AI Tech International Journal
Vol. 1, No. 1, July-December, 2023
Journal homepage: https://techaijournal.com

A Cognitive Approach to Cyber Defense and Building and Testing a Prototype Can Make the Internet a Safer Place.

B Pavan Kumar[1], Dr Ashish Chandra Swami[2], Dr Sikhakolli Gopi Krishna[3]

[1]Research Scholar, Department of Computer Science & Engineering, JS University, Shikohabad, Uttar Pradesh
[2]Associate Professor, Supervisor, Department of Computer Science & Engineering, JS University, Shikohabad, Uttar Pradesh
[3]Professor, Co-Supervisor & Professor Sri Mittapalli College of Engineering, Guntur, Andhra Pradesh.

Article Info

Article history:
Received July 08, 2023
Revised September 13, 2023
Accepted September 18, 2023
Published September 29, 2023

Index Terms: Cybersecurity games, attack–defense decision making, financial incentives in cybersecurity, monetary penalties, human–Nash equilibrium comparison, Markov security games, network patching effectiveness, cognitive modeling, Instance-Based Learning (IBL) theory, human factors in cybersecurity, adversarial behavior analysis,

ABSTRACT

Cyberattacks on digital infrastructure are increasing rapidly, making it essential to understand how financial incentives, technical limitations, and environmental information shape the decisions of attackers and defenders. This thesis examined these aspects through a series of controlled behavioral experiments combined with computational cognitive modeling. Across three financial incentive–based studies, it was consistently observed that monetary rewards influenced players' decision strategies in non-intuitive ways. Instead of encouraging active engagement, financial rewards led human participants to reduce both attack and defense actions, suggesting that incentives may create a risk-averse mindset in cyber game environments. Conversely, financial penalties for analysts—particularly penalties for misses and false alarms—had a substantial impact on strategy selection, pushing analysts to become more vigilant but also causing greater deviation from the ideal Nash equilibrium. When participants faced real human opponents rather than algorithmic Nash players, these motivational effects intensified, demonstrating that human-human interaction is a major driver of strategic variability. The Instance-Based Learning (IBL) cognitive models developed in this research successfully reproduced these human patterns, underscoring the importance of cognitive factors—such as recency of feedback, frequency of outcomes, and memory activation—in shaping cyber decision-making under uncertainty.

Technical constraints were examined through Markov Security Games (MSGs), which capture the dynamic nature of network security as it changes in response to defender patching actions. The results showed clear asymmetries in how effective and less effective patching influenced attacker behavior. Effective patching reduced network vulnerability by 90%, significantly limiting opportunities for successful intrusions, while less effective patching still reduced vulnerability by 50%. Although defense proportions remained stable across vulnerable and non-vulnerable states, attack actions decreased substantially when the network entered a non-vulnerable state, indicating that attackers were sensitive to technical constraints. Deviations from Nash equilibrium were frequent across conditions, suggesting that human players do not always conform to theoretically optimal strategies in dynamic environments. Supporting these findings, IBL modeling revealed that successful patching was associated with high recency and frequency of patching successes for analysts, while attackers performed better when relying on older or less frequent information—highlighting divergent cognitive mechanisms driving the two roles. Less effective patching produced opposite patterns, demonstrating that technical environments fundamentally shape cognitive processing and decision patterns.

Environmental factors were explored by manipulating the availability of interdependence information—knowledge about opponent actions and payoffs. When both players had full visibility into each other's intentions, behaviors,

and rewards, they significantly increased their attack and defense frequencies. This shows that transparency amplifies competitive behavior, as players react more aggressively to their opponent's visible strategies. In contrast, when players lacked interdependence information, decisions became more conservative and predictable. These findings emphasize the role of situational awareness in cybersecurity decision-making and highlight that information availability can either stabilize or destabilize strategic behavior.

Overall, the extended findings of this thesis demonstrate that real-world cybersecurity behavior is shaped by a complex interplay of human cognition, financial and motivational structures, network-level technical conditions, and contextual information. By integrating cognitive models with prototype-based experiments, this research provides a deeper understanding of how defenders and attackers actually make decisions—not just how they should behave according to theory. These insights offer valuable implications for designing adaptive, behavior-aware cybersecurity systems, improving defense mechanisms, and developing training programs that account for real human cognitive tendencies rather than assuming purely rational decision-making.

Index Terms: Cybersecurity games, attack–defense decision making, financial incentives in cybersecurity, monetary penalties, human–Nash equilibrium comparison, Markov security games, network patching effectiveness, cognitive modeling, Instance-Based Learning (IBL) theory, human factors in cybersecurity, adversarial behavior analysis,

INTRODUCTION

The Cybersecurity has become increasingly critical due to the rapid rise in the frequency, complexity, and precision of cyberattacks (Symmetry, 2019). A cyberattack involves attempts to gain unauthorized access to systems, manipulate data, or disrupt digital infrastructure (UpGuard, 2020). According to Symantec's 2019 report, global financial losses from data breaches were projected to reach $5 trillion by 2020. Attackers continue to exploit modern tools and technologies to identify system vulnerabilities, making both technical defenses and human decision-making essential for effective cybersecurity management (Symmetry, 2019). Despite this, organizations often underinvest in the human element until after an attack occurs (NIST, 2017; Anwar et al., 2016).

Cybersecurity interactions are frequently modeled as non-cooperative games using behavioral game theory (Dutt, Ahn, & Gonzalez, 2013). Traditional game-theoretic methods rely on static or fully informed games, which oversimplify the fast-paced, uncertain nature of real-world cybersecurity environments (Roy et al., 2010). These models typically assume Nash equilibrium strategies, but computing equilibria becomes increasingly complex in repeated and incomplete-information games. As a result, such models overlook the role of motivational, technical, and contextual factors in shaping attacker–defender behavior.

Behavioral Game Theory (BGT; Camerer, 2003) addresses these limitations by focusing on human decision-making, incentives, and bounded rationality. Security games—simple 2 × 2 attacker–defender interactions—capture dynamic choices where attackers decide to "attack" or "not attack," and defenders choose "defend" or "not defend." These games demonstrate how outcomes depend on the strategic decisions of both players and have been widely used to model real-world networks (Lye & Wing, 2005; Alpcan & Bäpper, 2011).

Building on these foundations, this thesis investigates how repeated attacker–defender interactions are influenced by technological constraints (e.g., patching accuracy), motivational factors (e.g., costs and rewards), and environmental elements (e.g., information availability). The study employs Instance-Based Learning Theory (IBLT), a cognitive framework shown to accurately model human decisions under uncertainty (Gonzalez & Dutt, 2011), to explain how attackers and defenders adapt their behavior over time.

LITERATURE REVIEW

Research Recent research in cybersecurity shows that traditional technical defenses alone are no longer sufficient to secure cyberspace. As cyber threats become more complex, frequent, and adaptive, researchers have turned toward building experimental prototypes—such as cyber ranges, virtualized network environments, attack–defense simulators, and digital twins—to observe how attackers behave and how defenders respond under realistic but controlled conditions. These prototype-based systems make it possible to test patching mechanisms, incident response strategies, intrusion detection tools, and automated defense algorithms with real user interaction data.

Parallel to these developments, cognitive approaches have gained significant importance, as they offer insights into how humans think, learn, and make decisions during cyber incidents. Models such as Instance-Based Learning (IBL), ACT-R,

SOAR, and reinforcement learning frameworks are increasingly being used to predict attacker intent, support defender decision-making, and reduce analyst workload during threat monitoring. Research shows that human factors—including perception, attention span, memory decay, cognitive biases, stress, fatigue, and experience levels—directly influence the speed and accuracy of cyber defense actions. For example, analysts may miss attacks due to information overload, while attackers often rely on trial-and-error learning patterns similar to reinforcement learning algorithms.

By combining prototype tools with cognitive models, researchers can design defense mechanisms that not only respond to threats but also **anticipate attacker strategies**, detect anomalies earlier, allocate resources more efficiently, and deliver personalized training for security teams. Behavioral data collected from prototypes can be fed into cognitive models to improve their predictive accuracy, creating a continuous feedback loop of learning and adaptation. This integrated approach—merging technical experimentation with human-centered cognitive science—is proving vital for building safer, more resilient, and behavior-aware cyberspace systems capable of defending against modern, evolving cyber threats.

METHODOLOGY

The Most existing cybersecurity literature remains highly specialized, largely policy-oriented, and predominantly centered around the United States. While some American strategists focus on China or Russia, the U.S. remains the primary reference point, producing a significant gap between this literature and broader international relations theories. Outside of U.S. military institutions and think tanks, scholarly work is scattered and lacks coherence.

Constructivist approaches—particularly securitization theory—have been used by several academics to examine how cyber threats are socially constructed. These studies offer meaningful insight into how countries perceive digital threats and respond through policy, but comparative research in non-U.S. contexts remains limited. Post-structuralist literature on "Postmodern War" also contributes by analyzing the interplay between information, warfare, and technology. Despite cybersecurity's growing prominence, theoretical engagement within security studies remains surprisingly limited. One explanation is that technological threats do not easily fit traditional definitions of "security," which emphasize urgency and extraordinary state responses.

This study relied heavily on primary sources—Government of India documents, Ministry of Defence and Ministry of Home Affairs releases, and other allied departmental reports. Secondary sources included scholarly books, peer-reviewed papers, national and international newspaper articles, and educational institution websites.

Notable works such as *Cyberpower and National Security*, *Cyberspace and National Security*, and *Cyber War* explain how cyberspace is weaponized against nations. Other texts, including *Cyber Security: The Essential Body of Knowledge* and *The Basics of Cyber Warfare*, help establish foundational understanding. Miriam Dunn Cavelty's literature and publications from IDSA cyber experts were also extensively examined.

Cyber threats remain in a formative stage—dynamic, fast-evolving, and driven by non-state actors, hackers, and malicious groups who continuously create new attack strategies. This study focused on India's cyber challenges, structural needs, and policy gaps. Cybersecurity debates remain contentious because every user—government, private sector, and citizens—is a stakeholder and faces varying threats like malware, worms, or data breaches. Opinions differ on whether the internet should remain fully open or be heavily regulated by the state.

This research analyzed the validity of claims on both sides, limiting its scope to national security concerns within India's proposed cyber security framework. Since the bill remains unratified by Parliament, the study focuses on the current policy ecosystem without imposing a temporal boundary.

As cyberattacks rise globally, cybersecurity has evolved into a central element of diplomacy and interstate relations. States now view cyberspace as a strategic domain where they must defend national interests, leverage digital tools for geopolitical advantage, form alliances, and counter hostile actors. National approaches differ significantly: countries like China and Russia advocate strong state control and internet sovereignty, while the U.S. and others emphasize openness, human rights, and free expression. Alongside states, civil society, global institutions, and private corporations also influence the governance of cyberspace, often resisting excessive state intervention and advocating for an open internet.

THESIS ORGANISATION

The Over thousands of years, humanity has progressed from primitive tools to sophisticated information and communication technologies. Modern innovations have greatly enhanced convenience and efficiency in daily life. Yet, as new technologies emerge, some individuals misuse them in pursuit of personal benefit, leading to negative consequences.

One of the most influential technological developments in recent years is the rise of mobile technology and cybersecurity systems. Researchers studying mobile phone usage over the past decade have noted that smartphones support communication, reduce loneliness, and help people maintain social connections. Today, smartphones have become essential components of everyday living, functioning not only as communication devices but also as tools for organizing tasks, engaging on social media, and accessing entertainment.

Despite these advantages, teenagers are particularly vulnerable to the negative effects of excessive smartphone use. Their ongoing search for identity and meaningful social connections often leads to overreliance on mobile devices. Smartphone addiction is increasingly recognized in Western countries as a significant psychological, social, and physical concern. Excessive usage can disrupt attention, reduce academic focus, and contribute to mental health problems such as anxiety, depression, and insomnia. Health risks have also been highlighted, with the World Health Organization classifying mobile phone radiation as "possibly carcinogenic" (Group 2B).

Teenagers often feel pressured to stay continuously connected to peers, causing them to blur the line between virtual and real-life interactions. As a result, many adolescents become detached from their surroundings despite constant digital connectivity. Both educators and parents report that excessive smartphone use negatively affects student behavior, personality development, and academic performance. Many teens spend substantial time calling, texting, and interacting on their phones, indicating a growing dependence that requires greater awareness and guidance.

RESULTS

According The experimental findings show that financial incentives, technical constraints, and environmental information significantly influence attacker and defender decision-making across all stages of cyber interaction. Monetary rewards consistently led to a reduction in both attack and defense actions, suggesting that incentivized players adopt more risk-averse strategies. In contrast, financial penalties increased defensive vigilance but also produced larger deviations from Nash equilibrium strategies, highlighting how stress and fear of loss distort human decision-making under uncertainty.

Prototype-based experiments demonstrated that patching effectiveness strongly shaped attacker behavior, with successful patching reducing attack attempts by more than half and altering the timing and frequency of intrusion attempts. In less effective patching scenarios, attackers maintained higher activity levels, showing that partial patching does not sufficiently discourage adversarial behavior. Analysts, however, displayed relatively stable defensive patterns across both vulnerable and patched network states, indicating a cautious "defense-first" mindset even when the system was non-vulnerable.

Cognitive models based on Instance-Based Learning Theory (IBL) accurately replicated human decisions across diverse experimental conditions, demonstrating high predictive validity and strong generalization across incentive and penalty environments. The models revealed that recency, frequency of past outcomes, cognitive noise, and attention levels were dominant contributors to both attacker and defender choices. High recency and frequency weights produced more rational, stable decisions, while increased cognitive noise resulted in greater deviations from optimal behavior.

Environmental information also played a significant role: when attackers and defenders had access to interdependence information (opponent payoffs and actions), both parties increased their activity levels. Attackers became more opportunistic, while analysts adopted more aggressive or anticipatory defense strategies.

Overall, the results indicate that combining cognitive insights with prototype-based experimentation is essential for designing adaptive, human-aware cyber defense mechanisms. This integrated approach not only improves prediction of adversarial behavior but also provides a foundation for training systems, automated decision support tools, and behavior-driven cybersecurity policies.

DISCUSSION

Moore's law The findings of this study underline the intricate interaction between human cognition, financial incentives, technical limitations, and situational awareness in cybersecurity decision-making. Reward and penalty structures caused significant deviations from Nash equilibrium predictions, demonstrating that traditional game-theoretic models alone cannot account for real human behavior in cyber conflict. While monetary rewards reduced both attack and defense actions—indicating a shift toward risk-averse strategies—penalties encouraged greater defensive vigilance but also introduced cognitive stress and decision variability. Prototype-based experiments showed that patching effectiveness plays a central role in shaping attacker choices, with higher patching success markedly reducing attack attempts. Defenders, however, maintained relatively stable behavior across different network states, suggesting a reliance on habitual or cautious

strategies rather than adaptive responses. Cognitive modeling with Instance-Based Learning (IBL) successfully reproduced these behavioral patterns, validating its usefulness for studying cybersecurity decisions under varying incentive conditions. Overall, integrating cognitive modeling with prototype-driven experimentation demonstrates that effective cyber defense must move beyond purely technical optimization. Behavior-aware defense strategies that account for human decision tendencies are essential for creating adaptive, predictive, and resilient security systems.

CONCLUSION

The This chapter outlines the rationale for using the Fogg Behavior Model (FBM) as the theoretical basis for designing the cybersecurity intervention. According to FBM, behavior occurs when ability, cue, and motivation converge. The model identifies three main motivators, each with two dimensions, and the researcher emphasized **anticipation** as the primary motivator for this intervention. The use of **Cyber Suraksha**, a serious game, was integrated to enhance user engagement and learning. Carpenter (2019) identifies four key variables driving security awareness programs: information dissemination, compliance, behavior shaping, and culture molding. Knowledge alone is insufficient; users lacking interest in cybersecurity are unlikely to act proactively. Efforts to improve security behavior must align with human nature, highlighting the need for SETA software designed with behavioral principles in mind. The primary aim of this thesis is to establish a framework for enhancing the cybersecurity practices of Indian smartphone users, leveraging optimism and fear as guiding principles.

The chapter also discusses the **SMITE IP traceback approach**, which combines MPLS and SDN to trace IP addresses effectively in autonomous systems. SMITE offers multiple advantages: it enables packet-level attack detection, reduces CAM/TCAM memory usage in switches, accelerates MPLS tag matching in the dataplane, and decreases overall power consumption. The approach enhances SDN network performance, and an IPv4 SMITE application example is described. While modifications for IPv6 networks are feasible by using dual MPLS labels, the source codes of OpenvSwitch and Ryu controllers have been adjusted to preserve the first bit of the source IP in the Reserved Flag (RF) field. Given the rise in distributed denial-of-service (DDoS) attacks, it is crucial to promptly notify authorities of any such incidents to ensure network security.

REFERENCES

[1]. S. Savage, D. Wetherall, A. Karlin, and T. Anderson, “Practical network support for IP traceback,” SIGCOMM Comput. Commun. Rev., vol. 30, no. 4, 295–306, Aug. 2000, issn: 0146-4833. doi: 10.1145/347057.347560. [Online]. Available: https://doi.org/10.1145/347057.347560 (pages 7, 9, 20).

[2]. Open Networking Foundation (ONF), Openflow switch specification version 1.3.0, 2015. [Online]. Available: https://www.opennetworking.org/wp-content/ uploads / 2014 / 10 / openflow - spec - v1 . 3 . 0 . pdf (visited on 06/01/2017)(pages 7, 11, 27, 43, 52, 69).

[3]. F. X. Wibowo, M. A. Gregory, K. Ahmed, and K. M. Gomez, “Multi-domain software defined networking: Research status and challenges,” Journal of Network and Computer Applications, vol. 87, pp. 32–45, 2017, issn: 1084-8045.

[4]. K. Katsalis, B. Rofoee, G. Landi, J. Riera, K. Kousias, M. Anastasopoulos, L. Kiraly, A. Tzanakaki, and T. Korakis, “Implementation experience in multi-domain SDN: Challenges, consolidation and future directions,” Computer Networks, vol. 129, pp. 142–158, 2017,

[5]. Mininet Team, Mininet: An instant virtual network on your laptop (or other pc), 2017. [Online]. Available: http://mininet.org/ (visited on 01/15/2017) (pages 9, 10, 43, 52, 54, 69, 76, 87, 96).

[6]. Linux Foundation, Openvswitch an open virtual switch, 2016. [Online]. Available: http://openvswitch.org/ (visited on 06/01/2016) (pages 9, 10, 43, 52, 69, 76, 87, 96).

[7]. Ryu Development Team, 2017, Ryu- Content-Based Software Defined Networking Framework-Build SDN Agilely, 2017. [Online]. Available: http://osrg.github. io/ryu/ (visited on 01/01/2017) (pages 9–11, 43, 52, 55, 59, 69, 76, 87, 96).

[8]. Ryu Development Team, 2015, Ryu documentation - release 3.25, 2015. [Online]. Available: http://ryu-devel.readthedocs.io/en/latest/ofproto_ref. html (visited on 01/01/2017) (pages 9–11, 43, 52, 59, 69, 76, 87, 96).

[9]. Open Networking Foundation(ONF), “Openflow switch specification, version 1.3.0 (wire protocol 0x04),” OpenFlow Switch Specification, 2012. [Online]. Available: https://opennetworking.org/wp-content/uploads/2014/10/openflowspec-v1.3.0.pdf (visited on 01/15/2017) (pages 9, 10, 76, 96).

[10]. P. Jakma and D. Lamparter, “Introduction to the quagga routing suite,” IEEE Network,vol. 28, no. 2, pp. 42–48, 2014. doi: 10.1109/MNET.2014.6786612 (pages 9, 43, 69, 76).

[11]. C. Cortes and V. Vapnik, “Support-vector networks,” in Machine Learning, 1995, pp. 273–297. doi: 10.1007/BF00994018. [Online]. Available: https://doi.org/ 10.1007/BF00994018 (pages 10, 36, 43, 78).

[12]. scikit-learn 0.23.1, Support Vector Machines (SVM), 2019. [Online]. Available:https : / / scikit - learn . org / stable /

modules / svm . html (visited on 04/16/2020) (pages 10, 36, 38, 43, 78, 85, 87, 90, 91, 96).

[13]. M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the kdd cup 99 data set," in 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications, 2009, pp. 1–6. doi: 10.1109/CISDA.2009.

[14]. Canadian Institute for Cybersecurity, University of New Brunswick(UNB), NSLKDD dataset, 2009. [Online]. Available: https://www.unb.ca/cic/datasets/ nsl.html (visited on 2020) (pages 10, 17, 34–36, 79).

[15]. P. Hadem, D. K. Saikia, and S Moulik, "I-SMITE: An IP traceback mechanism for inter-AS SDN networks using BGP," International Journal of Security and Networks (IJSN), Inderscience, vol. x, no. x, pp. xxx–xxx, 2020 (Accepted). doi: 10.1504/IJSN.2021.117864 (pages 11, 31, 59).

[16]. P. Hadem, D. K. Saikia, and S. Moulik, "An SDN-based intrusion detection systemusing SVM with selective logging for IP traceback," Computer Networks, vol. 191, p. 108 015, 2021, issn: 1389-1286. doi: https : / / doi . org / 10 . 1016 / j .comnet . 2021 . 108015. [Online]. Available: https : / / www . sciencedirect .com/science/article/pii/S1389128621001274 (page 11).

[17]. A. Viswanathan, R. Callon, and E. C. Rosen, "Multiprotocol label switching architecture," Internet RFCs, ISSN 2070-1721, vol. RFC 3031, 2001. doi: 10.17487/ RFC3031. [Online]. Available: https : / / tools . ietf . org / html / rfc3031(pages 11, 29, 44, 59).

[18]. L. Ghein, MPLS Fundamentals: A Comprehensive Introduction to MPLS Theory and Practice. Cisco Press, Indianapolis, 2006 (pages 11, 29, 44, 59).

[19]. Y. Rekhter, T. Li, and S.Hares, "A border gateway protocol 4 (bgp-4)," Internet RFCs, ISSN 2070-1721, vol. RFC 4271, 2006. [Online]. Available: https://tools.ietf. org/html/rfc4271 (pages 11, 31, 59, 62, 98).

[20]. C. Fernandez and J. L. Munoz, "Software defined networking (SDN) with OpenFlow 1.3 OpenvSwitch and ryu," UPC Telematics Department_PhD Thesis, 2015 (pages 11, 48).

[21]. H.-J. Liao, C.-H. Richard Lin, Y.-C. Lin, and K.-Y. Tung, "Intrusion detection system: A comprehensive review," Journal of Network and Computer Applications, vol. 36, no. 1, pp. 16–24, 2013, issn: 1084-8045. doi: https://doi.org/10.1016/j. jnca.2012.09.004.