthcoming.

Shin, J., L. R. Carley, G. B. Dobson, and K. M. Carley. 2022. "Leveraging OSIRIS to simulate real-world ransomware attacks on organization". In *2022 Winter Simulation Conference (WSC) Poster Session*.

Shin, J., L. R. Carley, G. B. Dobson, and K. M. Carley. 2023a. "Beyond Accuracy: Cybersecurity Resilience Evaluation of Intrusion Detection System against DoS Attacks using Agent-based Simulation". In *2023 Winter Simulation Conference (WSC)*, 118–129. IEEE.

Shin, J., L. R. Carley, G. B. Dobson, and K. M. Carley. 2023b. "Modeling and Simulation of the Human Firewall Against Phishing Attacks in Small and Medium-Sized Businesses". In *2023 Annual Modeling and Simulation Conference (ANNSIM)*, 369–380. IEEE.

Shin, J., G. B. Dobson, K. M. Carley, and L. R. Carley. 2022. "OSIRIS: Organization Simulation in Response to Intrusion Strategies". In *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*, 134–143. Springer.

Shin, J., G. B. Dobson, L. R. Carley, and K. M. Carley. 2023. "Revelation of System and Human Vulnerabilities Across MITRE ATT&CK Techniques with Insights from ChatGPT". *CASOS technical report (2023)*.

Strom, B. E., A. Applebaum, D. P. Miller, K. C. Nickels, A. G. Pennington and C. B. Thomas. 2018. "MITRE ATT&CK: Design and Philosophy". Technical Report 10AOH08A-JC, The MITRE Corporation, McLean, VA.

The DFIR Report 2021, Nov. "Continuing the bazar ransomware story". https://thedfirreport.com/2021/11/29/continuing-the-bazar-ransomware-story/.

The DFIR Report 2023, Feb. "Collect, exfiltrate, sleep, repeat". https://thedfirreport.com/2023/02/06/collect-exfiltrate-sleep-repeat/.

Trimpop, R. M., J. H. Kerr, and B. Kirkcaldy. 1998. "Comparing personality constructs of risk-taking behavior". *Personality and Individual differences* 26(2):237–254.

Vernon-Bido, D., J. J. Padilla, S. Y. Diallo, H. Kavak and R. J. Gore. 2016. "Towards modeling factors that enable an attacker.". In *SummerSim*, 46.

Yadav, T. and A. M. Rao. 2015. "Technical aspects of cyber kill chain". In *Security in Computing and Communications: Third International Symposium, SSCC 2015, Kochi, India, August 10-13, 2015. Proceedings 3*, 438–452. Springer.

AUTHOR BIOGRAPHIES

JEONGKEUN SHIN is a Ph.D. student in the Department of Electrical and Computer Engineering at Carnegie Mellon University. He is a member of the Center for Computational Analysis of Social and Organization Systems (CASOS). His research includes modeling and simulation of human and organizational behaviors relevant to cybersecurity. He holds a Bachelor's degree in Computer Science from the University of Michigan and a Master's degree in Electrical and Computer Engineering from Carnegie Mellon University. His email address is jeongkes@andrew.cmu.edu.

GEOFFREY B. DOBSON is a Systems Scientist at the Center for Computational Analysis of Social and Organizational Systems at Carnegie Mellon University's School of Computer Science. His research focuses on modeling and simulating the human behavioral and social aspects of cyber conflict. He is an officer in the United States Air Force Reserve stationed at the Air Force Research Laboratory, Wright-Patterson Air Force Base, OH where he oversees a research portfolio focused on human performance in cyber missions. His email address is gdobson@cs.cmu.edu.

L. RICHARD CARLEY received an S.B. in 1976, an M.S. in 1978, and a Ph.D. in 1984, all from the Massachusetts Institute of Technology. He is the professor of Electrical and Computer Engineering Department at Carnegie Mellon University (CMU) in Pittsburgh, Pennsylvania. Dr. Carley's research interests include analog and RF integrated circuit design in deeply scaled CMOS technologies, and novel micro-electromechanical and nano-electro-mechanical device design and fabrication. For the past several years, Dr. Carley has studied the design of efficient RF Power Amplifiers in advanced BiCMOS technologies. Dr. Carley has been granted 27 patents, authored or co-authored over 250 technical papers, and authored or co-authored over 20 books and/or book chapters. He has won numerous awards including Best Technical Paper Awards at both the 1987 and the 2002 Design Automation Conference (DAC), a Most Influential Paper award from DAC, and a Best Panel Session award at ISSCC in 1993. In 1997, Dr. Carley co-founded the analog electronic design automation startup, Neolinear, which was acquired by Cadence in 2004. His email address is lrc@andrew.cmu.edu.

KATHLEEN M. CARLEY (H.D. University of Zurich, Ph.D. Harvard, S.B. MIT) is a Professor of Societal Computing, Software and Societal Systems Department (S3D), Carnegie Mellon University; Director of the Center for Computational Analysis of Social and Organizational Systems (CASOS), Director of the Center for Informed Democracy and Social Cybersecurity (IDeaS), and CEO of Netanomics. Her research blends computer science and social science to address complex real world issues such as social cybersecurity, disinformation, disease contagion, disaster response, and terrorism from a high dimensional network analytic, machine learning, and natural language processing perspective. She and her groups have developed network and simulation tools, such as ORA, that can assess network and social media data. Her email address is kathleen.carley@cs.cmu.edu.