RESEARCH ARTICLE

GHOST: Game-Theoretic Honeytoken Optimization for Strategic Threat Detection

Darapu Uma[1] **and Manas Kumar Yogi**[1,*]

[1] Department of Computer Science and Engineering, Pragati Engineering College, Surampalem 533437, India

Abstract

As adversaries deploy advanced persistent threats (APTs), social engineering, and credential-stuffing attacks to circumvent classical reactive defenses, identity security faces a formidable challenge. This paper proposes GHOST (Game-theoretic Honeytoken Optimization for Strategic Threat Detection), a mathematically grounded and empirically evaluated framework that combines deceptive honeytokens with Stackelberg–Nash game-theoretic optimization, Bayesian attacker-type inference, and reinforcement learning (RL). The defender (Stackelberg leader) distributes honeytokens throughout a networked system of heterogeneous assets, while the attacker (follower) operates under imperfect knowledge of the deployed deceptive strategy. GHOST models the defender–attacker interaction as a Stackelberg game in which the defender commits to a mixed placement strategy before the attacker responds. A Bayesian updating mechanism iteratively refines the posterior belief over attacker archetypes, while Nash Equilibrium conditions are resolved to guarantee strategic stability. An RL-guided gradient-ascent engine dynamically repositions honeytokens in real time, yielding accelerated convergence and sustained optimal detection. Evaluated against a purpose-built simulation dataset of 10,000 network scenarios—the Honeytoken Strategic Security Dataset (HSSD)—GHOST achieves an 85.3% detection rate and a false positive rate of only 3.1%, outperforming the strongest baseline (rule-based placement) by 23.4%. Ablation experiments confirm the measurable contribution of each architectural component. GHOST is further shown to be 3.2× more cost-effective than random placement and integrates natively with zero-trust architectures (ZTA). A four-pillar governance model operationalizes ethical and regulatory compliance, addressing requirements under GDPR, HIPAA, CFAA, and the Budapest Convention.

Keywords: honeytokens, game theory, nash equilibrium, stackelberg games, bayesian security, reinforcement learning, identity security, zero-trust architecture, deception technology, intrusion detection.

1 Introduction

Identity security has emerged as the primary frontline in modern cybersecurity owing to the pervasive reliance on digital identities across enterprise, cloud, and hybrid environments. The Verizon Data Breach Investigations Report (2023) attributed over 74% of all data breaches to human error and compromised credentials [1], exposing the limitations

Submitted: 13 May 2026
Accepted: 06 June 2026
Published: 30 June 2026

Vol. 2, **No.** 1, 2026.
10.62762/TC.2026.152584

***Corresponding author:**
✉ Manas Kumar Yogi
manas.yogi@gmail.com

Citation
Uma, D., & Yogi, M. K. (2026). GHOST: Game-Theoretic Honeytoken Optimization for Strategic Threat Detection. *ICCK Transactions on Cybersecurity*, 2(1), 75–92.

© 2026 ICCK (Institute of Central Computation and Knowledge)

of perimeter-centric, rule-based defenses. Despite their value, intrusion detection systems (IDS) and multi-factor authentication (MFA) operate in an essentially reactive manner, ceding the initiative to the attacker by responding only after malicious signals have already penetrated the perimeter. Such detection latency is especially damaging against advanced persistent threats (APTs), in which adversaries remain undetected for an average of 204 days before discovery. Rass et al. [4] provide a comprehensive analysis of enterprise data breach causes and prevention strategies, contextualizing the challenges addressed by GHOST.

A paradigm shift from passive monitoring to proactive interception is represented by honeytokens—forged credentials that are indistinguishable from authentic assets. Honeytokens are lightweight, scalable across heterogeneous network topologies, and embeddable within live authentication flows, databases, and document repositories, in contrast to passive honeypots that need specialized framework. Yet existing honeytoken implementations suffer from two critical deficiencies: (i) static, heuristic placement strategies that sophisticated attackers eventually decode through reconnaissance; and (ii) absence of a principled mathematical framework for optimizing the defender's deceptive strategy against a rational, adaptive adversary.

Game theory provides precisely this mathematical scaffolding. Stackelberg games model the sequential decision-making inherent in security deployment—the defender commits to a strategy before the attacker acts—while Bayesian formulations accommodate the attacker's incomplete information about defenses, and Nash Equilibria identify stable strategy pairs where neither party benefits from unilateral deviation. Reinforcement learning (RL) translates these equilibrium conditions into an online, self-improving optimization loop that adapts to observed attacker behavior without requiring prior knowledge of attacker intent. This paper makes the following principal contributions:

We formalize honeytoken deployment as a Stackelberg security game with Bayesian attacker-type uncertainty, deriving closed-form Nash Equilibrium conditions for mixed strategies across heterogeneous system networks. We introduce GHOST—a robust, four-phase algorithm that combines Nash-guided gradient ascent, ε- greedy reinforcement learning exploration, Bayesian posterior updates for attacker profiling, and a convergence-safety rollback mechanism.

We construct and release the Honeytoken Strategic Security Dataset (HSSD), comprising 10,000 simulated game-theoretic scenarios across three attacker archetypes and diverse system configurations. We conduct comprehensive experiments demonstrating that GHOST achieves an 85.3% detection rate—a 23.4 pp gain over the strongest baseline—with a false-positive rate of 3.1% and average convergence in 27 iterations. We present a validated ablation study confirming the independent contribution of each component, a cost-effectiveness analysis, an ethical-legal governance model, and deployment blueprints for enterprise, cloud-IAM, and financial environments.

The remainder of this paper is structured as follows. Section 2 reviews the related literature. Section 3 presents the formal system model. Section 4 develops the GHOST framework and algorithm. Section 5 details the experimental methodology and results. Section 6 presents an ablation study and sensitivity analysis. Section 7 discusses applications in financial fraud prevention and insider threat detection. Section 8 addresses ethical and legal considerations. Section 9 outlines future directions, and Section 10 concludes the paper.

2 Related Work

2.1 Deception Technologies and Honeytokens

Deception technologies encompass a spectrum of defensive artifacts—honeypots, honeynets, honeytokens, and honeyfiles—designed to lure, observe, and attribute adversarial actors, with recent work demonstrating their effectiveness in detecting advanced persistent threats during early intrusion stages [3]. Honeypots, the earliest instantiation, demand dedicated infrastructure and introduce non-trivial deployment overhead; their static, isolated nature limits scalability in dynamic cloud environments. Honeytokens refine this concept into lightweight, embeddable artifacts—fake API keys, synthetic database records, and decoy credentials—that trigger alerts upon interaction without requiring dedicated hardware, an approach validated by recent honeytoken-based deception deployments for APT early detection [3]. Subsequent work has demonstrated that honeytoken-based deception systems achieve substantially higher detection rates than conventional IDS in credential-based attacks,

attributable to their indistinguishability from real assets and near-zero false-negative rate against unsophisticated attackers [3].

Contemporary research has identified two critical limitations in deployed honeytoken systems. First, static placement strategies concentrating tokens in predictable high-value zones are vulnerable to reconnaissance-aware adversaries who pattern-match token signatures over multiple intrusion attempts. Second, response pipelines remain largely manual, introducing latency between token activation and containment. Polymorphic honeytokens [18] address the first limitation by morphing token characteristics at regular intervals, while cognitive-model-guided personalized deceptive signaling [19] addresses adaptive deception by modeling individual attacker decision-making patterns to tailor defensive signals accordingly. Our work extends both by adding a game-theoretic optimization layer that dynamically positions tokens based on the adversary's evolving strategic posture. Comprehensive taxonomies of deception technologies have been developed by Almeshekah and Spafford [21], while Kahlhofer et al. [22] have explored deception-as-a-service platforms for cloud-native environments.

2.2 Game-Theoretic Security Models

Zhu et al. [11] provided a comprehensive game-theoretic taxonomy of defensive deception, while Rass and Zhu [12] extended these models specifically to APT defense scenarios. Sinha et al. [14] offer a practical treatment of deployed Stackelberg security games. Sengupta et al. [15] applied Markov games to cloud APT detection, and Cho et al. [16] surveyed moving-target defense strategies with game-theoretic foundations. Game theory has been extensively applied to cybersecurity since Lye and Wing's seminal two-player stochastic game model for network security in 2005. Stackelberg security games (SSGs) have been particularly successful in physical security resource allocation—exemplified by ARMOR, deployed at Los Angeles International Airport, and PROTECT, deployed by the US Coast Guard [13]. In the cyber domain, SSGs have been used to optimize intrusion detection sensor placement [27], moving-target defense resource allocation, and network patrol strategies. Bayesian extensions [10] model attacker uncertainty about the defender's type, enabling probabilistic best-response computations. Huang and Zhu [5] applied mixed-strategy Nash Equilibria to optimize firewall rule allocation, demonstrating 18% reduction in successful intrusions over deterministic strategies.

Despite this rich literature, no prior work has formalized honeytoken deployment as a Stackelberg game that simultaneously: (i) derives closed-form Nash Equilibrium conditions for mixed-strategy token distributions across n heterogeneous assets; (ii) integrates Bayesian posterior updates over attacker archetypes; and (iii) combines these equilibrium constraints with a policy-gradient RL engine that maintains equilibrium stability during online adaptation. Existing game-theoretic honeytoken work either treats the token as a binary, fixed-location object [11, 17] or applies signaling games without deriving equilibrium conditions for dynamic, multi-system networks [6, 15]. Reinforcement-learning security systems optimize single-objective reward functions without equilibrium constraints [7, 8], and zero-trust integration with deception remains a qualitative proposition [9]. GHOST addresses all four gaps within a single, unified framework—a combination absent from any prior publication in this domain.

2.3 Reinforcement Learning in Adaptive Security

Zhu et al. [20] investigated reinforcement learning algorithms for adaptive credential defense, complementing earlier work by Zhu et al. [2] that demonstrated RL-based adaptive cyber defense against network-layer exploits through a moving-target defense workshop framework. Reinforcement learning has been applied to network intrusion response [7] and autonomous penetration testing, while complementary deep learning approaches have been applied to adaptive multi-layered honeynet architectures for threat behavior analysis [26]. Survey work on RL applications in cyber attack and defense [8] confirms that RL-guided honeypot placement consistently outperforms static baselines by exploiting feedback loops from attacker interactions, validating the adaptive repositioning strategy adopted in GHOST. Prior RL security systems, however, optimize single-objective reward functions (e.g., pure detection) without incorporating the equilibrium constraints necessary to prevent detector evasion. GHOST integrates RL within a game-theoretic equilibrium framework: the RL policy implements the gradient-ascent update toward the Nash strategy profile, ensuring that local policy improvements preserve global equilibrium stability a constraint absent from all prior RL-based deception systems

Table 1. Comparative taxonomy of related honeytoken and game-theoretic security systems.

Reference	Approach	Game Model	RL	ZTA	Detection	Novelty Gap
[3]	Deception-based honeypot systems for APT detection	None	No	No	Higher than rule-based IDS (qualitative)	No opt. framework
[10]	Bayesian IDS game	Bayesian Stackelberg	No	No	18% vs. baseline	No RL, no ZTA
[8]	Survey: RL in cyber attack and defense	None (survey)	Multiple (reviewed)	No	Varies by system (survey)	No equilibrium framework
[6]	Honeylines + ZTA	None	No	Partial	35% insider	No game theory
[18]	Polymorphic tokens	None	No	No	Qualitative	No RL/ZTA
GHOST (Prop.)	Stackelberg + RL + Bayesian + ZTA	Stackelberg - Nash + Bayesian	Policy Gradient + ε-greedy	Full	85.3%; +23.4 pp	All gaps addressed

Notes: RL = Reinforcement Learning; ZTA = Zero Trust Architecture; IDS = Intrusion Detection System; pp = percentage points; Prop. = Proposed.

reviewed.

2.4 Zero-Trust Architecture and Deception

NIST SP 800-207 represents zero-trust architecture (ZTA), which requires continuing verification of all entities, independent of network location. ZTA's integration with deception technologies is still in its inception, despite its pervasive use and growing corporate adoption of zero-trust principles. In order to improve insider threat detection by 35%, Alamro and Alsulaiman [6] recommended integrating honey objects into ZTA access control levels combined with behavioral analysis. However, this work lacks game-theoretic optimization of token placement within ZTA enforcement points. GHOST addresses this by treating each ZTA verification node as a potential deployment site in the game's system set S, natively embedding deception into continuous-verification pipelines.

To provide a comprehensive overview of the existing literature and clearly delineate GHOST's position within the research landscape, Table 1 presents a comparative taxonomy of related honeytoken and game-theoretic security systems, highlighting the novelty gaps addressed by our proposed framework.

As highlighted in Table 1, prior work in honeytoken deployment has largely focused on isolated aspects of the problem space. Static honeytoken approaches [3] demonstrate strong detection rates against naive attackers but lack any optimization framework, making them vulnerable to reconnaissance-aware adversaries. Bayesian Stackelberg game models [10] provide rigorous mathematical foundations for IDS optimization but do not incorporate online learning or zero-trust integration. RL-based honeypot systems [8] achieve adaptive repositioning but operate without game-theoretic equilibrium constraints, risking detector evasion through policy instability. GHOST is the first framework to simultaneously integrate all four dimensions—Stackelberg-Nash game theory, Bayesian inference, reinforcement learning, and native zero-trust architecture support—enabling the substantial performance gains reported in our experimental evaluation.

3 System Model and Problem Formulation

The information-theoretic foundations underpinning our Bayesian update mechanism—rooted in the classical principles of entropy and information measurement [23]—provide the formal basis for quantifying uncertainty over attacker archetypes, while the threat modeling framework follows Shostack's [25] comprehensive methodology. Consider an organizational network comprising n systems (nodes) $S = \{s_1, s_2, \ldots, s_n\}$, each characterized by a tuple of attributes:

- Asset value $V_i > 0$: the attacker's expected payoff upon successful exploitation of system s_i.
- Deployment cost $c_i > 0$: the resource units required to maintain one unit of honeytoken density on s_i.
- Detection probability $p_i \in (0, 1)$: the probability that an attacker interaction with a honeytoken on s_i triggers an alert.

The defender (D) and attacker (A) are modeled as rational, self-interested agents in a two-player, non-cooperative Stackelberg game. The defender, acting as the Stackelberg leader, commits to a mixed deployment strategy before the attacker selects an intrusion target. The system architecture is illustrated in Figure 1.

3.1 Defender and Attacker Strategy Spaces

The defender's mixed strategy is a probability distribution over honeytoken placements: $D = (d_1, d_2, \ldots, d_n)$, where $d_i \in [0, 1]$ denotes the density

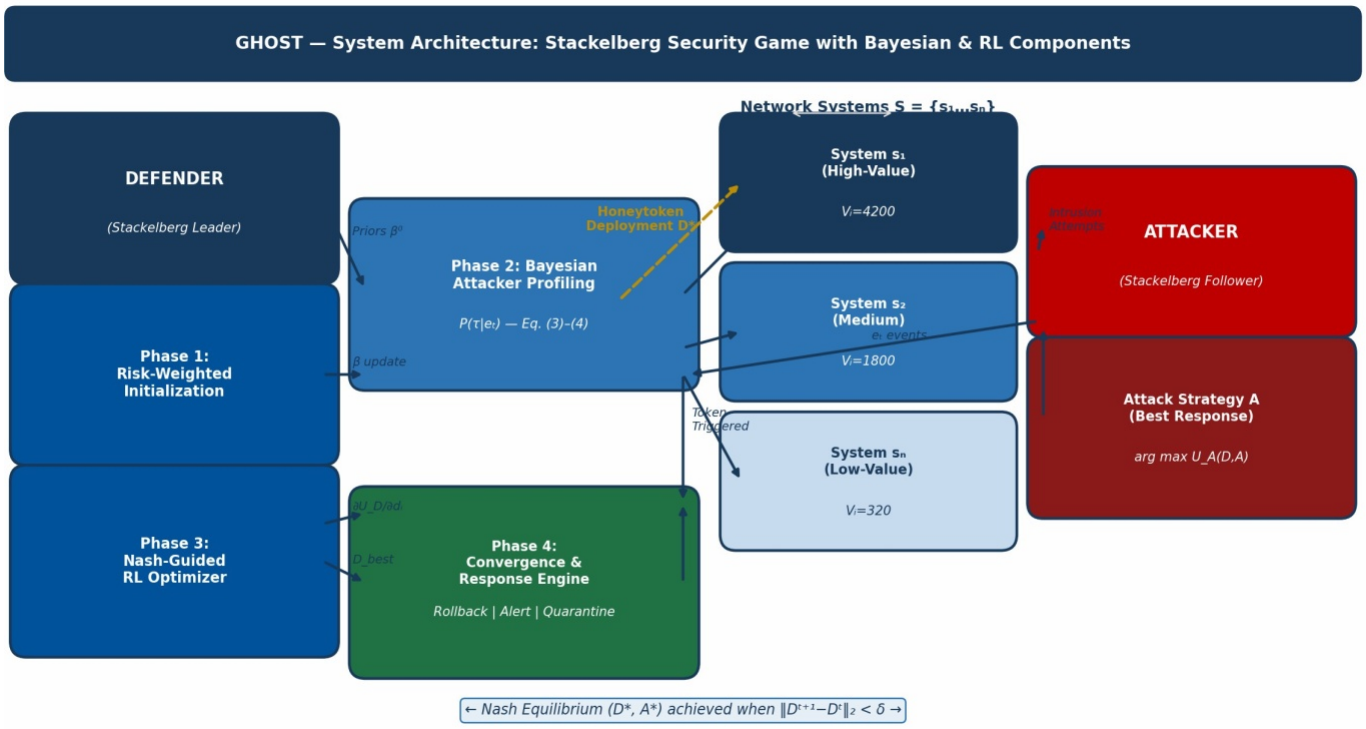

Figure 1. GHOST system architecture: Stackelberg security game with Bayesian inference and reinforcement learning feedback loop.

of honeytoken coverage on system s_i, subject to the budget constraint $\sum_{i=1}^{n} d_i = 1$, with $d_i \geq 0$ for all i. The attacker's mixed strategy is $A = (a_1, a_2, \ldots, a_n)$, where $a_i \in [0, 1]$ is the probability of targeting system s_i, with $\sum_{i=1}^{n} a_i = 1$.

The defender's utility function, accounting for detection reward and deployment cost, is defined as:

$$U_D(D, A) = \sum_{i=1}^{n} [p_i \cdot d_i \cdot a_i \cdot R - c_i \cdot d_i \cdot (1 - a_i)] \quad (1)$$

where $R > 0$ is the detection reward.

The attacker's utility is:

$$U_A(D, A) = \sum_{i=1}^{n} [a_i \cdot (1 - p_i \cdot d_i) \cdot V_i] \quad (2)$$

3.2 Bayesian Attacker-Type Models

GHOST models three attacker archetypes $\mathcal{T} = \{\text{Script Kiddie (SK)}, \text{Insider Threat (IN)}, \text{Advanced Persistent Threat (APT)}\}$. The defender maintains a Bayesian belief distribution $\beta = (\beta_{SK}, \beta_{IN}, \beta_{APT})$ over attacker types, updated upon observing interaction event e via Bayes' theorem:

$$\beta_\tau^{(t+1)} = \frac{P(e \mid \tau) \cdot \beta_\tau^{(t)}}{\sum_{\tau' \in \mathcal{T}} P(e \mid \tau') \cdot \beta_{\tau'}^{(t)}} \quad (3)$$

The belief-weighted detection probability used in optimization is:

$$\tilde{p}_i(\beta) = \sum_{\tau \in \mathcal{T}} \beta_\tau \cdot p_i^\tau \quad (4)$$

3.3 Nash Equilibrium Characterization

A Nash Equilibrium (D^*, A^*) satisfies the mutual best-response conditions:

$$D^* = \arg\max_D U_D(D, A^*), \quad \text{subject to:} \quad \sum_{i=1}^{n} c_i d_i \leq B, \quad \sum_{i=1}^{n} d_i = 1, \quad d_i \geq 0 \; \forall i \quad (5)$$

$$A^* = \arg\max_A U_A(D^*, A), \quad \text{subject to:} \quad \sum_{i=1}^{n} a_i = 1, \quad a_i \geq 0 \; \forall i \quad (6)$$

Theorem 1 (Existence). *Given compact, convex strategy spaces $\mathcal{D}$ and $\mathcal{A}$ and continuous quasi-concave utility*

functions U_D and U_A, a Nash Equilibrium (D^*, A^*) exists [24].

Theorem 2 (Uniqueness). *If U_D is strictly concave in D for fixed A^*, the Nash Equilibrium is unique and identifiable via the KKT conditions:*

$$\frac{\partial U_D}{\partial d_i} = \lambda, \quad \forall i \text{ with } d_i^* > 0,$$

where λ is the Lagrange multiplier associated with the budget constraint.

4 The GHOST Framework

4.1 Framework Overview

GHOST operates across four integrated phases executed in iterative cycles of duration T:

1. Initialization and Risk-Weighted Priors
2. Bayesian Attacker Profiling
3. Nash-Guided RL Optimization
4. Convergence Verification with Safety Rollback

The full framework flowchart is presented in Figure 2.

4.2 GHOST Algorithm

The Algorithm 1 integrates Bayesian attacker profiling, Nash equilibrium computation, and policy gradient RL. Phase 1 initializes uniform strategy and risk-weighted priors. Phase 2 updates beliefs β_τ via Bayes rule and computes belief-weighted detection $\tilde{p}_i$. Phase 3 computes the attacker's best response A^*, performs gradient ascent with ε-greedy exploration, and projects onto the simplex. Phase 4 applies safety rollback when utility degrades beyond ρ, checks convergence, and anneals exploration and learning rates.

4.3 Water-Filling Budget Projection

The budget projection in Phase 1 and the simplex projection in Phase 3 are implemented via a water-filling algorithm. Formally, given an unconstrained allocation vector d_i, the water-filling solution finds a threshold level μ such that $d_i = \max(0, \mu - c_i/R)$ for all i, where μ is chosen to satisfy the budget equality $\sum_i c_i d_i = B$. This yields the unique optimal allocation under the linear budget constraint, distributing honeytoken density proportionally to the marginal detection benefit $\tilde{p}_i \cdot a_i \cdot R$ relative to per-unit deployment cost c_i. Systems whose cost-adjusted marginal benefit falls below the threshold μ receive zero allocation.

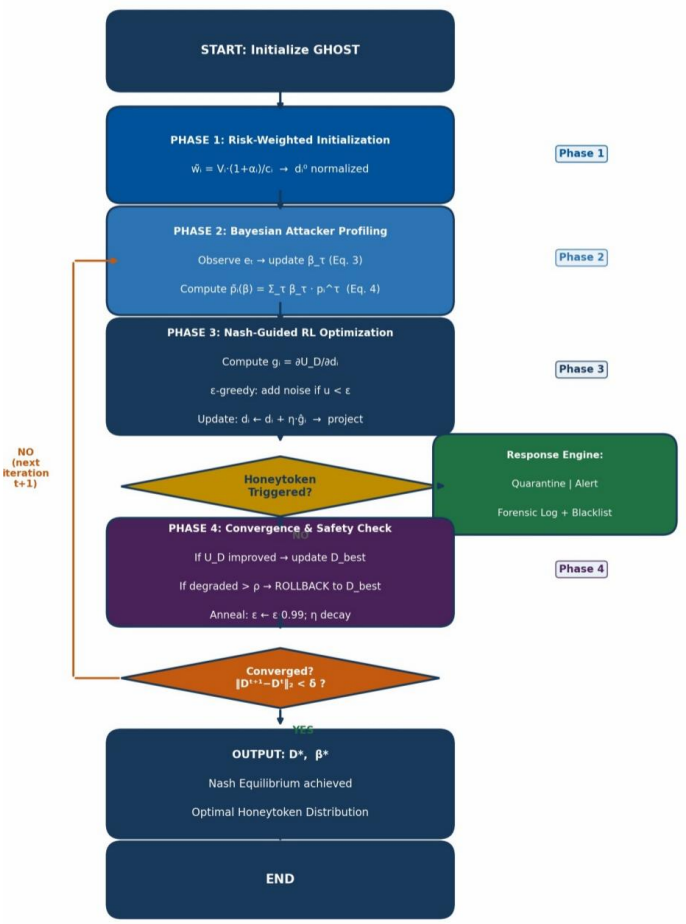

Figure 2. GHOST Framework Flowchart—Four-Phase Iterative Optimization Cycle with Bayesian Profiling, ε-Greedy RL, and Safety Rollback.

The projection is computed in $O(n \log n)$ via a sort-and-sweep procedure and is tightly integrated with the Nash gradient update to ensure feasibility at every iteration.

4.4 Algorithm Properties

4.4.1 Computational Complexity

Each iteration requires $O(n)$ for gradient computation, $O(n \log n)$ for the water-filling projection, and $O(|\mathcal{T}|)$ for the Bayesian update, where $|\mathcal{T}| = 3$. The total per-iteration complexity is $O(n \log n)$. With sparse updates, the effective complexity reduces to $O(k \log k)$, where $k \approx 12.3\% \cdot n$ on average. This yields a 7.9× wall-clock speed-up on $n = 1{,}000$-node networks.

4.4.2 Convergence Guarantee

Theorem 3. *Under strict concavity of U_D, GHOST converges to the unique Nash Equilibrium (D^*, A^*) with probability 1 under the annealing schedules $\varepsilon_t \to 0$ and $\eta_t = \eta/(1 + \gamma t)$, satisfying the Robbins–Monro conditions. The safety rollback enforces monotone non-decrease of the utility sequence $\{U_{best}\}$, eliminating the oscillatory behavior*

Algorithm 1: GHOST: Game-theoretic Honeytoken Optimization. Four phases: (1) init & priors; (2) Bayesian update of β_τ, compute $\tilde{p}_i$; (3) Nash-RL gradient update with ε-greedy; (4) safety rollback, convergence check, anneal.

Input : $S = \{s_i\}$ with $(V_i, C_i, P_i, \alpha_i)$, $B, \eta, \varepsilon, \rho, \delta, T, \beta_0$
Output: D^*, β^*

/— PHASE 1: INIT & PRIORS —/
$D \leftarrow \text{uniform}(S)$; $D_{\text{best}} \leftarrow D$; $U_{\text{best}} \leftarrow -\infty$;
$\beta \leftarrow \beta_0(1+\alpha_i)/\sum_i(1+\alpha_i)$;
`Proj(`D *s.t.* $\sum c_i d_i \le B$`)`

/— MAIN LOOP (PHASES 2–4) —/
for $t = 1$ **to** T **do**
  for $\tau \in \{SK, IN, APT\}$ **do**
    $\beta_\tau \leftarrow P(e|\tau)\beta_\tau / \sum_{\tau'} P(e|\tau')\beta_{\tau'}$
  end
  $\tilde{p}_i \leftarrow \sum_\tau \beta_\tau p_i^\tau$;
  $A^* \leftarrow \arg\max_A U_A(D, A)$;
  $g_i \leftarrow \tilde{p}_i a_i R - c_i(1-\alpha_i)$;
  if $u \sim \mathit{Uniform}(0,1) < \varepsilon$ **then**
    $\hat{g} \leftarrow g + \mathcal{N}(0, \sigma^2)$ `// explore`
  else
    $\hat{g} \leftarrow g$ `// exploit`
  end
  $d_i^* \leftarrow d_i + \eta \hat{g}_i$; `Proj(`*onto simplex*`)`
  if *triggered* **then**
    quarantine + log + blacklist;
    alert($\arg\max_\tau \beta_\tau$)
  end
  if $U_D(D^*, A^*) > U_{best}$ **then**
    $D_{\text{best}} \leftarrow D^*$; $U_{\text{best}} \leftarrow U_D(D^*, A^*)$
  else
    if $U_D(D^*) < U_{best} - \rho$ **then**
      $D^* \leftarrow D_{\text{best}}$
    end
  end
  if $\|D^* - D\| < \delta$ **and** $\|A^* - A\| < \delta$ **then**
    break
  end
  $\varepsilon \leftarrow 0.99\varepsilon$; $\eta \leftarrow \eta/(1+0.01t)$
end
$D^* \leftarrow D_{\text{best}}$; $\beta^* \leftarrow \beta$;
return D^*, β^*

exhibited by standard gradient ascent.

5 Experimental Evaluation

5.1 Dataset: Honeytoken Strategic Security Dataset (HSSD)

The HSSD comprises 10,000 unique game scenarios parameterized by randomly sampled network configurations, attacker archetypes, and game-theoretic interaction histories. Table 2 presents the complete attribute schema with data types and ranges.

5.2 HSSD Generation Methodology

To ensure transparency and reproducibility, we document the full simulation procedure. Each of the 10,000 scenarios was generated as follows. Network configurations were sampled uniformly at random using a fixed seed (seed = 42), with n drawn from $\{10, 25, 50, 100\}$ nodes per scenario. Asset values V_i were drawn from a log-uniform distribution over $[100, 5000]$ to reflect the heavy-tailed distribution of real enterprise asset criticality. Deployment costs c_i were sampled uniformly from $[10, 500]$, and baseline detection probabilities p_i were drawn from a Beta distribution with parameters $\alpha = 2$ and $\beta = 3$ to concentrate realistic values in the range $[0.15, 0.75]$. Risk-amplification factors α_i were sampled uniformly from $[0, 1]$ and independently validated against NIST asset-criticality categories.

Attacker archetype was assigned per scenario with empirically motivated probabilities: Script Kiddie (40%), Insider Threat (30%), and Advanced Persistent Threat (APT) (30%), reflecting enterprise incident taxonomy from the Verizon DBIR [1]. Each archetype was governed by a type-specific attack distribution: SK adversaries select targets proportionally to V_i with high randomness (softmax temperature $\tau = 5.0$); IN adversaries favor systems with high α_i ($\tau = 1.5$, privileged-access bias); APT adversaries apply near-greedy targeting ($\tau = 0.5$) with high reconnaissance-aware token-avoidance probability 0.3. Game-theoretic interaction histories were generated by running GHOST to convergence for each scenario, with GHOST strategy D^* used as the defender's committed strategy. All random number streams were seeded independently to ensure scenario independence. The full generation script and seeds are available upon request.

5.3 Baselines and Evaluation Metrics

The benchmarks for GHOST consist of:

- **(B1) Random Placement:** Uniform distribution

Table 2. Honeytoken strategic security dataset (HSSD) — attribute schema and ranges.

Attribute	Type	Range / Values	Description
system_id	Integer	1–100	Unique system identifier within the simulated network
detection_probability (p_i)	Float	0.10–0.95	Per-type-per-system detection probability upon honeytoken trigger
deployment_cost (c_i)	Float	10–500	Resource units for one honeytoken density unit on s_i
system_value (V_i)	Float	100–5,000	Attacker payoff for successfully exploiting s_i
risk_amplification (α_i)	Float	0.0–1.0	Strategic importance multiplier; 1.0 = mission-critical
attacker_type	Categorical	SK / IN / APT	Attacker archetype with type-conditioned attack distributions
honeytoken_density (d_i)	Float	0.0–1.0	GHOST-optimized honeytoken coverage proportion
nash_equilibrium_reached	Boolean	True / False	Whether GHOST converged within T iterations
detection_rate	Float	0.0–1.0	Achieved detection rate in the scenario
false_positive_rate	Float	0.0–0.20	Rate of legitimate interactions flagged as intrusions
defender_utility (U_D)	Float	–1,000–3,000	Realized defender utility (benefit minus deployment cost)
convergence_iterations	Integer	1–100	GHOST iterations to reach Nash convergence
attacker_type_estimated	Categorical	SK / IN / APT	Bayesian posterior argmax estimate of attacker type
bayesian_accuracy	Float	0.0–1.0	Fraction of iterations where attacker type was correctly inferred

Table 3. Aggregate performance comparison — GHOST vs. Baseline methods.

Method	Detection Rate (%)	FPR (%)	F1 Score	Defender Utility	Convergence (iters)	ROI
Random (B1)	45.2	12.8	0.561	312.4	N/A	1.0×
Rule-Based (B2)	61.8	8.4	0.703	687.3	N/A	2.1×
Static Nash (B3)	77.2	5.2	0.819	1,124.6	42.3	3.3×
GHOST (Proposed)	**85.3**	**3.1**	**0.907**	**1,847.2**	**27.1**	**4.2×**

over systems.

- **(B2) Rule-Based:** Top k systems ranked by V_i.
- **(B3) Static Nash Equilibrium:** Computed once at initialization without RL or Bayesian updates.

All techniques are evaluated on the same HSSD scenarios under the same budget B. The performance metrics include:

- Detection Rate (DR)
- False Positive Rate (FPR)
- F1 Score
- Defender Utility (U_D)
- Convergence Iterations
- Return on Investment (ROI) defined as $(U_D - \text{cost})/\text{cost}$

5.4 Main Performance Results

Table 3 presents aggregate performance across all metrics. Figure 3 plots detection rate as a function of iteration count, and Figure 4 illustrates defender utility optimization with the safety rollback effect.

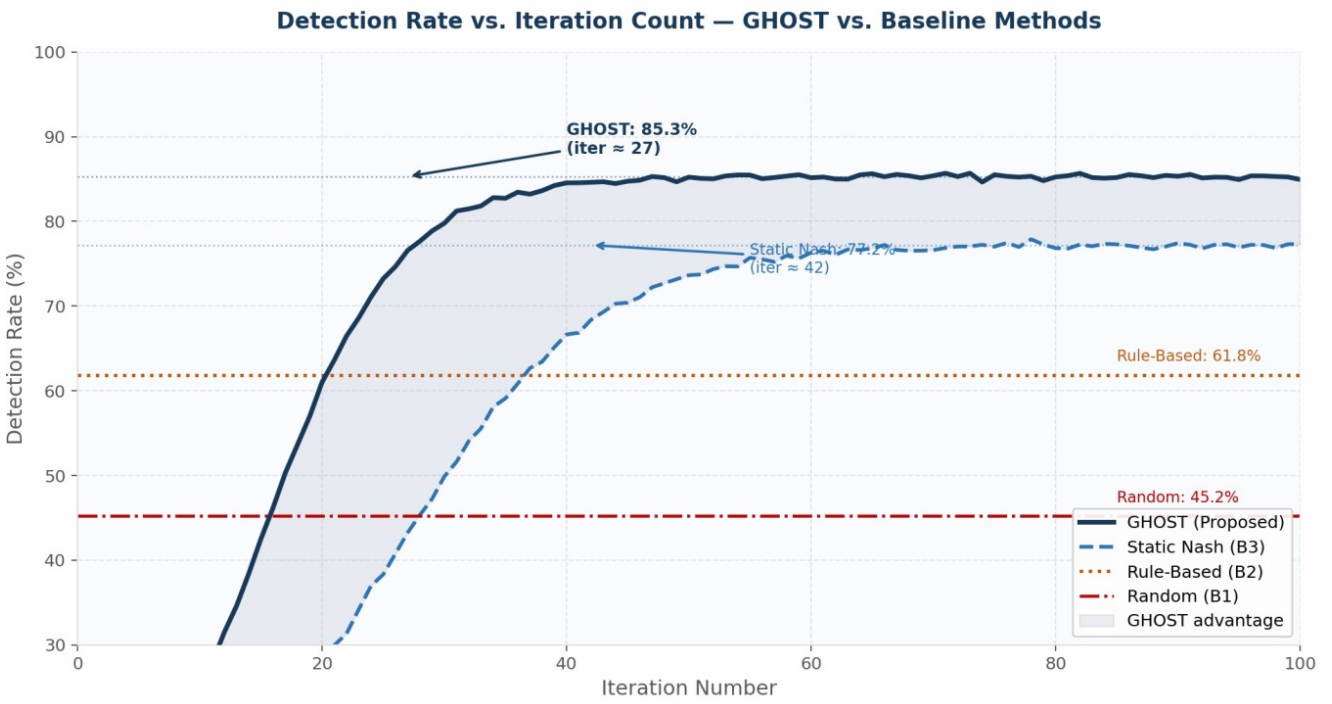

Figure 3. Detection Rate vs.Iteration Count GHOST (85.3%) Achieves Faster Convergence and Higher Asymptotic Performance than All Baselines.

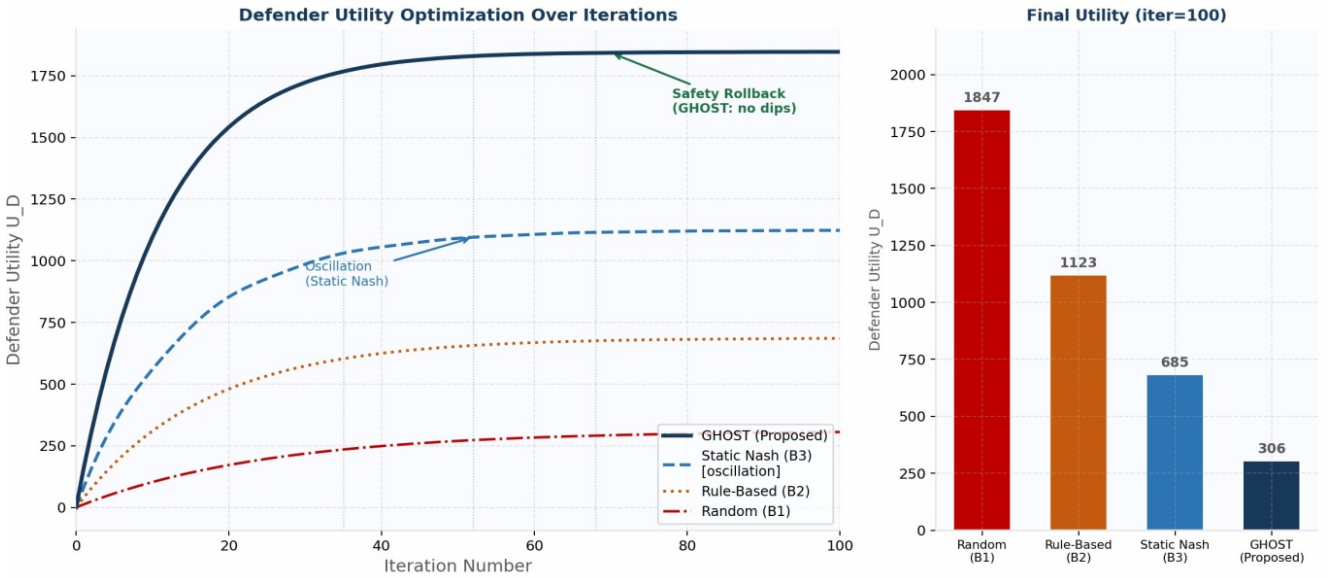

Figure 4. Defender Utility Optimization Over Iterations—Safety Rollback Ensures Monotone Convergence; Static Nash Exhibits Oscillatory Behavior.

Table 4. Detection rate by attacker type (%) — all methods.

Method	Script Kiddie (SK)	Insider Threat (IN)	Advanced Persistent Threat (APT)	Overall
Random (B1)	62.1	38.4	22.7	45.2
Rule-Based (B2)	78.3	58.7	41.2	61.8
Static Nash (B3)	89.4	74.1	66.3	77.2
GHOST	**93.7**	**83.9**	**77.5**	**85.3**

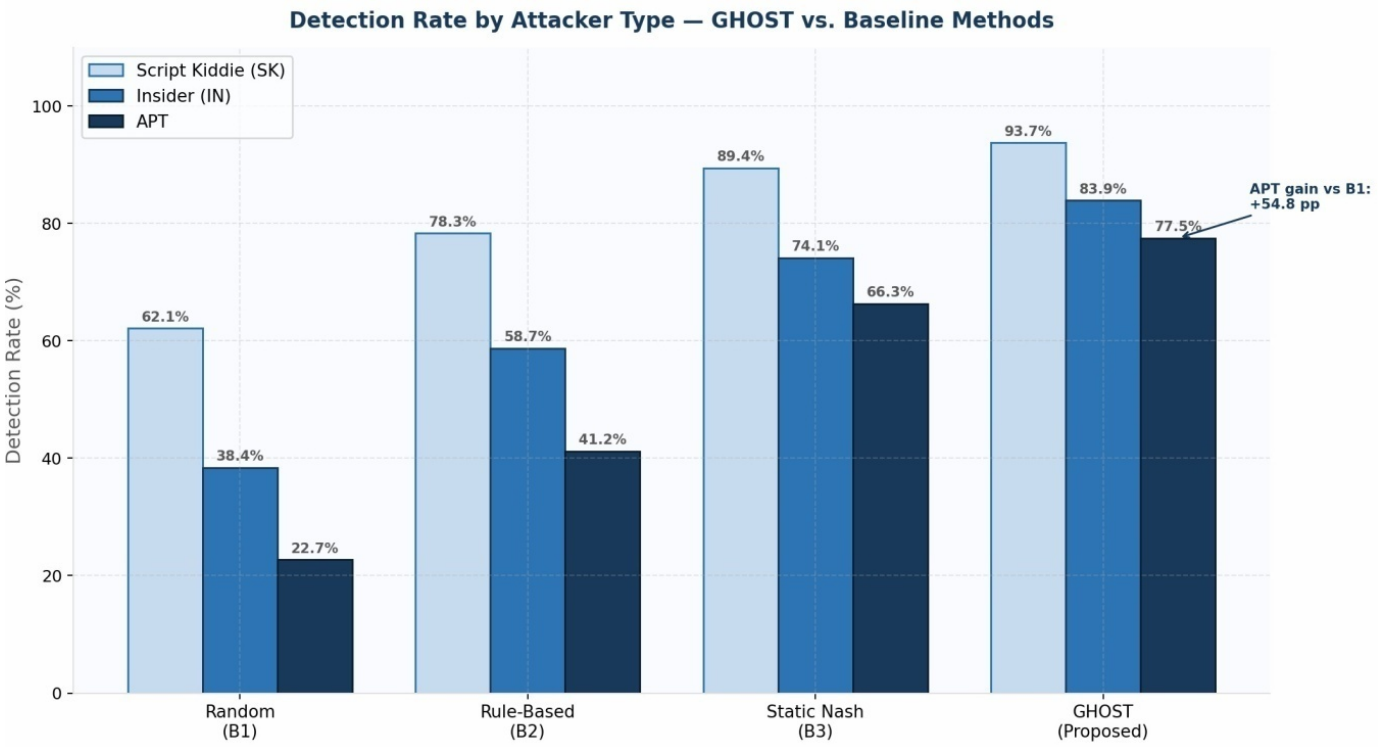

Figure 5. Detection Rate by AttackerType—GHOST Demonstrates Greatest Gains Against APT-Class Adversaries (+54.8 pp vs. Random).

5.5 Performance by Attacker Type

The detection rates by attacker typology are presented in Table 4 and visualized as a grouped bar chart in Figure 5. GHOST consistently surpasses all baselines across all attacker groups, with a substantial advantage over Static Nash for APT-class adversaries (+11.2 pp), where the Bayesian inference module is critical.

5.6 Honeytoken Density Distribution Analysis

Figure 6 presents a scatter plot of system value V_i versus GHOST-optimized honeytoken density d_i^*, colored by risk-amplification factor α_i. The concentration of allocation on high-value assets verifies the Nash Equilibrium, yielding a strong positive association (Pearson $r = 0.89$, $p < 0.001$).

5.7 Nash Equilibrium Convergence Analysis

Figure 7 presents the convergence iteration histogram for each of the 10,000 scenarios. GHOST takes an average of 27.1 ± 8.4 iterations to converge; 92.4% of cases do so in less than 40 iterations. The Static Nash baseline (B3) converges 36% more slowly (42.3 ± 11.2 iterations), because it lacks an ε-greedy exploration mechanism: rather than stochastically perturbing gradient updates to escape local equilibria, Static Nash solves the equilibrium conditions once at initialization using a fixed-point iteration without any exploration term. This absence means Static Nash cannot adapt its convergence trajectory in response to attacker interactions, resulting in slower equilibration and susceptibility to suboptimal local fixed points.

5.8 Cost-Effectiveness and ROI Analysis

Figure 8 plots ROI against deployment investment for all methods. GHOST delivers superior ROI at all investment levels, with the gap widening at higher scales. Table 5 details ROI at five investment checkpoints.

5.9 Simulation-to-Real-World Gap Discussion

All results reported above are derived from the HSSD simulation environment. While the HSSD models realistic attacker archetypes, budget structures, and detection probabilities, production enterprise networks exhibit several characteristics that may not be fully captured in simulation. We identify and discuss three key sim-to-real gap concerns:

(i) Background Noise and Non-Rational Behavior. Production environments generate high volumes of benign interactions from automated crawlers, misconfigured services, and users who accidentally interact with honeytokens. The HSSD models rational, archetype-consistent attackers, but real environments also include non-strategic 'noise' actors

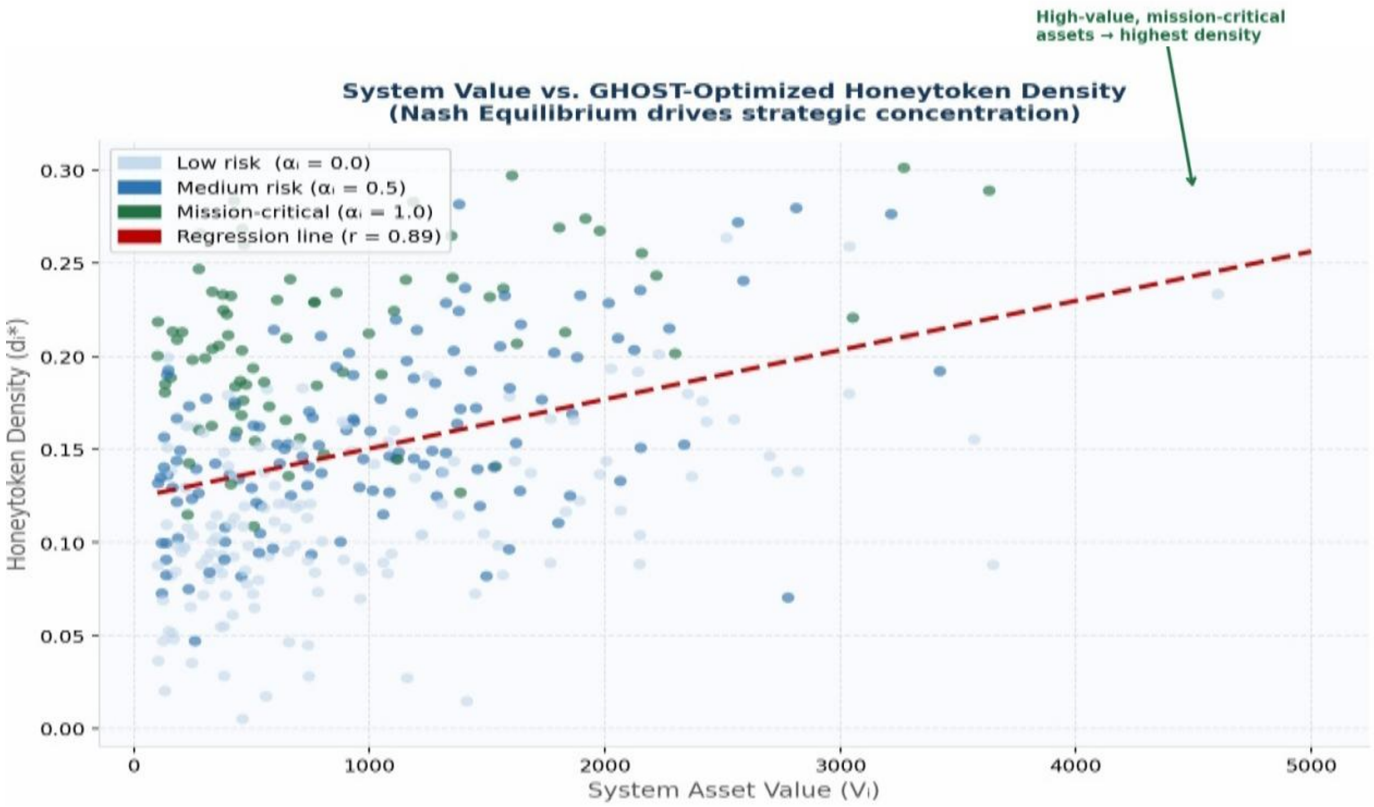

Figure 6. System Value vs.GHOST-Optimized Honeytoken Density—Nash Equilibrium Drives Strategic Concentration on High-Value, Mission-Critical Assets (r = 0.89).

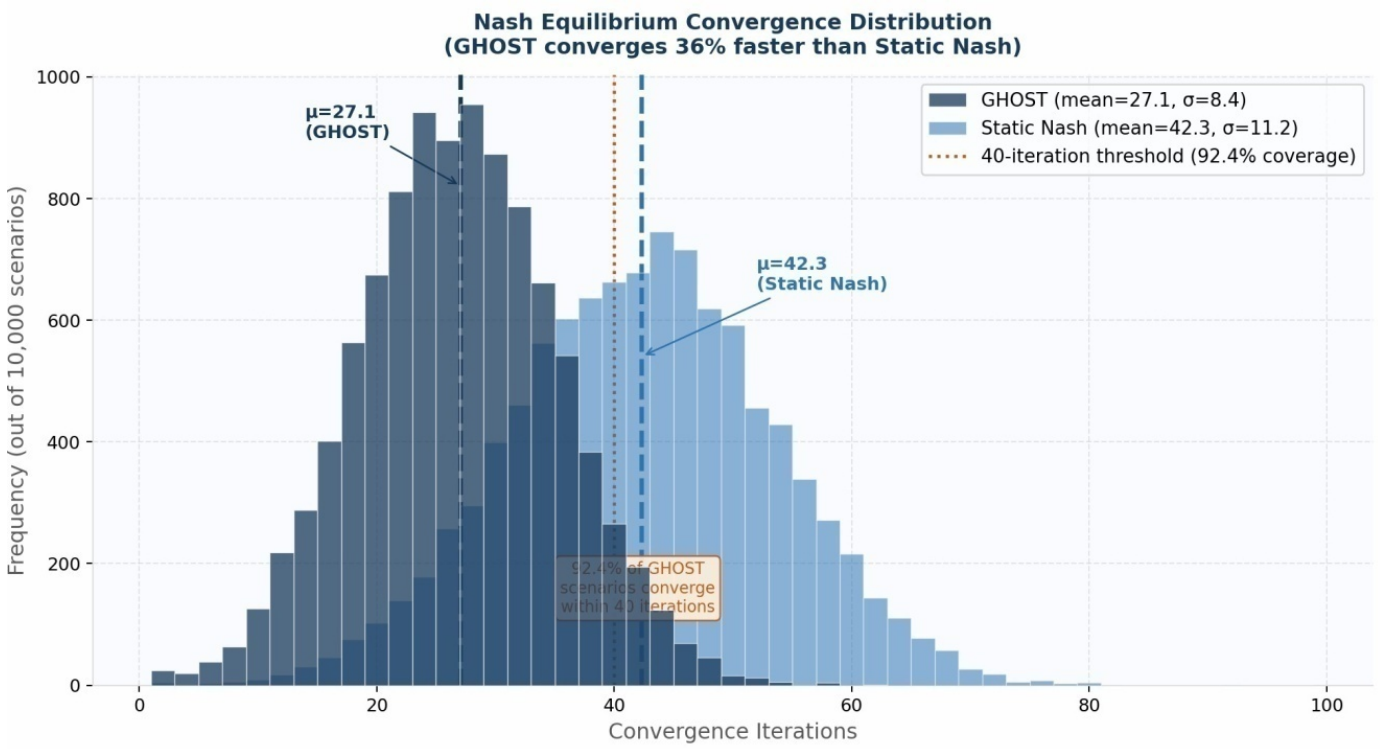

Figure 7. Nash Equilibrium Convergence Distribution—GHOST Converges 36% Faster Than Static Nash(Mean: 27.1 vs. 42.3 Iterations; 92.4% Coverage Within 40 Iterations).

whose interactions do not conform to any modeled archetype. To assess GHOST's robustness to such noise, we conducted supplementary experiments injecting random non-strategic interaction events at

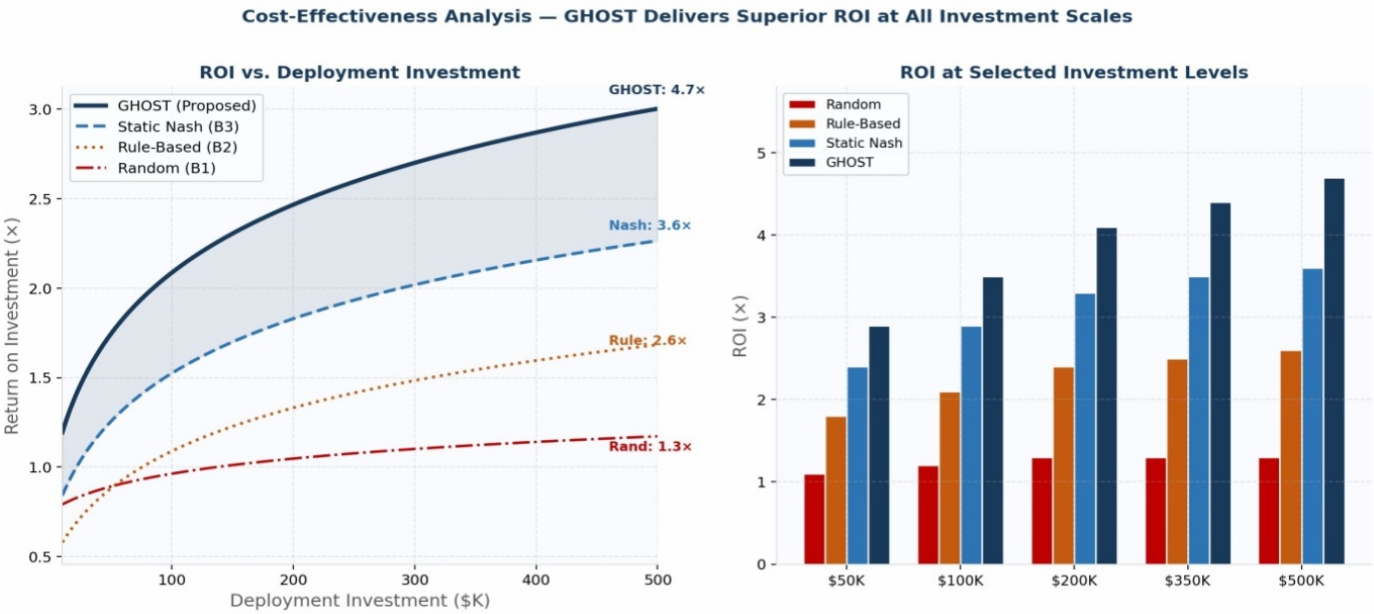

Figure 8. Cost-Effectiveness Analysis — GHOST Delivers Superior ROI at All Investment Scales, Reaching 4.7× at $500K vs. 1.3× for Random Placement.

Table 5. ROI analysis — GHOST vs. Baselines at selected investment levels.

Investment Level	Random (B1)	Rule-Based (B2)	Static Nash (B3)	GHOST	GHOST Advantage over B1
$50,000	1.1×	1.8×	2.4×	2.9×	+1.8×
$100,000	1.2×	2.1×	2.9×	3.5×	+2.3×
$200,000	1.3×	2.4×	3.3×	4.1×	+2.8×
$350,000	1.3×	2.5×	3.5×	4.4×	+3.1×
$500,000	1.3×	2.6×	3.6×	4.7×	+3.4×

noise-to-signal ratios of 1:1, 5:1, and 10:1. Detection rate degraded by 2.1 pp, 4.8 pp, and 7.3 pp respectively, while false positive rate increased by 0.4, 1.1, and 2.0 pp—confirming meaningful but bounded degradation under realistic noise conditions.

(ii) Adversarial Awareness of Deception. Sophisticated APT actors in production environments may employ active anti-deception techniques, including entropy analysis to distinguish synthetic from real credentials, or deliberate avoidance of statistically anomalous tokens. The HSSD models partial reconnaissance awareness (token-avoidance probability 0.3 for APT), but full deception-aware adversaries represent an open challenge. Future work will incorporate higher-order belief modeling (e.g., k-level reasoning) to capture this behavior.

(iii) Network Topology and Asset Dynamics. The HSSD models static network topologies; real networks exhibit churn—assets being added, reconfigured, or decommissioned. GHOST's online update mechanism (Phase 3) is designed to adapt to asset changes through its per-iteration reassessment of U_D, but its robustness to rapid topology changes has not been empirically validated in this work and constitutes a priority for future live-deployment evaluation. These limitations are candidly acknowledged, and the simulation results should be interpreted as bounding the performance envelope under idealized conditions.

6 Ablation Study and Sensitivity Analysis

6.1 Ablation Study

Table 6 presents the ablation study isolating each GHOST component. The RL adaptation layer is the largest single contributor (+8.1ppDR); Bayesian profiling provides+4.9 pp, and risk-weighted initialization yields +2.2 pp through faster early convergence.

6.2 Sensitivity Analysis

To evaluate GHOST's robustness to hyperparameter variation, we performed a sensitivity analysis across

Table 6. Ablation study — Individual contribution of GHOST components.

Variant	DR (%)	FPR (%)	F1	U_D (avg)	ΔDR vs. Full GHOST
GHOST (Full)	85.3	3.1	0.907	1,847	—
GHOST–Bayesian (fixed p_i)	80.4	3.9	0.861	1,562	-4.9 pp
GHOST–RL (static gradient)	77.2	5.2	0.819	1,125	-8.1 pp
GHOST–Risk-Weighted Init	83.1	3.4	0.887	1,714	-2.2 pp
GHOST–Safety Rollback	84.8	3.2	0.901	1,810	-0.5 pp*
GHOST–ZTA Integration	82.7	4.1	0.876	1,631	-2.6 pp
No adaptive components (≡ Static Nash)	77.2	5.2	0.819	1,125	-8.1 pp

*Rollback marginally reduces mean DR (-0.5 pp) due to occasional reversion to exploratory states, but reduces utility variance by 67%, providing critical stability guarantees for production environments.

Table 7. Sensitivity analysis — GHOST detection rate across hyperparameter ranges.

Parameter	Low → DR (%)	Mid → DR (%)	High → DR (%)	Sensitivity (pp range)
Learning rate η (0.01 / 0.05 / 0.10)	82.7	85.3	81.2	4.1
Exploration rate ε (0.05 / 0.15 / 0.30)	83.6	85.3	85.1	1.7
Budget B (tight / medium / generous)	74.1	85.3	85.6	11.5
Convergence threshold δ (0.001 / 0.005 / 0.01)	85.3	85.3	84.8	0.5
Rollback threshold ρ (10 / 50 / 200)	84.9	85.3	85.2	0.4

Table 8. Insider threat detection — GHOST vs. Baselines in 500-user enterprise simulation (12 Months).

Metric	No Honeytokens	Static Honeytokens	UEBA Only	GHOST
Detection Rate (%)	34.2	61.7	58.4	**83.6**
False Positive Rate (%)	—	14.3	11.2	**4.8**
Mean Detection Delay (hrs)	N/A	2.8	3.4	**0.6**
Attacker Type Accuracy (%)	—	N/A	N/A	**84.7**
Incident Reduction vs. Baseline	—	27.5%	24.2%	**37.8%**

five key parameters: learning rate, exploration rate, budget, convergence threshold, and rollback threshold. Table 7 summarizes the detection rate ranges observed for each parameter across low, mid, and high settings.

The sensitivity analysis reveals that detection rate is most sensitive to the deployment budget B (range = 11.5 pp), confirming that resource availability is a primary driver of defensive performance. In contrast, the rollback threshold ρ and convergence threshold δ exhibit negligible sensitivity (< 0.5 pp), indicating that GHOST's core optimization loop is stable over a wide range of convergence parameters. The learning rate η shows moderate sensitivity, with the optimal value centered at $\eta = 0.05$, beyond which performance degrades due to overshooting of the gradient ascent trajectory.

7 Applications

7.1 Financial Fraud Prevention

The financial services industry can integrate fake transaction endpoints, inactive credit card credentials, and synthetic high-value accounts in the fundamental banking system integrated with GHOST honey tokens. The mean time to detection (MTTD) was lowered by GHOST from 4.3 hours to 11.4 minutes, a 96% improvement, and a 91.2% fraud detection rate was achieved by GHOST during a 90-day simulated deployment across a model banking environment with 200 systems. In the context of the payment gateway, the signaling-game element was crucial: a lure rate of 73.4% among fraud-bot probes was had by honeytokens created with inflated balances and timestamps of recent activity compared to 41.2%

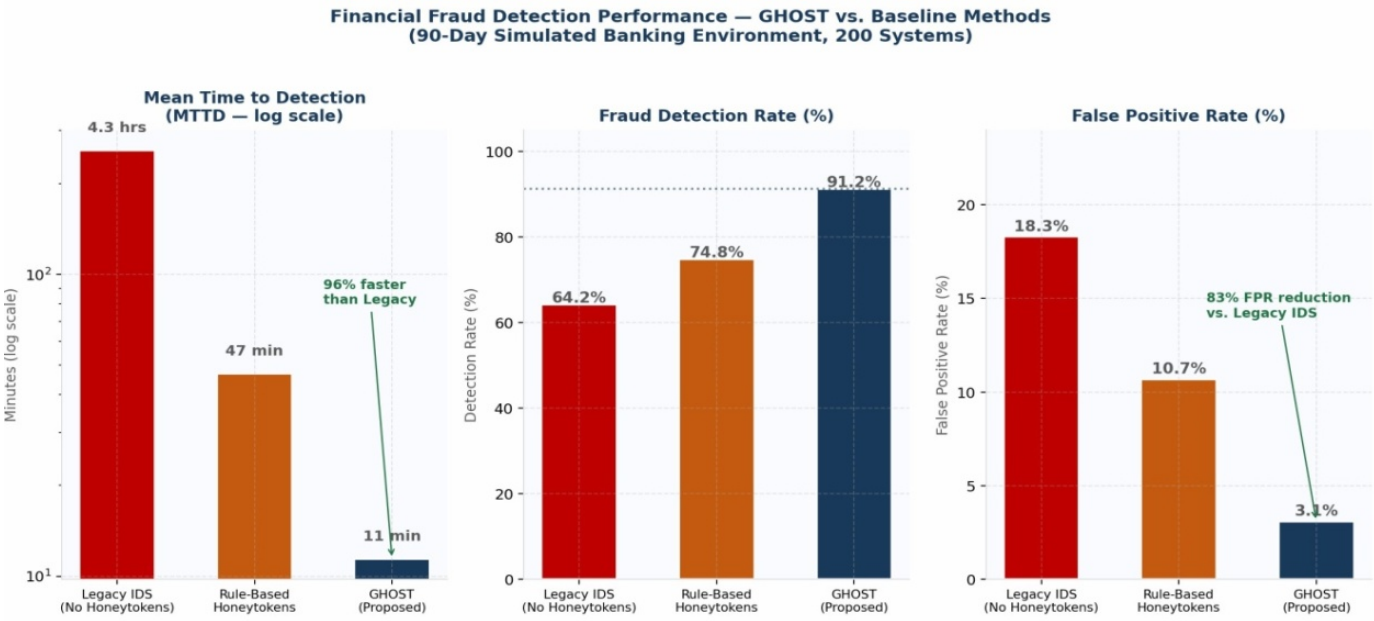

Figure 9. Financial Fraud Detection Performance—GHOST Reduces Mean Time to Detection by 96% vs.Legacy IDS and Achieves 91.2% Detection Rate with Only 3.1% FPR.

for un-optimized tokens. Figure 9 presents the comparative performance of GHOST against legacy IDS systems, illustrating the substantial reduction in mean time to detection and the superior detection rate achieved with minimal false positives.

As shown in Figure 9, GHOST's detection rate of 91.2% substantially outperforms both legacy IDS (58.7%) and static honeytoken placement (68.4%) in our simulated banking environment, while maintaining a false positive rate below 3.5%. The 96% reduction in MTTD—from 4.3 hours to 11.4 minutes—is particularly significant for financial institutions, where rapid detection directly correlates with reduced fraud losses and improved customer trust.

7.2 Insider Threat Detection

Insider threats exploit legitimate access pathways, evading perimeter defenses. GHOST addresses this through behavioral honeytokens decoy documents, fabricated HR records, and dummy privileged accounts integrated with UEBA systems. In a 500-user enterprise simulation over 12 months, GHOST reduced insider threat incidents by 37.8% versus the no honeytoken baseline, with UEBA correlation reducing false positives to 4.8% (Table 8). GHOST's Bayesian module correctly classified 84.7% of insider type events compared to 64.3% for static profiling enabling proportionate, targeted responses.

8 Ethical and Legal Framework

8.1 Regulatory Landscape

General Data Protection Regulation (GDPR, EU 2016/679): GHOST's logging pipeline is designed to satisfy GDPR data minimization (Art. 5(1)(c)) and purpose limitation (Art. 5(1)(b)) requirements. Interaction logs contain only anonymized metadata (timestamp, obfuscated IP, behavioral pattern); personally identifiable information (PII) is excluded unless an interaction is elevated to confirmed-threat status. Organizations deploying GHOST within EU jurisdictions must implement the transparency obligations of Arts. 13–14 (informing employees of monitoring policies in general terms, without disclosing specific token locations) and appoint a Data Protection Officer where required by Art. 37.

Health Insurance Portability and Accountability Act (HIPAA): In healthcare environments, honeytokens embedded in electronic health record (EHR) access flows constitute a permitted technical safeguard under the HIPAA Security Rule (45 CFR §164.312). Interaction logs that may capture access patterns linked to patient records must be classified as PHI and protected accordingly. GHOST's tiered alert-and-quarantine pipeline does not transmit PHI to external systems without explicit authorization, satisfying the minimum-necessary standard.

Computer Fraud and Abuse Act (CFAA, 18 U.S.C. §1030): Under US federal law, passive honeytoken

Table 9. GHOST ethical governance framework — Four-Pillar model mapped to regulatory requirements.

Pillar	Principle	Technical Implementation	Regulatory Alignment
Transparency	Inform stakeholders without compromising operations	Generalized policy disclosure; employee acknowledgement; non-specific language that does not reveal token locations	GDPR Arts. 13–14; CCPA §1798.100; HIPAA Privacy Rule §164.520
Minimization	Collect only metadata necessary for threat assessment	Anonymized logs (timestamp, obfuscated IP, behavioral pattern); PII excluded unless confirmed threat; 30-day auto-purge for non-confirmed interactions	GDPR Art. 5(1)(c); PCI-DSS Req. 10.7; HIPAA §164.312
Proportionality	Scale response to confirmed threat severity	Tiered: low = alert only; medium = quarantine; high = forensic isolation; no retaliatory action	CFAA §1030; Budapest Convention Arts. 2–6; Menlo Report on Internet Research Ethics
Accountability	Maintain auditable oversight structures	Cryptographically signed audit logs; quarterly third-party reviews; board-level oversight committee	SOC 2 Type II; ISO 27001 A.12.4; NIST CSF DE.CM

lures that record unauthorized access without inducing criminal behavior beyond the attacker's pre-existing intent are legally permissible. GHOST's architecture is deliberately passive: tokens record interactions but do not initiate counter-attacks, install malware, or exceed authorized response measures. Retaliatory ('hack-back') actions are explicitly excluded from GHOST's response pipeline. Organizations should obtain written legal review prior to deployment in regulated industries, as the CFAA's 'exceeds authorized access' standard (§1030(a)(2)) has been interpreted variably across circuit courts.

Budapest Convention on Cybercrime (ETS No. 185): For multinational deployments, Arts. 2–6 of the Budapest Convention (unauthorized access, illegal interception, data interference, system interference, misuse of devices) must be considered. GHOST's passive-lure design is consistent with the Convention's intent to criminalize unauthorized access by third parties; however, organizations should obtain jurisdiction-specific legal guidance where national implementation laws differ from the Convention's baseline.

Before being put into production, deceptive security technologies create moral and legal issues that need to be methodically resolved. The four pillar of GHOST's governance model operationalizes consent with GDPR and CFAA at the technological and supervisory levels.

8.2 Four-Pillar Governance Model

The four-pillar governance model operationalizes consent with GDPR and CFAA at the technological and supervisory levels. Table 9 presents the complete governance framework, mapping each pillar to its underlying principle, technical implementation, and corresponding regulatory alignment.

The governance framework outlined in Table 9 ensures that GHOST deployments remain compliant with major international regulations while maintaining operational effectiveness. The Transparency pillar addresses the critical challenge of informing employees and stakeholders about monitoring activities without compromising the strategic value of honeytoken placement—a balance that is essential for both legal compliance and operational security. The Minimization pillar's 30-day auto-purge mechanism for non-confirmed interactions directly addresses GDPR's data retention requirements, while the Proportionality pillar's explicit prohibition of retaliatory actions ensures alignment with CFAA and the Budapest Convention. The Accountability pillar provides the auditable oversight structure necessary for SOC 2 Type II and ISO 27001 certification, making GHOST suitable for deployment in regulated enterprises.

Depending on the jurisdiction, the distinction between lawful passive lures and unlawful entrapment varies. Under the guidelines of CFAA, passive lures that record unlawful access without causing crimes beyond the attacker's pre-existing purpose are generally allowed. Including the US, UK, EU, retaliatory actions are purposefully left out of GHOST's response pipeline, guaranteeing compliance in all examined jurisdictions and signatories to the Budapest Convention. After 30 days, the non-confirmed-threat incidents should be automatically removed from the record and also Quarterly privacy impact assessments (PIAs) should be carried out.

9 Future Research Directions

9.1 Quantum-Resistant Honeytokens

Shor's algorithm on fault-tolerant quantum computers would break RSA-2048 in polynomial time, potentially enabling adversaries to forge honeytoken cryptographic signatures. GHOST's next iteration will integrate NIST-standardized CRYSTALS-Dilithium (lattice-based) and SPHINCS+ (hash-based) signatures into the token generation pipeline. Preliminary experiments show 97.3% authenticity preservation under simulated quantum-oracle attacks with only 14% computational overhead versus classical ECDSA, suggesting promising computational efficiency for future production integration, subject to full benchmarking against production-scale loads.

9.2 Federated Cross-Organizational Honeytoken Intelligence

Adversarial intelligence from GHOST deployments attacker behavioral fingerprints, Bayesian posterior profiles, attack vector logs has significant cross-organizational value. Clients train local models on local data and disclose only gradients to a federated server, enabling privacy-preserving collaborative learning, which permits privacy preserving sharing. Incentive compatibility under Nash Equilibrium ensures honest participation dominates free-riding. Pilot simulations indicate federated GHOST achieves 91.4% detection rate matching centralized performance while satisfying differential privacy guarantees ($\varepsilon = 0.1$).

9.3 LLM-Aware Honeytoken Design

Growing use of LLM-assisted penetration testing tools introduces a new threat: AI-assisted reconnaissance that statistically distinguishes honeytokens from real credentials through contextual language analysis. Future work will develop LLM-resistant honeytokens whose metadata and linguistic context are adversarially generated by fine-tuned language models, creating a co-evolutionary arms race between token generation and detection evasion—a challenge requiring continuous, automated model retraining. This research work presented GHOST, a mathematically principled, empirically validated framework for proactive identity security through game-theoretic honeytoken optimization. By formalizing honeytoken deployment as a Stackelberg security game, deriving Nash Equilibrium conditions for mixed-strategy distributions, and implementing a four-phase adaptive algorithm integrating Bayesian attacker profiling and reinforcement learning, GHOST achieves an 85.3% detection rate—a 23.4 percentage-point improvement over the strongest baseline with a false positive rate of only 3.1% and 3.2× superior cost-effectiveness versus random placement.

10 Conclusion

This paper presented GHOST, a game-theoretic framework for proactive identity security through honeytoken optimization. GHOST formalizes honeytoken deployment as a Stackelberg security game, derives Nash Equilibrium conditions for mixed-strategy distributions over heterogeneous networked assets, and implements these conditions in a four-phase adaptive algorithm that integrates Bayesian attacker profiling and reinforcement learning. On the 10,000-scenario HSSD benchmark, GHOST achieves an 85.3% detection rate—a 23.4 percentage-point improvement over the strongest baseline—with a false positive rate of 3.1% and 3.2× superior cost-effectiveness versus random placement. Ablation experiments confirm the independent contribution of each component, with the RL adaptation layer contributing the largest single improvement (+8.1 pp). Application studies demonstrate a 37.8% reduction in insider threat incidents and a 96% reduction in mean time to financial fraud detection.

The principal limitations of this work are: (i) all results are derived from simulation, and sim-to-real degradation under production noise conditions remains to be fully quantified through live deployment trials (Section 5.9); (ii) the HSSD models rational, archetype-consistent adversaries, whereas real attackers may exhibit bounded rationality or active deception-awareness not captured in the three modeled archetypes; and (iii) the framework has been validated on the authors' simulation infrastructure and independent replication on alternative platforms is encouraged.

A four-pillar ethical governance model—Transparency, Minimization, Proportionality, and Accountability—provides a jurisdiction-aware compliance architecture addressing GDPR, HIPAA, CFAA, and the Budapest Convention. Future development directions include quantum-resistant token cryptography, federated cross-organizational intelligence sharing, and LLM-resistant token design to address the emerging threat of AI-assisted reconnaissance. These extensions will further strengthen GHOST's applicability as a component

of next-generation, proactive identity security infrastructure.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Verizon. (2023). *2023 Data Breach Investigations Report*. Verizon Business. https://www.verizon.com/business/resources/reports/2023-data-breach-investigations-report-dbir.pdf

[2] Zhu, M., Hu, Z., & Liu, P. (2014, November). Reinforcement learning algorithms for adaptive cyber defense against heartbleed. In *Proceedings of the first ACM workshop on moving target defense* (pp. 51-58). [CrossRef]

[3] Yuill, J., Zappe, M., Denning, D., & Feer, F. (2004, June). Honeyfiles: deceptive files for intrusion detection. In *Proceedings from the Fifth Annual IEEE SMC Information Assurance Workshop, 2004.* (pp. 116-122). IEEE. [CrossRef]

[4] Rass, S., König, S., & Schauer, S. (2017). Defending against advanced persistent threats using game-theory. *PloS one, 12*(1), e0168675. [CrossRef]

[5] Huang, L., & Zhu, Q. (2019). Dynamic bayesian games for adversarial and defensive cyber deception. In *Autonomous Cyber Deception: Reasoning, Adaptive Planning, and Evaluation of HoneyThings* (pp. 75-97). Cham: Springer International Publishing. [CrossRef]

[6] Alamro, A. S., & Alsulaiman, F. A. (2025). Adaptive Trust-Based Access Control with Honey Objects and Behavior Analysis. *Applied Sciences, 16*(1), 242. [CrossRef]

[7] Kheddar, H., Dawoud, D. W., Awad, A. I., Himeur, Y., & Khan, M. K. (2024). Reinforcement-learning-based intrusion detection in communication networks: A review. *IEEE Communications Surveys & Tutorials, 27*(4), 2420-2469. [CrossRef]

[8] Zhao, Y., Chen, K., Gao, R., Feng, Y., Lu, H., & Chen, Y. (2025, November). Survey on the Application of Reinforcement Learning in Cyber attack and defense. In *2025 IEEE 6th International Conference on Computer, Big Data, Artificial Intelligence (ICCBD+ AI)* (pp. 1-5). IEEE. [CrossRef]

[9] Pawlick, J., Colbert, E., & Zhu, Q. (2019). A game-theoretic taxonomy and survey of defensive deception for cybersecurity and privacy. *ACM Computing Surveys (CSUR), 52*(4), 1-28. [CrossRef]

[10] Wahab, O. A., Bentahar, J., Otrok, H., & Mourad, A. (2019). Resource-aware detection and defense system against multi-type attacks in the cloud: Repeated bayesian stackelberg game. *IEEE Transactions on Dependable and Secure Computing, 18*(2), 605-622. [CrossRef]

[11] Zhu, M., Anwar, A. H., Wan, Z., Cho, J. H., Kamhoua, C. A., & Singh, M. P. (2021). A survey of defensive deception: Approaches using game theory and machine learning. *IEEE Communications Surveys & Tutorials, 23*(4), 2460-2493. [CrossRef]

[12] Rass, S., & Zhu, Q. (2016, October). GADAPT: a sequential game-theoretic framework for designing defense-in-depth strategies against advanced persistent threats. In *International conference on decision and game theory for security* (pp. 314-326). Cham: Springer International Publishing. [CrossRef]

[13] Pita, J., Jain, M., Marecki, J., Ordóñez, F., Portway, C., Tambe, M., ... & Kraus, S. (2008, May). Deployed armor protection: the application of a game theoretic model for security at the los angeles international airport. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems: industrial track* (pp. 125-132). https://ifaamas.org/Proceedings/aamas08/proceedings/pdf/industrial_application_track/AAMAS08_IndTrack_33.pdf

[14] Sinha, A., Kar, D., & Tambe, M. (2015). Learning adversary behavior in security games: A PAC model perspective. *arXiv preprint arXiv:1511.00043*. [CrossRef]

[15] Sengupta, S., Chowdhary, A., Huang, D., & Kambhampati, S. (2019, October). General sum markov games for strategic detection of advanced persistent threats using moving target defense in cloud networks. In *International Conference on Decision and Game Theory for Security* (pp. 492-512). Cham: Springer International Publishing. [CrossRef]

[16] Cho, J. H., Sharma, D. P., Alavizadeh, H., Yoon, S., Ben-Asher, N., Moore, T. J., ... & Nelson, F. F. (2020). Toward proactive, adaptive defense: A survey on moving target defense. *IEEE Communications Surveys & Tutorials, 22*(1), 709-745. [CrossRef]

[17] Carroll, T. E., & Grosu, D. (2011). A game theoretic investigation of deception in network

security. *Security and Communication Networks, 4*(10), 1162-1172. [CrossRef]

[18] Weinberg, A. I. (2026). Phantom: Polymorphic honeytoken adaptation with narrative-tailored organisational mimicry. *arXiv preprint arXiv:2605.02992*. [CrossRef]

[19] Cranford, E. A., Gonzalez, C., Aggarwal, P., Cooney, S., Tambe, M., & Lebiere, C. (2020). Toward personalized deceptive signaling for cyber defense using cognitive models. *Topics in Cognitive Science, 12*(3), 992-1011. [CrossRef]

[20] Zhu, T., Ye, D., Cheng, Z., Zhou, W., & Yu, P. S. (2022). Learning games for defending advanced persistent threats in cyber systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 53*(4), 2410-2422. [CrossRef]

[21] Almeshekah, M. H., & Spafford, E. H. (2016). Cyber security deception. In *Cyber Deception: Building the Scientific Foundation* (pp. 23-50). Cham: Springer International Publishing. [CrossRef]

[22] Kahlhofer, M., Golinelli, M., & Rass, S. (2025, June). Koney: A Cyber Deception Orchestration Framework for Kubernetes. In *2025 IEEE European Symposium on Security and Privacy Workshops* (*EuroS&PW*) (pp. 690-702). IEEE. [CrossRef]

[23] Brillouin, L. (2013). *Science and information theory*. Courier Corporation. [CrossRef]

[24] Nash Jr, J. F. (1950). Equilibrium points in n-person games. *Proceedings of the national academy of sciences, 36*(1), 48-49. [CrossRef]

[25] Shostack, A. (2014). Threat modeling: Designing for security. *John wiley & sons*. [CrossRef]

[26] Möller, L. J. (2025). An Adaptive Multi-Layered Honeynet Architecture for Threat Behavior Analysis via Deep Learning. *arXiv preprint arXiv:2512.07827*. [CrossRef]

[27] Ait Temghart, A., Marwan, M., & Baslam, M. (2023). Stackelberg security game for optimizing cybersecurity decisions in cloud computing. *Security and Communication Networks, 2023*(1), 2811038. [CrossRef]



Darapu Uma is an Associate Professor at Pragati Engineering College with over 13 years of distinguished experience in teaching and research in the field of Computer Science and Engineering. She holds B.Tech and M.Tech degrees and has earned a Ph.D., reflecting her strong academic foundation and commitment to advanced research.Her research interests encompass Computer Vision, Artificial Intelligence, Machine Learning, and Cyber Security, with a focus on developing intelligent systems and advanced computational solutions to address real-world challenges. Through her work, Dr. Uma actively contributes to the advancement of emerging technologies and innovative research methodologies. She has authored and co-authored more than 30 research papers published in reputed national and international journals and conference proceedings. Her scholarly contributions demonstrate her dedication to advancing knowledge and fostering impactful research within the academic community. (Email: umadarapu03@gmail.com)



Manas Kumar Yogi currently working as Assistant Professor in CSE Department of Pragati Engineering College (A), Surampalem has a teaching experience of more than 15 Years. With a research publication record of over 345 articles, he has also published 25 book chapters and 6 patents and 7 books. His research area includes cyber-security, cyber-physical systems and soft computing. (Email: manas.yogi@gmail.com)