Proceedings of the 2025 Winter Simulation Conference
E. Azar, A. Djanatliev, A. Harper, C. Kogler, V. Ramamohan, A. Anagnostou, and S. J. E. Taylor, eds.

A BASELINE SIMULATION OF HYBRID MISINFORMATION AND SPEARPHISHING CAMPAIGNS IN ORGANIZATIONAL NETWORKS

Jeongkeun Shin[1], Han Wang[2], L. Richard Carley[1], and Kathleen M. Carley[1]

[1]CASOS Center, Carnegie Mellon University, Pittsburgh, PA, USA
[2]Information Systems Program, Carnegie Mellon University, Pittsburgh, PA, USA

ABSTRACT

This study presents an agent-based simulation that examines how pre-attack misinformation amplifies the effectiveness of spearphishing campaigns within organizations. A virtual organization of 235 end user agents is modeled, each assigned unique human factors such as Big Five personality traits, fatigue, and job performance, derived from empirical data. Misinformation is disseminated through Facebook, where agents determine whether to believe and spread false content using regression models from prior psychological studies. When agents believe misinformation, their psychological and organizational states degrade to simulate a worst-case scenario. These changes increase susceptibility to phishing emails by impairing security-related decision-making. Informal relationship networks are constructed based on extraversion scores, and network density is varied to analyze its effect on misinformation spread. The results demonstrate that misinformation significantly amplifies organizational vulnerability by weakening individual and collective cybersecurity-relevant decision-making, emphasizing the critical need to account for human cognitive factors in future cybersecurity strategies.

1 INTRODUCTION

The nature of modern warfare has fundamentally changed compared to the past. The sound of the battlefield is no longer filled solely with gunfire and artillery, and strong firepower or military strategy alone no longer guarantees victory. Instead, warfare is increasingly being shaped by invisible flows of information that begin even before the first shot is fired. These “silent wars” are becoming more prominent, where an adversary’s internal cohesion is disrupted and morale is weakened, ultimately diminishing the will to fight without the need for physical confrontation. In today’s highly developed digital world, false information is being widely disseminated through social networking services and messaging platforms with the intention of promoting specific narratives or undermining the legitimacy of opposing forces. These operations go beyond simple opinion manipulation. They are actively used to create confusion in military decision-making, provoke fear among civilians, and incite distrust among targeted groups, serving as effective tactical tools in modern information warfare.

This strategy is also likely to be widely adopted by cybercriminals looking to target private companies in the near future. Traditional cyberattacks have mainly focused on exploiting various system vulnerabilities identified within the target organization’s computing infrastructure. However, with the continued advancement of security systems and the proactive efforts of operating system and network developers, it has become increasingly difficult to breach systems through technical vulnerabilities alone. As a result, attackers have increasingly turned to social engineering strategies that aim to bypass technical defenses by manipulating users to make poor decisions or take unsafe actions. This approach, which targets human cognitive vulnerabilities, has emerged as an effective alternative or complement to purely technical attacks. In response, many organizations have implemented various measures such as cybersecurity training and phishing simulations in an effort to defend against social engineering threats.

979-8-3315-8726-0/25/$31.00 ©2025 IEEE

However, even such training and preventive measures can be rendered ineffective in the face of persuasive social cyber attacks (Carley 2020). For example, the spread of false information through social networks or messaging platforms, such as scandals involving organizational leaders, rumors of financial instability, or large-scale layoffs following mergers, can cause psychological unrest and anxiety among employees. This emotional disruption impairs individual judgment and weakens their vigilance toward security, significantly increasing their susceptibility to social engineering attacks. As a result, attackers are more likely to succeed in breaching the organization by exploiting these vulnerabilities through phishing attempts, malicious link clickbait, or credential theft. In particular, the rapid advancement of artificial intelligence technologies in recent years has fundamentally transformed the way misinformation is created and disseminated. With the increasing sophistication of deepfake technologies, it has become possible to generate visual and audio content that is nearly indistinguishable from reality. Moreover, large language models enable the effortless creation of highly persuasive and professionally written false narratives. This has led to the emergence of misinformation that is far more credible and convincing than ever before. When such content infiltrates an organization, it becomes increasingly difficult for individual members to critically evaluate or verify its authenticity. Consequently, AI-generated misinformation not only increases the likelihood of successful social engineering attacks, but also poses a growing threat by undermining the effectiveness of traditional security education.

To address this emerging threat, this study leverages an agent-based modeling and simulation approach to explore how hybrid cyberattacks, combining misinformation with traditional technical and social engineering methods, can amplify organizational damage. We model a medium-sized organization, incorporating a range of human factors, and construct an informal relationship network among employees based on social networking service interactions. Through this simulated network, we analyze how misinformation spreads, how many individuals are influenced by it, and to what extent such misinformation, introduced prior to a cyberattack, can impact the overall scale of damage. The goal is to quantitatively explore the human vulnerability amplification effect caused by pre-attack misinformation in organizational contexts.

2 RELATED WORKS

Agent-based modeling and simulation (ABMS) (Macal and North 2009) has been widely used in the field of cybersecurity to replicate security problems within virtual environments and to experiment with various defense strategies to identify optimal solutions (Kavak et al. 2021; Vestad and Yang 2024). Well-designed and validated simulation models can be effectively leveraged to systematically experiment with a wide range of complex and unconventional scenarios that are difficult to test in real-world environments (Carley et al. 2006). Through such simulations, it becomes possible to identify optimal cybersecurity solutions. This approach has been widely adopted to replicate and analyze various cybersecurity situations that are otherwise difficult to experiment with in reality due to practical, financial, or ethical constraints, enabling low-cost and risk-free exploration within a virtual environment.

Research on ABMS in the field of cybersecurity has evolved in line with the changing nature of cyberattacks. During the period when technically driven attacks such as Distributed Denial of Service (DDoS) were predominant, ABMS was primarily used to replicate realistic computer network infrastructures and to model how DDoS attacks operate within those environments (Chen, Longstaff, and Carley 2004). Through such simulations, ABMS was leveraged to analyze how collaborative or distributed defense strategies among networked devices could effectively respond to DDoS attacks (Gorodetski et al. 2001). In more recent years, when simulating technical attacks such as DDoS, ABMS has been commonly employed to assess how intrusion detection systems (IDS), developed using various machine learning algorithms, perform once deployed within an organization. Specifically, these simulations aim to evaluate the extent to which such systems can mitigate damage and enhance organizational resilience (Shin et al. 2023a).

As social engineering became widely adopted in cyberattacks, ABMS began to incorporate not only technical components, such as devices and networks, but also human and organizational behavioral characteristics into simulation models. Blythe et al. implemented human agents within their simulation and modeled how factors such as fatigue and stress influenced behaviors such as ignoring security warnings or underestimating risks (Blythe et al. 2011). Their study examined how these cognitive factors affect the overall cybersecurity posture of an organization. Burns et al. integrated various social science theories to simulate human agents' attitudes toward cybersecurity within organizational contexts. Their research explored how differences in these attitudes impacted susceptibility to phishing attacks and, ultimately, influenced the overall scale of cybersecurity damage (Burns et al. 2017). Shin et al. developed the OSIRIS framework (Shin et al. 2022), in which each human agent was assigned a unique level of phishing susceptibility based on a regression model trained on empirical data (Shin, Carley, and Carley 2024). This approach enabled the simulation to more accurately reflect individual differences in vulnerability to phishing attacks. In addition, Shin et al. extended their model by introducing a mechanism in which human agents' phishing susceptibility dynamically changes over time due to memory recency effects following security training (Shin, Carley, and Carley 2023). They incorporated regression models from multiple empirical studies to simulate dynamic human factors, demonstrating how changes in fatigue, perceived vulnerability, and job performance influence phishing susceptibility and ultimately affect organizational cyberattack outcomes (Shin, Carley, and Carley 2025).

In the current landscape, hybrid attacks that combine technical cyberattacks with social cyberattacks (Carley 2020), such as misinformation campaigns, remain relatively rare. As a result, simulation-based research exploring such integrated future threats is still limited. A notable exception is the work by Padur et al., who simulated a hybrid attack combining DDoS and misinformation. In their model, malicious agents use reinforcement learning to determine when to launch the attack and how long to sustain it, with the goal of undermining user trust in the targeted service and influencing users to switch providers or act irrationally based on false information (Padur, Borrion, and Hailes 2025). While Padur et al. focused on the erosion of external user trust, this study shifts attention inward to the organization itself. Specifically, we examine how socially driven cyberattack campaigns can destabilize internal personnel, increase human vulnerability, and make individuals more susceptible to targeted social engineering attacks. Using agent-based simulation, we examine how this increased susceptibility influences the magnitude of damage in cyberattacks for data exfiltration within the organization. To the best of our knowledge, this is one of the first simulation-based studies to model the impact of socially driven hybrid attacks on internal organizational vulnerabilities.

3 HYBRID MISINFORMATION AND SPEARPHISHING CAMPAIGN

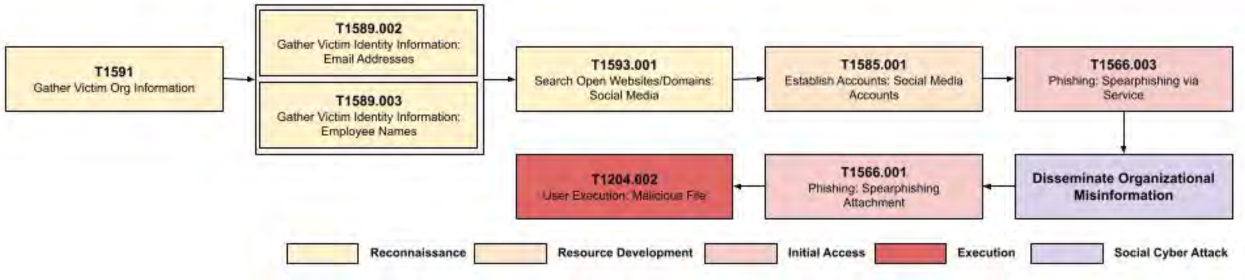
T1591
Gather Victim Org Information
T1589.002
Gather Victim Identity Information: Email Addresses
T1589.003
Gather Victim Identity Information: Employee Names
T1593.001
Search Open Websites/Domains: Social Media
T1585.001
Establish Accounts: Social Media Accounts
T1566.003
Phishing: Spearphishing via Service
Disseminate Organizational Misinformation
T1566.001
Phishing: Spearphishing Attachment
T1204.002
User Execution: Malicious File
Reconnaissance
Resource Development
Initial Access
Execution
Social Cyber Attack

Figure 1: Hybrid cyberattack campaign model.

Figure 1 illustrates the hybrid misinformation and spearphishing campaign simulated in this study. We used the MITRE ATT&CK techniques (Strom et al. 2018) to model the cyberattack scenario. The attack begins with the adversary selecting a target organization and gathering information about it **(T1591)**. The attacker then collects the names and email addresses of end user agents within the organization **(T1589.002, T1589.003)** and identifies their Facebook social media accounts **(T1593.001)**. Next, following a technique previously used by the Windshift group, the attacker creates a fake Facebook account using a fake persona **(T1585.001)**, and this account is used to send friend requests to the target organization's end user agents

(Karim 2018). Once the requests are accepted, the attacker engages with the users over time to build trust **(T1566.003)**. After establishing trust, the attacker begins posting plausible but false organizational information on the fake account. These posts are repeatedly disseminated to the connected end user agents. Once a sufficient level of misinformation exposure is achieved, the attacker sends phishing emails containing malicious attachments to all end user agents in the organization **(T1566.001)**. The campaign concludes as the attacker waits for recipients to mistakenly download and execute the malware **(T1204.002)**.

4 HUMAN AND ORGANIZATION MODEL

Eftimie et al. collected data from 235 employees prior to conducting a phishing campaign, including age, gender, and Big Five personality traits (openness, conscientiousness, extraversion, agreeableness, and neuroticism) (Eftimie, Moinescu, and Răcuciu 2022). After the campaign, they developed a regression model to analyze the relationship between these human factors and phishing susceptibility. We perform our attack simulations on a virtual organization that was replicated, calibrated, and validated by Shin et al. using the OSIRIS framework, based on the organization studied in Eftimie et al.'s empirical research (Shin et al. 2024; Shin et al. 2022). Shin et al. extended the human factor model by adding several dynamic human factors after recognizing that phishing susceptibility is not determined solely by static characteristics (Shin, Carley, and Carley 2025). In our study, we include two of these factors, job performance (Mwaisaka et al. 2019; Rehman et al. 2015; Basit et al. 2017) and fatigue (Tian et al. 2022; Åkerstedt et al. 2014), as illustrated in Figure 2.

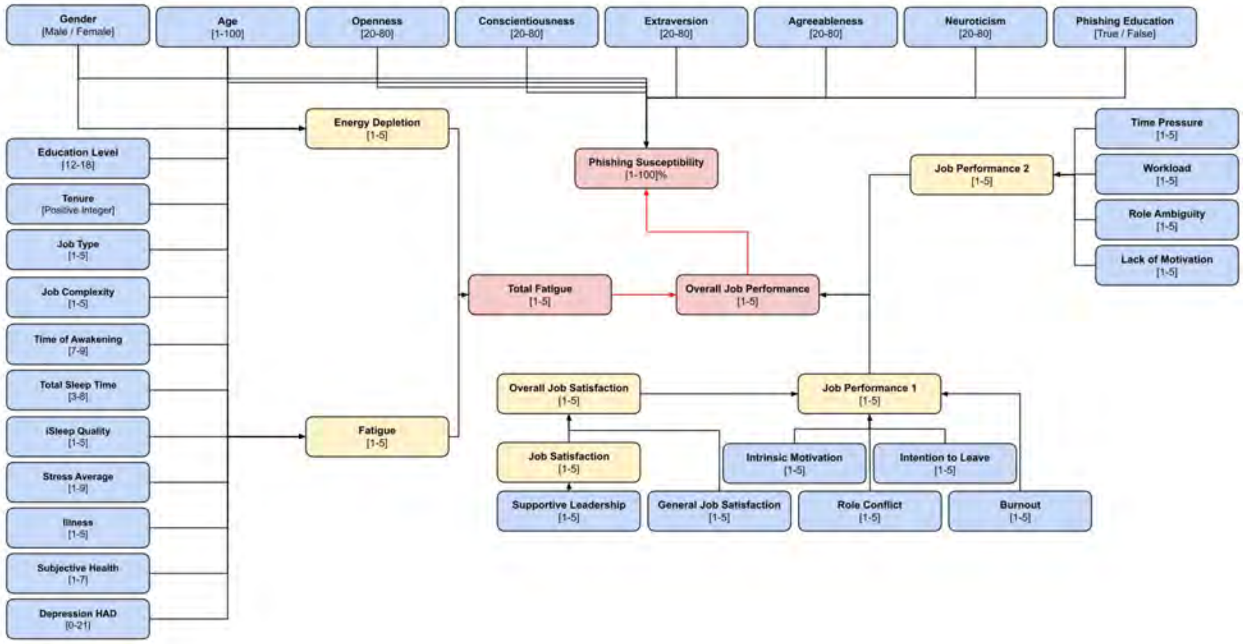

Figure 2: Comprehensive human factor model of end user agents (Shin, Carley, and Carley 2025).

We then configure job performance and fatigue to affect end user agents' decision-making during interactions with suspicious emails, as illustrated in Figure 3. There is no designated cyber defender agent in this virtual organization. Instead, following Shin et al.'s human firewall simulation model (Shin et al. 2023b), we assume that all end user agents are trained to broadcast organization-wide security alerts when encountering a suspicious email. Furthermore, each agent is designed to double-check previously shared alerts even if an externally sourced email is not initially perceived as suspicious. The decision-making processes in both steps are influenced by each end user agent's current job performance and fatigue level.

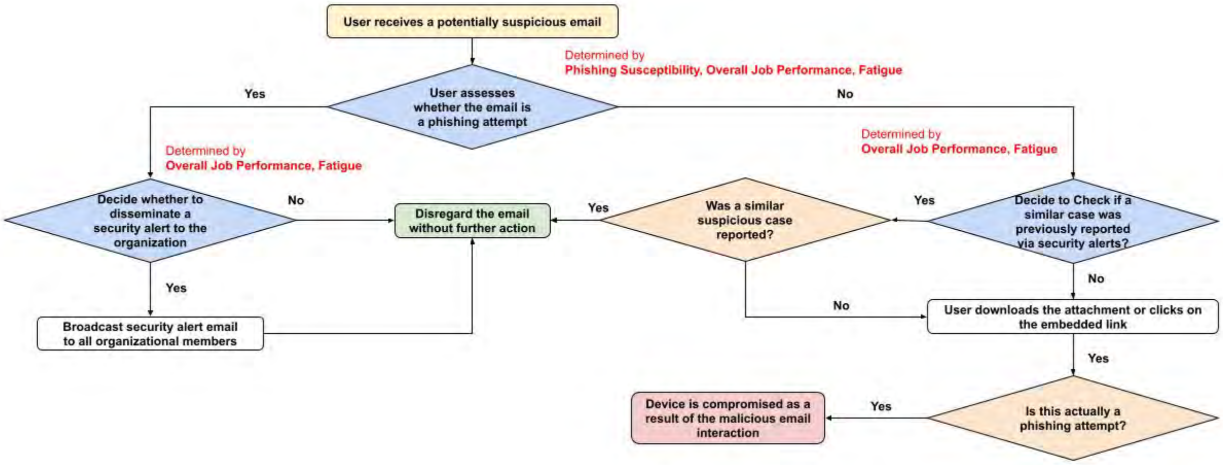
User receives a potentially suspicious email
Determined by
Phishing Susceptibility, Overall Job Performance, Fatigue
Yes
User assesses whether the email is a phishing attempt
No
Determined by
Overall Job Performance, Fatigue
Determined by
Overall Job Performance, Fatigue
Decide whether to disseminate a security alert to the organization
No
Disregard the email without further action
Yes
Was a similar suspicious case reported?
Yes
Decide to Check if a similar case was previously reported via security alerts?
Yes
No
Broadcast security alert email to all organizational members
No
User downloads the attachment or clicks on the embedded link
Yes
Device is compromised as a result of the malicious email interaction
Yes
Is this actually a phishing attempt?

Figure 3: Decision-making process of end user agents in response to suspicious emails.

4.1 Stranger Friend Request Acceptance Model on Social Networking Services

In this paper, we assume that all end user agents use Facebook as their primary social networking service platform to communicate with their friends. As described in the attack campaign in Section 3, the attacker agent creates a Facebook account using a fake persona and sends friend requests to all end user agents within the target organization. This step is essential because, under the Facebook content delivery system, a user's posts are significantly more likely to appear on the feeds of those they are connected with as friends.

To model whether each end user agent accepts or rejects a friend request based on their unique Big Five personality traits, we investigated empirical studies that offer regression models capable of predicting such social decisions. Leow and Wang proposed a regression model that predicts the likelihood of accepting a Facebook friend request from a stranger based on the stranger's gender, the personality cues conveyed in the message, and the degree of alignment between those cues and the recipient's own personality traits (Leow and Wang 2018). However, because our simulation does not model the attacker's persona or the message content based on empirical data, we could not directly apply this study. Instead, we adopted the regression model proposed by Freitag and Bauer (Freitag and Bauer 2016), which estimates the probability of trusting a stranger using Big Five personality traits on a 0-to-1 scale. We operationalize generalized interpersonal trust as the likelihood of a positive response to a stranger's social overture. In this case, the overture refers to accepting a friend request from an unknown individual on Facebook. Although Freitag and Bauer's study was based on a physical-world wallet return scenario (Freitag and Bauer 2016), we argue that similar trust dynamics apply in virtual social interactions involving unfamiliar individuals. Both situations involve deciding whether to trust a stranger by granting access to personal property or information, such as returning a lost wallet or accepting a friend request from an unknown contact. Thus, the wallet return scenario provides a practical measure of generalized interpersonal trust. Freitag and Bauer presented several regression models, but except for one, all include the language region as an independent variable (Freitag and Bauer 2016). Since the language region reflects cultural differences specific to the original sample, it is not directly applicable to our simulation. Therefore, we used the first model, which estimates trust using only the Big Five personality traits. We then converted the original 20-to-80 scale Big Five personality values to a 0-to-1 scale and applied the Freitag and Bauer's regression formula presented in Equation 1 (Freitag and Bauer 2016) to model the probability that each human agent accepts a Facebook friend request from a stranger.

$$\text{Trust in Strangers } (\%) = 22.81 \cdot A - 6.11 \cdot E - 1.75 \cdot N + 16.13 \cdot O - 28.86 \cdot C + 43.88 \tag{1}$$

4.2 Informal Relationship Network

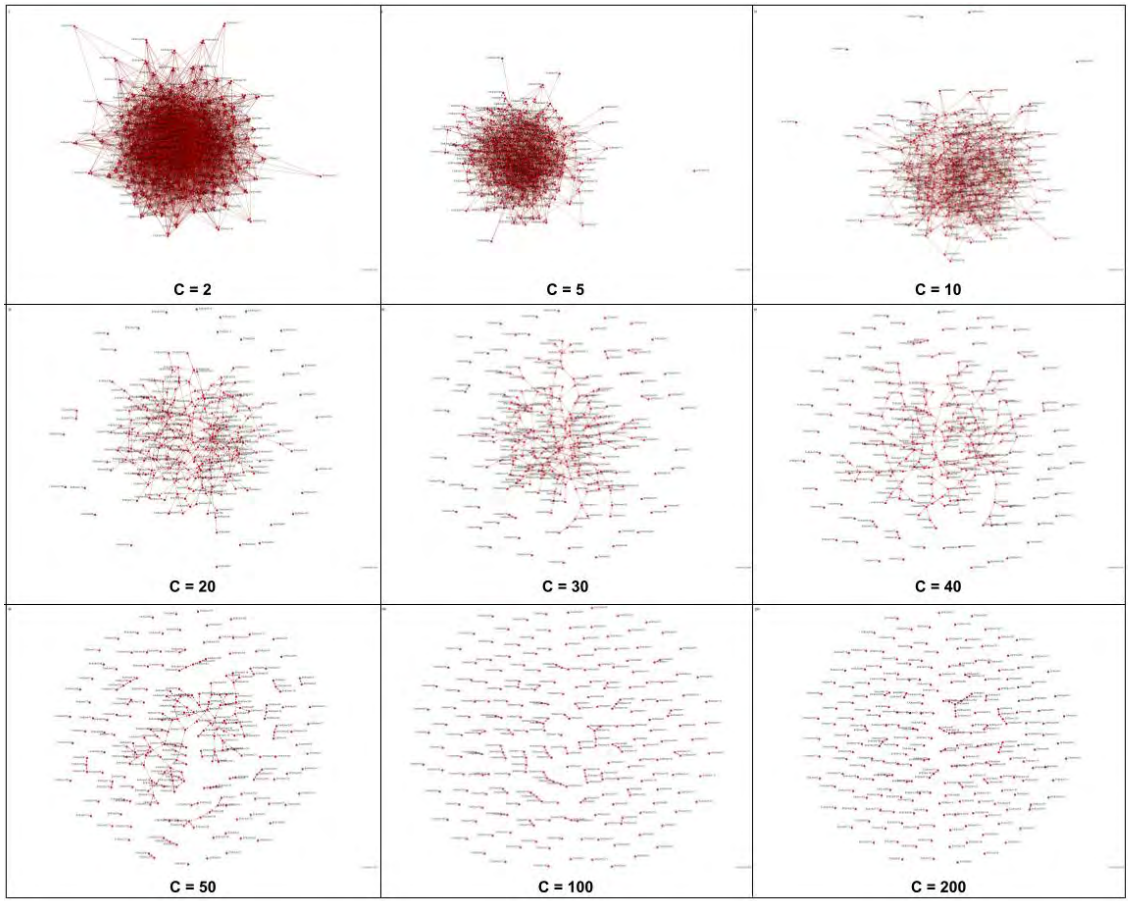

Figure 4: Informal relationship network visualized via ORA (Carley 2014; Altman et al. 2020).

In this study, we construct the informal relationship network within the organization based on each end user agent's Extraversion value. Previous research consistently shows that individuals with higher levels of extraversion tend to form more social connections and maintain broader interpersonal networks. For example, Barrick et al. found that extraverted team members are more likely to develop friendships and interpersonal bonds with their colleagues (Barrick et al. 1998). Similarly, Asendorpf and Wilpers demonstrated that highly extraverted individuals are more successful in making new friends and maintaining larger social networks (Asendorpf and Wilpers 1998). Feiler and Kleinbaum further observed that extraverts, compared to introverts, are significantly more likely to form extensive friendship ties (Feiler and Kleinbaum 2015). These findings suggest that extraversion plays a central role in shaping informal social structures within organizations, making it a suitable basis for modeling relationship networks in agent-based simulations.

Based on this theoretical foundation, we constructed the informal relationship network among the 235 end user agents in our virtual organization by modifying the Erdős–Rényi random network algorithm (Erdos and Rényi 1960). While Erdős and Rényi originally assumed a uniform connection probability p between all pairs of nodes, our model introduces heterogeneity in link formation based on each agent's Extraversion score. Specifically, we first normalized the original 20-to-80 scale Extraversion scores to a 0-to-1 scale. Then, as illustrated in Equation 2, we defined the probability P_{ij} of an informal relationship forming between agents i and j as the product of their normalized Extraversion values. The resulting probability was then divided by a scaling constant c, and the final probability of connection P_{ij}^{final} was calculated as shown in Equation 3.

$$P_{ij} = E_i \times E_j = \left(\frac{x_i - 20}{80 - 20}\right) \times \left(\frac{x_j - 20}{80 - 20}\right) = \frac{(x_i - 20)(x_j - 20)}{3600} \tag{2}$$

$$P_{ij}^{\text{final}} = \frac{P_{ij}}{c} \tag{3}$$

In this study, we varied the scaling constant c to generate informal relationship networks with different densities, in order to analyze how changes in network density affect the spread of misinformation and the subsequent impact of phishing email attacks. Specifically, we applied nine values of c (2, 5, 10, 20, 30, 40, 50, 100, and 200) to construct networks based on Equations 2 and Equation 3. Figure 4 presents visualizations of the resulting networks for each case, created using ORA software (Altman, Carley, and Reminga 2020). Table 1 summarizes the average number of informal relationships per end user agent, the total network density, and the number of isolated nodes for each case. As the value of c increases, both the average number of informal relationships per end user agent and the total network density tend to decrease, while the number of isolated nodes tends to increase.

Table 1: Informal relationship network characteristics under varying values of scaling constant c.

c	**Mean # of Links Per Each End User Agent**	**Total Network Density**	**# of isolated nodes**
2	28.31	0.120967	0
5	11.19	0.047827	1
10	5.69	0.024332	5
20	2.74	0.011711	27
30	1.96	0.008365	42
40	1.64	0.007019	55
50	1.18	0.005055	77
100	0.45	0.001928	147
200	0.30	0.001273	178

4.3 Misinformation Belief & Spread Model

During the simulation, each end user agent who encounters a piece of misinformation on Facebook undergoes two decision-making processes. The first is whether to believe the misinformation. The second is whether to engage with the post (e.g., sharing, liking, commenting, or reacting), thereby potentially spreading it to their Facebook friends. To model the first process, belief in misinformation, we applied the regression model developed by Lai et al., which examines the relationship between Big Five personality traits and rumor belief using data from 11,551 participants (Lai et al. 2020). However, since the intercept value was not reported in their model, we estimated it indirectly. We calculated the weighted sum of the beta coefficients and the mean values of each personality trait, then subtracted this from the overall mean of the rumor belief scores to approximate the constant (2.09). As illustrated in Equation 4, we used the estimated intercept and beta coefficients to calculate the belief score of each agent on a scale of 1 to 5. This score was then normalized to 0-1 scale, to obtain the probability that the agent believes misinformation disseminated by the attacker agent. Belief in rumors is determined probabilistically for each exposure. An agent believes the misinformation if a random draw is less than or equal to their belief probability. Although this allows for probabilistic belief, the outcome for each event is still binary. Future research could model susceptibility as a continuous variable to better capture risk across the population.

$$\text{Probability of Believing a Rumor} = \frac{(0.10 \times E + 0.02 \times A + 0.01 \times C + 0.13 \times N - 0.04 \times O + 2.09) - 1}{(5-1)} \tag{4}$$

When an end user agent believes the misinformation disseminated by the attacker, several independent human factors presented in Figure 2 are negatively affected. In reality, the extent to which each human factor is influenced may vary depending on individual characteristics. However, since the objective of this study is to evaluate how severely misinformation can amplify the impact of subsequent spearphishing attacks, we assume a worst-case scenario. Specifically, we assume that once an end user agent believes the misinformation, all associated human factors immediately shift from their optimal values to their worst possible states. The specific changes in each human factor are summarized in Table 2. These changes directly affect fatigue levels and overall job performance, which in turn negatively influence the decision-making process described in Figure 3. As a result, users become less capable of recognizing suspicious emails and are less likely to engage in security-supportive behaviors, such as alerting other members of the organization or double checking ambiguous emails, even when they do not appear overtly malicious. Although this assumption may not reflect all real-world cases, it is intentionally designed to represent a worst-case scenario in which belief in misinformation causes immediate and comprehensive degradation of psychological and organizational resilience. This approach allows us to better understand the upper bounds of risk amplification through social cyber attacks.

Table 2: Changes in human factor values before and after belief in misinformation.

Human Factor	Value Before Belief in Misinformation	Value After Belief in Misinformation
iSleep Quality	5	1
Stress Average	1	9
Subjective Health Rating	7	3
Depression HAD	0	10
Supportive Leadership	5	1
General Job Satisfaction	5	1
Intrinsic Motivation	5	1
Intention to Leave	1	5
Burnout	1	5
Lack of Motivation	5	1

To model the second decision-making process, engagement with the misinformation post, we leveraged the regression model developed by Buchanan and Benson based on an empirical study involving 357 Facebook users (Buchanan and Benson 2019). In their study, they proposed a model that predicts the likelihood that end user agents will generate organic reach through actions such as sharing, liking, commenting, or reacting to a misinformation post. This prediction was based on the Big Five personality traits, risk propensity, and trust condition. Risk propensity in this context refers to an individual's general tendency to take risks in daily life. The trust condition indicates whether the misinformation post was shared by a close friend, which represents high trust, or by someone recently added as a Facebook friend but not well known to the user, which represents low trust. Since the beta coefficient for risk propensity was reported as 0.00 in their model, we excluded this variable from our simulation. In our implementation, we assigned a trust condition value of 0 when the user directly encountered the misinformation from the attacker agent, and a value of 1 when the post was viewed through another end user agent within the organization who was connected as a Facebook friend. As shown in Equation 5, we calculated the organic reach score of each agent, which ranged from 4 to 20, based on their Big Five personality traits and trust condition. As illustrated in Equation 6, this score was then normalized using Min-Max normalization to derive the probability of spreading misinformation by organic reach. In our simulation, when an end user agent generates organic reach in response to a specific misinformation post, Facebook's algorithm automatically exposes the post to other end user agents within the organization who are connected as Facebook friends. Although we assume all users interact only via Facebook, real-world users and attackers operate across multiple platforms. Thus, actual misinformation spread depends on both organizational usage patterns and attack strategies.

$$\text{Organic Reach Score} = 9.70 + 0.03 \times E - 0.12 \times A - 0.03 \times C + 0.02 \times N - 0.01 \times O + 1.55 \times \text{TC} \quad (5)$$

$$\text{Probability of Spreading Misinformation} = \frac{\text{Organic Reach Score - 4}}{16} \quad (6)$$

5 VIRTUAL EXPERIMENTS

In this section, we describe the details of our virtual experiment. We first simulated a baseline attack campaign, which involved only a spearphishing attack targeting all end user agents within the organization. Subsequently, we simulated a hybrid cyberattack campaign, which began with the spread of misinformation followed by a spearphishing attack, using nine different informal relationship networks. In total, we defined 10 experimental cases, and for each case, we conducted 100 simulation runs, resulting in 1,000 simulations overall. Throughout the simulation, we collected data on the number of agents who encountered misinformation, those who believed it, and those who were ultimately deceived by the spearphishing email and executed the malware. An overview of the virtual experiment design is summarized in Table 3.

Table 3: Virtual experiment summary.

Category	Variable	Description
Independent Variable	Informal Relationship Network (scaling constant c)	Network density is controlled by adjusting the scaling constant c. Values of c used in this study include 2, 5, 10, 20, 30, 40, 50, 100, and 200.
Dependent Variable	Number of agents who encountered misinformation	The number of end user agents exposed to the misinformation post shared on Facebook.
	Number of agents who believed misinformation	The number of end user agents who accepted the misinformation as true.
	Number of agents tricked by spearphishing	The number of end user agents who were deceived by the spearphishing email and executed the attached malware.
Control Variable	Virtual Organization	A simulated organization consisting of 235 end user agents, each with distinct human factors such as phishing susceptibility, job performance, and fatigue level.
	Cyberattack Campaign Type	Two types of attack campaigns were simulated: (1) a baseline spearphishing campaign targeting all users, and (2) a hybrid campaign beginning with misinformation dissemination followed by spearphishing.
	Number of Simulations Per Case	100

5.1 Simulation Results & Statistical Analysis

As illustrated in Figure 5, the denser informal relationship networks, represented by lower values of the scaling constant c, resulted in greater exposure to and belief in misinformation. This, in turn, led to higher success rates for phishing attacks. Compared to the baseline spearphishing campaign without misinformation, the hybrid campaign resulted in an increase of 4 to 7 additional phishing victims on average. This corresponds to approximately a 40% to 75% increase in overall impact, demonstrating the amplification effect of misinformation on cyberattack outcomes.

The regression analysis results presented in Tables 4–6 demonstrate that the mean number of informal relationships per agent is a strong and statistically significant predictor of all three outcome variables: exposure to misinformation, belief in misinformation, and phishing victimization. Specifically, each additional informal connection increases the number of exposed agents by 4.51 (β = 0.83), the number of believers by 2.00 (β = 0.80), and the number of phishing victims by 0.11 (β=0.26). These results suggest that informal relationship network density within the virtual organization influences the amplification of the overall effectiveness of hybrid misinformation and spearphishing attacks.

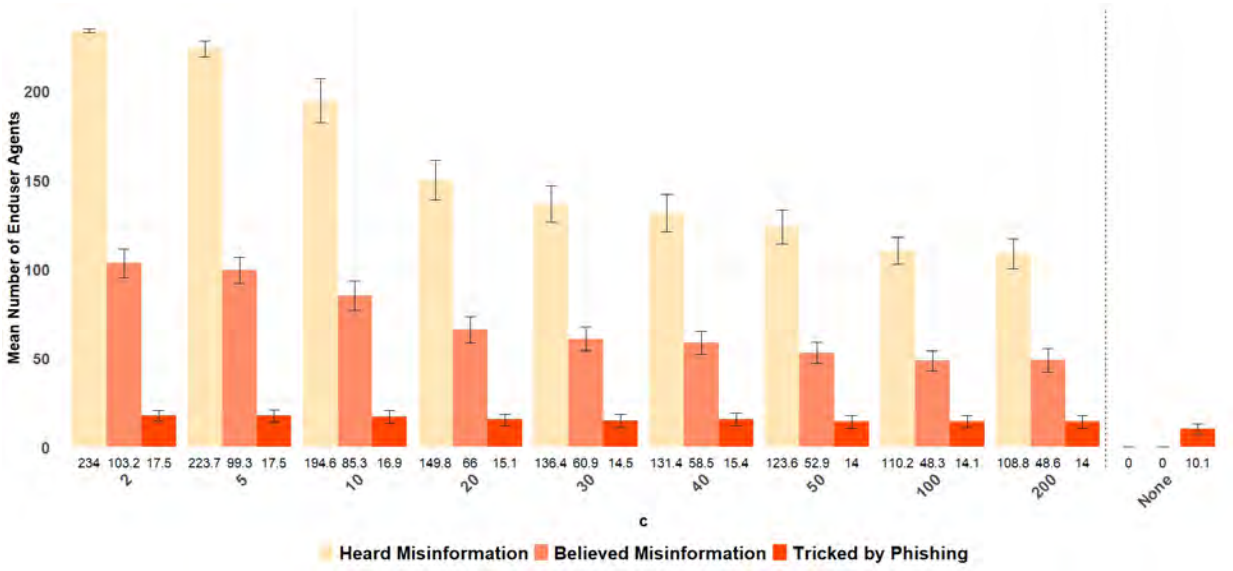

Figure 5: Simulation results.

Table 4: Regression analysis 1 - correlation between mean number of informal relationships to the number of end user agents exposed to misinformation.

	B	*SD B*	*β*	*t*	*p*
(Constant)	130.15	1.04		125.09	< 0.0001
Mean Number of Informal Relationships per Each Agent	4.51	0.09	0.83	45.13	< 0.0001

Table 5: Regression analysis 2 - correlation between mean number of informal relationships to the number of end user agents believed misinformation.

	B	*SD B*	*β*	*t*	*p*
(Constant)	57.34	0.51		111.00	< 0.0001
Mean Number of Informal Relationships per Each Agent	2.00	0.04	0.80	40.34	< 0.0001

Table 6: Regression analysis 3 - correlation between mean number of informal relationships to the number of end user agents tricked by phishing.

	B	*SD B*	*β*	*t*	*p*
(Constant)	14.74	0.14		100.83	< 0.0001
Mean Number of Informal Relationships per Each Agent	0.11	0.01	0.26	8.14	< 0.0001

6 CONCLUSION AND FUTURE WORKS

In this paper, we used agent-based simulations to examine how hybrid misinformation and spearphishing campaigns amplify phishing victimization across virtual organizations with varying degrees of informal relationship network density. Our findings show that the presence of misinformation prior to technical attacks significantly amplifies organizational vulnerability, increasing phishing success rates by up to 75% compared to baseline conditions. These results highlight the critical importance of addressing human cognitive factors and pre-attack psychological manipulation in cybersecurity strategies. In the future, we plan to expand this research in several directions. First, we will incorporate alternative network generation algorithms to explore how different social structures affect misinformation spread. Second, the current spearphishing model only captures the number of users deceived. Future simulations will implement more realistic attack chains to evaluate how early stage misinformation affects overall organizational damage. Third, we plan to assess the effectiveness of technical and organizational defense strategies against misinformation. Finally, while the current study assumes an immediate and extreme shift in human factors as a worst-case scenario, which may overestimate the impact compared to real-world situations, we acknowledge this limitation and will incorporate empirically grounded models of how misinformation gradually influences human emotions and cognitive states in future work.

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

Åkerstedt, T., J. Axelsson, M. Lekander, N. Orsini, and G. Kecklund. 2014. "Do Sleep, Stress, and Illness Explain Daily Variations in Fatigue? A Prospective Study". *Journal of Psychosomatic Research* 76(4):280–285.

Altman, N., K. M. Carley, and J. Reminga. 2020. "ORA User's Guide 2020". *Carnegie Mellon University, Pittsburgh, PA, Institute for Software Research International, Technical Report* 2:2.

Asendorpf, J. B., and S. Wilpers. 1998. "Personality Effects on Social Relationships". *Journal of Personality and Social Psychology* 74(6):1531.

Barrick, M. R., G. L. Stewart, M. J. Neubert, and M. K. Mount. 1998. "Relating Member Ability and Personality to Work-Team Processes and Team Effectiveness". *Journal of Applied Psychology* 83(3):377.

Basit, A., Z. Hassan *et al.* 2017. "Impact of Job Stress on Employee Performance". *International Journal of Accounting and Business Management* 5(2):13–33.

Blythe, J., A. Botello, J. Sutton, D. Mazzocco, J. Lin, M. Spraragen *et al.* 2011. "Testing Cybersecurity with Simulated Humans". In *Proceedings of the AAAI Conference on Artificial Intelligence*, Volume 25, 1622–1627.

Buchanan, T., and V. Benson. 2019. "Spreading Disinformation on Facebook: Do Trust in Message Source, Risk Propensity, or Personality Affect the Organic Reach of "Fake News"?". *Social Media + Society* 5(4):2056305119888654.

Burns, A., C. Posey, J. F. Courtney, T. L. Roberts, and P. Nanayakkara. 2017. "Organizational Information Security as a Complex Adaptive System: Insights from Three Agent-Based Models". *Information Systems Frontiers* 19:509–524.

Carley, K. M. 2014. "ORA: A Toolkit for Dynamic Network Analysis and Visualization".

Carley, K. M. 2020. "Social Cybersecurity: An Emerging Science". *Computational and Mathematical Organization Theory* 26(4):365–381.

Carley, K. M., D. B. Fridsma, E. Casman, A. Yahja, N. Altman, L.-C. Chen, *et al.* 2006. "BioWar: Scalable Agent-Based Model of Bioattacks". *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans* 36(2):252–265.

Chen, L.-C., T. A. Longstaff, and K. M. Carley. 2004. "Characterization of Defense Mechanisms Against Distributed Denial of Service Attacks". *Computers & Security* 23(8):665–678.

Eftimie, S., R. Moinescu, and C. Răcuciu. 2022. "Spear-Phishing Susceptibility Stemming from Personality Traits". *IEEE Access* 10:73548–73561.

Erdos, P., and A. Rényi. 1960. "On the Evolution of Random Graphs". *Publicationes Mathematicae, Institutum Hungaricum Academiae Scientiarum* 5(1):17–60.

Feiler, D. C., and A. M. Kleinbaum. 2015. "Popularity, Similarity, and the Network Extraversion Bias". *Psychological Science* 26(5):593–603.

Freitag, M., and P. C. Bauer. 2016. "Personality Traits and the Propensity to Trust Friends and Strangers". *The Social Science Journal* 53(4):467–476.

Gorodetski, V. I., O. Karsayev, A. Khabalov, I. Kotenko, L. J. Popyack, and V. Skormin. 2001. "Agent-Based Model of Computer Network Security System: A Case Study". In *Information Assurance in Computer Networks: Methods, Models, and Architectures for Network Security. International Workshop MMM-ACNS 2001, St. Petersburg, Russia, May 21–23, 2001, Proceedings 1*, 39–50.

Karim, Taha 2018. "In the Trails of Windshift APT". https://gsec.hitb.org/sg2018/sessions/commsec-the-trails-of-windshift-apt/.

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):1–13.

Lai, K., X. Xiong, X. Jiang, M. Sun, and L. He. 2020. “Who Falls for Rumor? Influence of Personality Traits on False Rumor Belief”. *Personality and Individual Differences* 152:109520.

Leow, S., and Z. Wang. 2018. “You Don’t Know Me but Can I Be Your Friend? Accepting Strangers as Friends on Facebook”. *Social Networking* 8(01):52.

Macal, C. M., and M. J. North. 2009. “Agent-Based Modeling and Simulation”. In *2009 Winter Simulation Conference (WSC)*, 86–98 https://doi.org/10.1109/WSC.2009.5429318.

Mwaisaka, D., G. K’Aol, and C. Ouma. 2019. “Influence of Supportive Leadership Style on Employee Job Satisfaction in Commercial Banks in Kenya”. *Journal of Human Resource and Leadership* 4(1):44–66.

Padur, K., H. Borrion, and S. Hailes. 2025. “Using Agent-Based Modelling and Reinforcement Learning to Study Hybrid Threats”. *Journal of Artificial Societies and Social Simulation* 28(1).

Rehman, W. U., S. Y. Janjua, and H. Naeem. 2015. “Impact of Burnout on Employees’ Performance: An Analysis of Banking Industry”. *World Review of Entrepreneurship, Management and Sustainable Development* 11(1):88–105.

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. Pittsburgh, PA, USA.

Shin, J., L. R. Carley, and K. Carley. 2025. “Simulation of Human Organizations with Computational Human Factors Against Phishing Campaigns”. In *International Conference on Cyber Warfare and Security*, 614–623. Academic Conferences International Limited.

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)*, 1–13. Washington, D.C., USA.

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 https://doi.org/10.1109/WSC60868.2023.10408211.

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. Hamilton, ON, Canada.

Shin, J., G. B. Dobson, K. M. Carley, and L. R. Carley. 2022. “OSIRIS: Organization Simulation in Response to Intrusion Strategies”. In *Social, Cultural, and Behavioral Modeling. SBP-BRiMS 2022*, edited by R. Thomson, C. P. Dancy, and A. A. Pyke, Volume 13558 of *Lecture Notes in Computer Science*, 134–143. Cham, Switzerland.

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”. In *Technical Report*.

Tian, Q., J. Bai, and T. Wu. 2022. “Should We Be "Challenging" Employees? A Study of Job Complexity and Job Crafting”. *International Journal of Hospitality Management* 102:103165.

Vestad, A., and B. Yang. 2024. “A Survey of Agent-Based Modeling for Cybersecurity”. *Human Factors in Cybersecurity* 127:83–93.

AUTHOR BIOGRAPHIES

JEONGKEUN SHIN is a Societal Computing Ph.D. student at Carnegie Mellon University School of Computer Science in Pittsburgh, Pennsylvania, United States. His research interests include agent-based modeling and simulation for cybersecurity and human factors in cybersecurity. His email address is jeongkes@andrew.cmu.edu.

HAN WANG is an undergraduate student majoring in Information Systems at Carnegie Mellon University in Pittsburgh, Pennsylvania. Her research interests include UI/UX design and software development. Her email address is hanwang3@andrew.cmu.edu.

L. RICHARD CARLEY is the Professor of Electrical and Computer Engineering Department at Carnegie Mellon University in Pittsburgh, Pennsylvania, United States. His research interests include analog and RF integrated circuit design in scaled CMOS technologies, and algorithms and methodology for analyzing social media network data. His email address is lrc@andrew.cmu.edu.

KATHLEEN M. CARLEY 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), 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. Her email address is kathleen.carley@cs.cmu.edu.