# A Multi-Domain Cyber-Deception Bayesian Game for Resilient IoT Systems

Alioum Abdoulaye
*Laboratoire DAVID, Université Paris-Saclay,*
*Université de Versailles Saint-Quentin-en-Yvelines*
*Department of Mathematics and Computer Science,*
*University of Ngaoundere*
Versailles, France / Ngaoundere, Cameroon
aliyoumabdoulaye@gmail.com

Yacine Benallouche
*Orange Research*
*Université Sorbonne Paris Nord*
Paris, France
yacine.benallouche@orange.com

Abdelhak Mourad Gueroui
*Laboratoire DAVID, Université Paris-Saclay,*
*Université de Versailles Saint-Quentin-en-Yvelines*
Versailles, France
mourad.gueroui@uvsq.fr

Ado Adamou Abba Ari
*Laboratoire DAVID, Université Paris-Saclay,*
*Université de Versailles Saint-Quentin-en-Yvelines*
*LaRI Lab, University of Maroua*
Versailles, France / Maroua, Cameroon
ado-adamou.abba-ari@uvsq.fr

***Abstract*—Internet of Things (IoT) infrastructures are increasingly targeted by Advanced Persistent Threats (APTs) that progress methodically across the perception, network, and application layers. Cyber-deception, which strategically deploys decoys (honeypots) to mislead attackers, has emerged as a proactive defense paradigm. However, existing game-theoretic deception models suffer from three structural limitations: they treat each IoT layer in isolation, they assume zero-sum interaction, and they ignore the resilience requirements that govern critical IoT services. We introduce the Multi-Domain Cyber-Deception Bayesian Game (MD-CDBG), a stochastic Bayesian game with incomplete information that couples the three IoT layers under formal resilience and semantic-coherence constraints. We define a novel solution concept, the Resilience-constrained Bayesian Nash Equilibrium (R-BNE), which embeds a hard resilience constraint at every time step, and we solve it with a Bayesian Double Oracle algorithm whose inner oracle uses a Linear Complementarity Program (LCP) to correctly handle the non-zero-sum structure of the game. Through a Lagrangian decomposition, the three coupled sub-games are solved in parallel while enforcing the budget and coherence constraints, and a fallback isolation policy is triggered whenever resilience approaches its threshold. Extensive simulations against three baselines, including a recent state-of-the-art double-oracle deception approach, across three attacker profiles (opportunistic, persistent, and state-sponsored) show that the MD-CDBG reduces resilience-constraint violations by 65–72% while inducing rational attacker behavior, demonstrating that modeling deception, belief, and resilience jointly is essential for protecting multi-layer IoT systems.**

***Index Terms*—IoT security, cyber-deception, Bayesian game, Nash equilibrium, resilience, double-oracle, linear complementarity, honeypot, advanced persistent threat.**

## I. Introduction

The proliferation of Internet of Things (IoT) devices in industrial, medical, and urban infrastructures has dramatically expanded the attack surface available to sophisticated adversaries. Billions of resource-constrained sensors, gateways, and actuators now mediate critical services, and their heterogeneity and scale make conventional perimeter defenses insufficient [1]. Among the threats they face, Advanced Persistent Threats (APTs) are particularly dangerous: they operate methodically over long periods, escalating privileges layer by layer through the canonical three-tier IoT architecture, perception (sensors), network (gateways and routers), and application (cloud services and decision engines) [2].

Traditional intrusion-detection systems are inherently reactive: they detect attacks only after they manifest, by which point a persistent adversary may have already compromised critical assets. Cyber-deception offers a proactive alternative. By deploying decoys (honeypots) that mimic real assets, a defender can mislead the attacker, waste its resources, and reveal its intentions before genuine damage occurs [3], [4]. The core strength of cyber-deception is its proactive nature, transforming static defenses into dynamic, unpredictable, and more resilient security stances [5]. To be effective, however, deception must be optimized through rigorous decision-making under uncertainty, for which game theory provides a principled foundation [6].

Despite a rich literature, existing game-theoretic deception models exhibit three structural gaps when applied to IoT. First, they typically model a single domain or layer, ignoring the cross-layer coupling that characterizes APT progression in IoT; a honeypot strategy optimal for the network layer may be incoherent with the perception or application layer, making decoys easy to unmask. Second, many adopt a zero-sum formulation in which the attacker's gain is exactly the defender's loss; this assumption fails for IoT, where energy consumption, deployment cost, and resilience introduce defender objectives that are not mirrored by the attacker. Third, and most critically, prior work largely ignores resilience: the capacity of the

979-8-3195-0167-7/26/$31.00 ©2026 IEEE

system to maintain service continuity under attack, which is the paramount concern for critical IoT deployments [7].

Contributions. This paper addresses these gaps with the following contributions:

- We formulate the Multi-Domain Cyber-Deception Bayesian Game (MD-CDBG), a stochastic Bayesian game coupling the three IoT layers under formal resilience, semantic-coherence, and budget constraints (Section III).
- We define the Resilience-constrained Bayesian Nash Equilibrium (R-BNE), a solution concept that embeds a hard resilience constraint $\Omega(t) \geq \Omega_{\min}$ at every time step (Section V).
- We design R-BNE-DO, a Bayesian Double Oracle algorithm whose inner oracle solves the non-zero-sum game via a Linear Complementarity Program, coordinated by a Lagrangian decomposition over the three layers, and we prove its convergence (Section VI).
- We evaluate the MD-CDBG against three baselines, including the recent GameSec 2025 double-oracle deception model [5], across three attacker profiles, demonstrating a 65–72% reduction in resilience-constraint violations across all scenarios and substantially more rational attacker behavior, at a defender utility that matches or exceeds the strongest baseline (Section VII).

The remainder of the paper is organized as follows. Section II reviews related work and positions our contribution. Section III formalizes the MD-CDBG. Section IV defines the utility functions. Section V introduces the resilience function and the R-BNE. Section VI presents the R-BNE-DO algorithm and its convergence analysis. Section VII reports the experimental evaluation. Section VIII concludes.

## II. Related Work

### A. Game-Theoretic Cyber-Deception

Signaling games and Stackelberg games have been widely used to model the strategic interaction between a defender deploying decoys and an attacker probing a network [8]. Pawlick et al. [1] provide a comprehensive taxonomy of defensive deception, categorizing approaches by information structure and game type, and book-length treatments establish the scientific foundations of cyber-deception and of game theory and machine learning for security [9], [10]. Huang and Zhu [6] model adaptive honeypot engagement as a semi-Markov decision process solved by reinforcement learning. Two-layer signaling games have been proposed to defend cyber-physical systems against attacks that traverse both a cyber and a physical plane [11]. These models capture information asymmetry but generally treat a single domain and assume a static type for the adversary, leaving cross-layer coupling and resilience unaddressed.

### B. Double Oracle Methods for Large Games

The Double Oracle (DO) meta-algorithm [12] computes equilibria in large or infinite action spaces by iteratively expanding restricted strategy sets, invoking a best-response oracle for each player until no profitable deviation exists. DO and its variants have proven effective for honeypot-allocation and attack-graph games [13]–[15]. Most relevant to our work, Asghar et al. [5] model a dual-domain attack on cyber-physical sensors as a non-zero-sum game solved by double-oracle with a greedy, submodular defender oracle, extending their earlier zero-sum multi-domain deception formulation [16]. Crucially, although their game is not zero-sum, they prove it is strategically equivalent to a zero-sum game (in the sense of Moulin and Vial [17]), which lets them solve each restricted game with a single linear program and avoid the linear complementarity problem (LCP) altogether. This equivalence, however, is a property of their particular payoff structure: it does not survive once the defender's objective carries resilience, energy-budget, and cross-layer coherence penalties that the attacker does not share, as in resilient IoT. Their model also provides no resilience guarantee. We show in Section III that under these penalties the game is not reducible to zero-sum, so the LCP becomes necessary rather than optional.

### C. Bayesian Games and Belief Modeling

Games with incomplete information, introduced by Harsanyi [18], model uncertainty about an opponent's type through a prior belief updated by Bayes' rule. In cyber-security, the attacker's type (e.g., its sophistication or objective) is rarely known, motivating Bayesian formulations. Wan et al. [2] extend this idea with hypergame theory, allowing players to hold subjective and possibly inconsistent views of the same game, and apply it to defensive deception against APTs. However, these approaches maintain a belief over the attacker's type without coupling that belief to hard operational constraints such as resilience or a global deployment budget.

### D. Deception in IoT and Wireless Sensor Networks

A growing body of work targets deception in resource-constrained IoT and wireless sensor networks (WSNs). Anwar et al. [3] develop a game-theoretic framework for dynamic cyber-deception in the Internet of Battlefield Things. Sihomnou et al. [19], [20] study Bayesian and signaling games to defend IoT devices against energy-depletion and battery-drain denial-of-service attacks, emphasizing the lightweight constraints intrinsic to IoT. He et al. [7] propose a deep reinforcement learning approach to time-delay differential game deception for real-time resource deployment. These works recognize IoT-specific constraints but consider a single layer and do not formalize resilience as an equilibrium constraint.

### E. Learning-Based and Detection Approaches

Complementary to game theory, machine-learning detectors have been applied to multi-domain intrusion detection, including deep autoencoder models [21] and graph neural networks that deceive network reconnaissance [22]. Sayed et al. [4] combine game theory with learning to defend against zero-day attacks. Such detectors learn attacker behavior reactively but lack the forward-looking equilibrium reasoning and resilience guarantees that a game-theoretic formulation provides; integrating such a learned detector as an additional oracle within our framework is a promising direction (Section VIII).

TABLE I
POSITIONING OF THE MD-CDBG AGAINST REPRESENTATIVE PRIOR WORK.

| Work | Multi-layer | Non-zero-sum | Bayesian belief | Resil. constr. |
|---|---|---|---|---|
| Pawlick [1] | ✗ | ✓ | ✓ | ✗ |
| Huang [6] | ✗ | ✓ | ✗ | ✗ |
| Wan [2] | ✗ | ✓ | ✓ | ✗ |
| Anwar [3] | ✗ | ✓ | ✗ | ✗ |
| Asghar [5] | ✓ | ✗ | ✗ | ✗ |
| Nguemkam [13] | ✗ | ✓ | ✓ | ✗ |
| **MD-CDBG (ours)** | ✓ | ✓ | ✓ | ✓ |

The "non-zero-sum" column indicates whether the equilibrium is computed on the genuine non-zero-sum game. Asghar [5] formulate a non-zero-sum game but reduce it to a strategically equivalent zero-sum game; MD-CDBG instead solves the non-zero-sum game directly via the LCP.

### F. Positioning

To our knowledge, the MD-CDBG is the first model to (i) couple all three IoT layers in a single Bayesian deception game, (ii) preserve the non-zero-sum structure through an LCP-based oracle rather than a zero-sum proxy, and (iii) embed resilience as a hard constraint within the equilibrium concept itself. Table I summarizes how the MD-CDBG differs from representative prior work.

## III. System Model: MD-CDBG

### A. Game Definition

We model the defender–attacker interaction across the three IoT layers as a multi-domain stochastic Bayesian game:

$$\mathcal{G} = \langle \mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3, \mathcal{C}, \Omega \rangle, \tag{1}$$

where $\mathcal{G}_k$ is the sub-game at layer $k \in \{1,2,3\}$ (perception, network, application), $\mathcal{C} = \{C_1, C_2, C_3\}$ is the set of coupling constraints linking the sub-games, and $\Omega$ is the system resilience function.

Each sub-game is a tuple

$$\mathcal{G}_k = \langle \mathcal{N}_k, \hat{\mathcal{N}}_k, \mathcal{A}_k^D, \mathcal{A}_k^A, \Theta_A, u_k^D, u_k^A \rangle, \tag{2}$$

comprising $|\mathcal{N}_k|$ real nodes and $|\hat{\mathcal{N}}_k|$ decoy nodes, defender and attacker action sets $\mathcal{A}_k^D, \mathcal{A}_k^A$, a set of attacker types $\Theta_A = \{\theta_{\text{opp}}, \theta_{\text{pers}}, \theta_{\text{state}}\}$ corresponding to opportunistic, persistent, and state-sponsored adversaries, and layer-level utilities $u_k^D, u_k^A$. The defender chooses a mixed strategy $\sigma_k^D \in \Delta(\mathcal{A}_k^D)$; the attacker of type $\theta_A$ chooses $\sigma_k^A(\cdot|\theta_A) \in \Delta(\mathcal{A}_k^A)$. Figure 1 illustrates the overall architecture: the attacker progresses upward through the three layers (constraint $C_1$); decoys at adjacent layers must remain coherent ($C_2$); and the defender maintains a per-layer Bayesian belief $\mu_k^t(\theta_A)$ over the attacker type, all under a global budget ($C_3$) and a resilience constraint $\Omega(t) \geq \Omega_{\min}$.

### B. Coupling Constraints

The three sub-games are linked by three constraints that make the MD-CDBG a genuinely multi-domain game rather than three independent games:

- ($C_1$) APT progression: layer $k$ is accessible only if at least one node in layer $k-1$ is compromised. This enforces the sequential escalation characteristic of APTs: an attacker must breach the perception layer before the network layer, and the network before the application.
- ($C_2$) Semantic coherence: decoys at adjacent layers must be mutually consistent, $\text{Coh}(\hat{\mathcal{N}}_k, \hat{\mathcal{N}}_{k+1}) \geq \delta_{\min}$. An attacker that observes a network decoy inconsistent with the perception decoys it has already seen can unmask the deception.
- ($C_3$) Global budget: the total deployment cost across all layers must not exceed a budget $B$, i.e., $\sum_k f_4^k \leq B$, reflecting the finite operational resources of an IoT operator.

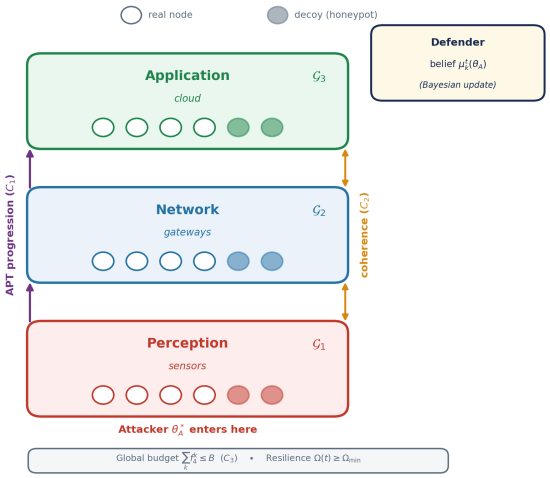



Fig. 1. Three-layer architecture of the MD-CDBG.

### C. State and Transition

The system state $\mathbf{s}(t) = (\mathbf{s}_1(t), \mathbf{s}_2(t), \mathbf{s}_3(t))$ records, for each layer, the set of compromised nodes. The compromise ratio of layer $k$ is $\delta_k(t) = |\text{compromised}_k|/|\mathcal{N}_k|$. At each step, the realized defender and attacker actions induce a stochastic transition $\mathbf{s}(t+1) = \mathcal{T}(\mathbf{s}(t), a^A(t), a^D(t))$: an attacker action targeting a real node succeeds with a technical success probability $p_{k,i}^A$ unless it is trapped by a decoy (probability $f_1^k$), while a defender isolation action can recover a compromised node.

### D. Bayesian Belief

Since the attacker's true type $\theta_A^*$ is unobservable, the defender maintains a belief $\mu_k^t(\theta_A) \in \Delta(\Theta_A)$ per layer $k$ and time $t$, initialized at the prior $\mu_k^0(\theta_A) = p(\theta_A)$ and updated after each observed action $a_k^A(t)$ via Bayes' rule:

$$\mu_k^{t+1}(\theta_A) = \frac{\sigma_k^A(a_k^A(t) \mid \theta_A)\, \mu_k^t(\theta_A)}{\sum_{\theta' \in \Theta_A} \sigma_k^A(a_k^A(t) \mid \theta')\, \mu_k^t(\theta')}. \tag{3}$$

**Lemma 1** (Belief convergence)**.** *Suppose the type-conditional strategies $\{\sigma_k^A(\cdot \mid \theta_A)\}_{\theta_A \in \Theta_A}$ are pairwise distinct (identifiability) and layer $k$ is visited infinitely often. Then, under the true type $\theta_A^*$, the belief $\mu_k^t(\theta_A^*) \to 1$ almost surely as $t \to \infty$; that is, the belief concentrates on the true type. The result follows because, for every $\theta \neq \theta_A^*$, the likelihood-ratio process $L_t(\theta) = \mu_k^t(\theta)/\mu_k^t(\theta_A^*)$ is a non-negative martingale under $\theta_A^*$*

*and hence converges almost surely; the strict positivity of the Kullback–Leibler divergence between distinct type-conditional distributions forces its limit to* 0*, so all false types vanish and* $\mu_k^t(\theta_A^*) \to 1$.

## IV. Utility Functions

### A. Decoy Quality and Detection

The realism of a layer-$k$ decoy deployment is captured by a composite quality $q_k(\sigma_k^D) \in [0,1]$ combining functional, statistical, and temporal realism:

$$q_k(\sigma_k^D) = \omega_1 q_k^{\text{func}} + \omega_2 q_k^{\text{stat}} + \omega_3 q_k^{\text{temp}}, \qquad \textstyle\sum_i \omega_i = 1, \tag{4}$$

where the functional realism $q_k^{\text{func}}(\sigma_k^D) = \sum_i \sigma_k^D(a_{k,i}^D)\, r_{k,i}^{\text{func}}$ is the strategy-weighted realism of the chosen decoy actions ($r_{k,i}^{\text{func}} \in [0,1]$ being the intrinsic realism of action $a_{k,i}^D$), and the statistical and temporal components measure how closely decoy traffic matches that of real nodes.

The attacker's detection capability combines an active and a passive component through a complementary-probability law:

$$d_k(\sigma_k^A, \sigma_k^D) = 1 - \big(1 - d_k^{\text{eff}}\big)\big(1 - d_k^{\text{vuln}}\big), \tag{5}$$

where the active component scales the attacker's probing investment by its type-specific efficiency,

$$d_k^{\text{eff}}(\sigma_k^A) = \sigma_k^A(a_k^{\text{probe}}) \cdot \psi_k(\theta_A), \tag{6}$$

and the passive component reflects the intrinsic detectability of imperfect decoys, $d_k^{\text{vuln}}(\sigma_k^D) = \kappa\,(1 - q_k)$ with $\kappa \in [0,1]$. Equation (5) expresses that a decoy escapes detection only if it survives both the active probe and the passive inspection, hence the product of the two survival probabilities.

### B. Deception Effectiveness

The deception effectiveness at layer $k$ multiplies the decoy ratio, the decoy quality, and the complement of the detection capability:

$$f_1^k(\sigma_k^D, \sigma_k^A) = \frac{|\hat{\mathcal{N}}_k|}{|\mathcal{N}_k| + |\hat{\mathcal{N}}_k|} \cdot q_k(\sigma_k^D) \cdot \big(1 - d_k(\sigma_k^A, \sigma_k^D)\big). \tag{7}$$

The three factors admit a clear reading: more decoys raise the probability that the attacker targets one; higher quality reduces the chance the attacker recognizes it; and lower detection capability increases the chance the decoy goes unnoticed.

### C. Semantic Coherence

The coherence constraint $C_2$ penalizes abrupt quality discontinuities between adjacent layers, which an attacker can exploit to unmask the deception. We measure coherence by the normalized similarity of the decoy quality at two adjacent layers:

$$\text{Coh}_k(\sigma_k^D, \sigma_{k+1}^D) = 1 - \frac{\big|\, q_k(\sigma_k^D) - q_{k+1}(\sigma_{k+1}^D) \,\big|}{\max\big(q_k,\, q_{k+1},\, \epsilon\big)}, \tag{8}$$

where $\epsilon > 0$ avoids division by zero. $\text{Coh}_k \in [0,1]$ equals 1 when the two layers deploy decoys of identical quality and decreases as their realism diverges; constraint $C_2$ requires $\text{Coh}_k \geq \delta_{\min}$ for all adjacent pairs.

### D. Defender Utility

**Definition 1** (Defender utility)**.** *The defender's utility at layer* $k$ *is a weighted linear combination of four partially conflicting objectives:*

$$u_k^D\big(\sigma_k^D, \sigma_k^A(\cdot|\theta_A)\big) = \underbrace{\alpha_1 f_1^k}_{\textit{deception}} + \underbrace{\alpha_2 \Omega_k(\sigma_k^D)}_{\textit{resilience}} - \underbrace{\alpha_3 f_3^k(\sigma_k^D)}_{\textit{energy}} - \underbrace{\alpha_4 f_4^k(\sigma_k^D)}_{\textit{cost}}, \tag{9}$$

*with* $\sum_i \alpha_i = 1$, $\alpha_i \geq 0$. *The energy overhead and operational deployment cost are the strategy-weighted sums of the per-action unit costs:*

$$f_3^k(\sigma_k^D) = \sum_i \sigma_k^D(a_{k,i}^D)\, e_{k,i}^D, \quad f_4^k(\sigma_k^D) = \sum_i \sigma_k^D(a_{k,i}^D)\, c_{k,i}^D, \tag{10}$$

*where* $e_{k,i}^D$ *and* $c_{k,i}^D$ *are the unit energy and deployment costs of defender action* $a_{k,i}^D$ *(both normalized to* $[0,1]$*).*

Because the true type is unobservable, the defender evaluates $u_k^D$ in expectation over its current belief, giving the belief-weighted local utility

$$\bar{u}_k^D\big(\sigma_k^D(\mu_k^t), \mu_k^t\big) = \mathbb{E}_{\theta_A \sim \mu_k^t}\big[u_k^D\big(\sigma_k^D, \sigma_k^A(\cdot|\theta_A)\big)\big], \tag{11}$$

and the global defender utility aggregates over layers:

$$U^D(\sigma^D(\mu^t), \mu^t) = \sum_{k=1}^{3} \beta_k\, \bar{u}_k^D\big(\sigma_k^D(\mu_k^t), \mu_k^t\big), \tag{12}$$

with $\sum_k \beta_k = 1$, $\beta_k > 0$ reflecting the relative criticality of each layer.

### E. Attacker Utility

The utility of attacker type $\theta_A$ at layer $k$ balances expected asset gain $g_k^A$, resource cost $c_k^A$, and detection risk $\rho_k^A$:

$$u_k^A(\sigma_k^A, \sigma_k^D \mid \theta_A) = \gamma_1 g_k^A - \gamma_2 c_k^A - \gamma_3 \rho_k^A, \tag{13}$$

with $\sum_i \gamma_i = 1$. The three components are

$$g_k^A = \sum_i \sigma_k^A(a_{k,i}^A)\, v_{k,i}^A\, (1 - f_1^k)\, p_{k,i}^A, \tag{14}$$

$$c_k^A = \sum_i \sigma_k^A(a_{k,i}^A)\, r_{k,i}^A, \tag{15}$$

$$\rho_k^A = \sum_i \sigma_k^A(a_{k,i}^A)\, f_1^k\, \lambda_{k,i}, \tag{16}$$

where $v_{k,i}^A$ is the value of the asset targeted by action $a_{k,i}^A$, $p_{k,i}^A$ its technical success probability, $r_{k,i}^A$ its resource cost, and $\lambda_{k,i}$ its detection sensitivity. Crucially, the deception term $f_1^k$ enters both $g_k^A$ (as a factor $1 - f_1^k$, reducing gain when the attacker hits a decoy) and $\rho_k^A$ (increasing the risk of being trapped). This dual effect forces the attacker to balance aggression against caution, producing genuinely mixed strategies at equilibrium. Figure 2 summarizes this per-turn interaction as a five-step loop: (1) the defender holds a belief $\mu_k^t(\theta_A)$ while the attacker has a true type $\theta_A^*$; (2) both players act simultaneously on $f_1^k$ (the defender pushing it up with better decoys, the attacker pushing it down through detection); (3) because $f_1^k$ both reduces the attacker's gain and raises its risk, neither player can play a pure strategy, (4) yielding a mixed

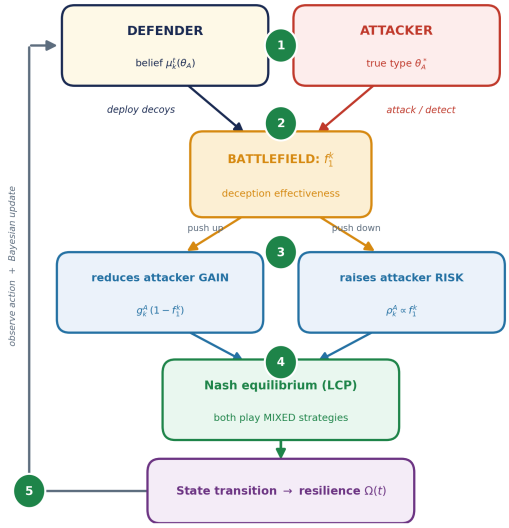



Fig. 2. Per-turn interaction between the two players (five-step loop).

Nash equilibrium computed by the LCP; the resulting state transition updates the resilience $\Omega(t)$; and (5) the observed attacker action feeds the Bayesian belief update for the next turn. The global attacker utility aggregates over layers subject to the $C_1$ accessibility indicator:

$$U^A(\sigma^A, \sigma^D \mid \theta_A) = \sum_{k=1}^{3} \beta_k^A \, \mathbb{E}_{\mathbf{s}}\big[u_k^A \cdot \mathbf{1}[\mathcal{A}_k^A(\mathbf{s}) \neq \emptyset]\big] \,. \quad (17)$$

**Remark 1** (Structural asymmetry)**.** *The defender's global utility $U^D$ is an expectation over attacker types (via the belief $\mu_k^t$), whereas the attacker's utility $U^A$ is an expectation over system states (which nodes are decoys). This double incomplete information is the precise source of the game's non-zero-sum structure: the resilience, energy-budget, and coherence terms in $U^D$ have no counterpart in $U^A$, so $u_k^D \neq -u_k^A$ in general. For the dual-domain game of [5] the defender and attacker payoffs differ only by a column-constant term, so the game is strategically equivalent to a zero-sum game and is solvable by a linear program. Here that property fails: the constraint-driven penalties in $U^D$ are not column-constant, the bimatrix $M_k^D + M_k^A$ is not rank-one, and no zero-sum reduction exists. The restricted game must therefore be solved as a genuine non-zero-sum bimatrix game via the LCP (Section VI).*

## V. Resilience and the R-BNE

### A. Resilience Function

Following the resistance–absorption–recovery decomposition common in cyber-physical resilience analysis [23]–[25], the local resilience at layer $k$ is a weighted sum of three components:

$$\Omega_k(t) = w_1 \cdot \mathrm{res}_k(t) + w_2 \cdot \mathrm{abs}_k(t) + w_3 \cdot \mathrm{rec}_k(t), \quad (18)$$

defined as

$$\mathrm{res}_k(t) = \mathbf{1}[\,1 - \delta_k(t) \geq \rho_k\,], \quad (19)$$

$$\mathrm{abs}_k(t) = \max(0,\ 1 - \delta_k(t)), \quad (20)$$

$$\mathrm{rec}_k(t) = \mathbf{1}[\,\delta_k(t)\,|\mathcal{N}_k|\,\tau_k \leq T_k^{\max}\,], \quad (21)$$

where $\rho_k$ is the resistance threshold, $\tau_k$ the per-node recovery time, and $T_k^{\max}$ the maximum acceptable recovery time. Resistance indicates whether the uncompromised fraction exceeds $\rho_k$; absorption measures the remaining healthy fraction; and recovery indicates whether the compromised nodes can be restored within the acceptable time budget. The weights $w_1, w_2, w_3$ adapt to the global compromise level $\bar{\delta}(t) = \frac{1}{3}\sum_k \delta_k(t)$: resistance dominates when the system is healthy, absorption when it is partially compromised, and recovery when damage is extensive. The global resilience is the weighted aggregation

$$\Omega(t) = \sum_{k=1}^{3} \beta_k^{\Omega}\, \Omega_k(t), \qquad \sum_k \beta_k^{\Omega} = 1. \quad (22)$$

### B. Solution Concept

**Definition 2** (Admissible policy set)**.** *The set $\Pi_{\Omega}^D$ of admissible defender policies comprises all belief-dependent behavioral policies $\sigma^D : \Delta(\Theta_A) \to \Delta(\mathcal{A}^D)$ that satisfy the resilience constraint $\Omega(t) \geq \Omega_{\min}(t)$ for all $t$ and the coupling constraints $\mathcal{C}$.*

**Definition 3** (R-BNE)**.** *A Resilience-constrained Bayesian Nash Equilibrium is a strategy profile $(\sigma^{D*}(\cdot), \sigma^{A*}(\cdot|\theta_A^*))$ satisfying, for all belief states $\mu^t \in \Delta(\Theta_A)$:*

*(i)* $$\sigma^{D*}(\mu^t) = \arg\max_{\sigma^D \in \Pi_{\Omega}^D} \mathbb{E}_{\theta_A \sim \mu^t}\big[U^D(\sigma^D(\mu^t), \sigma^{A*})\big] \,, \quad (23)$$

*(ii)* $$\sigma^{A*}(\cdot|\theta_A^*) = \arg\max_{\sigma^A \in \Delta(\mathcal{A}^A)} U^A(\sigma^{D*}, \sigma^A \mid \theta_A^*), \quad (24)$$

*(iii)* $$\Omega(t) \geq \Omega_{\min}(t) \quad \forall t \geq 0. \quad (25)$$

The R-BNE differs from a standard Bayesian Nash equilibrium [26] in three respects. First, the defender strategy $\sigma^{D*}(\mu^t)$ is a belief-dependent behavioral policy, mapping the current belief to a mixed action rather than being a fixed distribution. Second, the feasible set $\Pi_{\Omega}^D$ imposes a hard resilience constraint at every time step. Third, condition (ii) uses the attacker's true utility $U^A$, not the defender's constraint-penalized objective, ensuring that the attacker's rationality is modeled independently of the defender's internal penalties.

**Proposition 1** (Existence)**.** *If (E1) the admissible set $\Pi_{\Omega}^D$ is non-empty and (E2) the pure action sets are finite, then the MD-CDBG admits at least one R-BNE in mixed strategies.*

*Proof.* Over the finite horizon $T$, the observed action sequences are finitely many, so the set of beliefs reachable through (3) is finite; a pure policy is therefore a map between two finite sets, and there are finitely many of them. Restricting to those satisfying the resilience and coupling constraints leaves $\Pi_{\Omega}^D$, finite and non-empty by (E1). The attacker likewise selects among the finitely many exploitation paths of a finite graph. Mixed strategies thus range over finite-dimensional simplices, compact and convex. Expected utilities are multilinear

in the mixing probabilities, hence continuous in the profile and linear, a fortiori quasi-concave, in each player's own strategy. Nash's theorem for finite games applies, and the constraint $\Omega(t) \geq \Omega_{\min}$ is enforced by restricting the defender's support to $\Pi^D_\Omega$ rather than by penalizing the objective, so the equilibrium obtained is admissible by construction. □

Note that $\mathrm{res}_k$ and $\mathrm{rec}_k$ in (21) are indicators, hence discontinuous in $\delta_k$. This does not affect the argument, which rests on finiteness rather than on regularity; it does however make $\Omega$ non-differentiable, which is why the scheme of Section VI is combinatorial rather than first-order.

## VI. R-BNE-DO Algorithm

### A. Lagrangian Decomposition

The coupling constraints $C_2$ and $C_3$ are dualized via non-negative Lagrange multipliers $\nu_k$ (semantic coherence at layer $k$) and $\lambda$ (global budget). For fixed multipliers, the Lagrangian separates into three independent layer sub-problems:

$$\mathcal{L} = \sum_{k=1}^{3} \mathcal{L}_k(\sigma^D_k, \sigma^A_k, \lambda, \nu_k), \tag{26}$$

where each local Lagrangian augments the belief-weighted defender utility with the budget and coherence penalties:

$$\mathcal{L}_k = \beta_k \, \mathbb{E}_{\theta_A \sim \mu^t_k}\big[u^D_k\big] - \lambda f^k_4 - \nu_k \max(0, \delta_{\min} - \mathrm{Coh}_k)\,. \tag{27}$$

This separation is what enables the three sub-games to be solved in parallel. The dual problem

$$\min_{\lambda, \boldsymbol{\nu} \geq 0} \; \max_{\sigma^D \in \Pi^D_\Omega, \sigma^A} \; \mathcal{L}$$

is solved by dual gradient ascent: the multipliers are updated in the direction of their respective constraint violations,

$$\lambda^{t+1} = \Big[\lambda^t + \eta_t \Big( \textstyle\sum_k f^k_4 - B \Big)\Big]^+, \tag{28}$$

$$\nu^{t+1}_k = \big[\nu^t_k + \eta_t \big(\delta_{\min} - \mathrm{Coh}_k\big)\big]^+, \tag{29}$$

with step $\eta_t = \eta_0/\sqrt{t}$ satisfying the Robbins–Monro conditions [27] $\sum_t \eta_t = \infty$ and $\sum_t \eta_t^2 < \infty$. The multipliers act as adaptive shadow prices: $\lambda$ rises when the budget is exceeded, $\nu_k$ rises when inter-layer coherence falls below $\delta_{\min}$.

### B. Bayesian Double Oracle

The inner oracle (Algorithm 1) computes the local Bayesian Nash equilibrium of sub-game $\mathcal{G}_k$ for fixed multipliers. Because the MD-CDBG is non-zero-sum (Remark 1), the restricted-game equilibrium cannot be computed by a single minimax linear program; instead, we use a Linear Complementarity Program [28] with two distinct payoff matrices. Let $e_i$ denote the pure defender action $a^D_{k,i}$ and $e_j$ the pure attacker action $a^A_{k,j}$. The defender matrix $M^D_k$ encodes the belief-weighted Lagrangian payoff, and the attacker matrix $M^A_k$ encodes the attacker's true utility:

$$M^D_k[i,j] = \mathbb{E}_{\theta_A \sim \mu^t_k}\big[u^D_k(e_i, e_j \mid \theta_A)\big] - \lambda f^k_4(e_i) - \nu_k \max\big(0,\, \delta_{\min} - \mathrm{Coh}_k(e_i, \sigma^D_{k+1})\big), \tag{30}$$

$$M^A_k[i,j] = u^A_k(e_i, e_j \mid \theta^*_A). \tag{31}$$

**Algorithm 1** Bayes-DO: Bayesian Double Oracle for $\mathcal{G}_k$

**Require:** $\mathcal{G}_k$, multipliers $(\lambda, \nu_k)$, belief $\mu^t_k(\theta_A)$
**Ensure:** Local BNE $(\sigma^{D*}_k, \sigma^{A*}_k)$

```
Initialize R_D ← {a^D_{k,1}}, R_A ← {a^A_{k,1}}
repeat
    (σ̃^D_k, σ̃^A_k) ← NE(M^D_k|_{R_D×R_A}, M^A_k|_{R_D×R_A}) via LCP
    a^{D*}_k ← arg max_a E_{θ_A∼μ^t_k}[L_k(a, σ̃^A_k)] {ties: see below}
    a^{A*}_k ← arg max_a u^A_k(σ̃^D_k, a | θ*_A) {idem}
    if a^{D*}_k ∈ R_D and a^{A*}_k ∈ R_A then
        return (σ̃^D_k, σ̃^A_k) {exact equilibrium found}
    else
        R_D ← R_D ∪ {a^{D*}_k};   R_A ← R_A ∪ {a^{A*}_k}
    end if
until convergence
```

Two points distinguish them. First, $M^D_k$ is an expectation over the belief $\mu^t_k$ (the defender does not know the type), whereas $M^A_k$ uses the attacker's true type $\theta^*_A$. Second, the budget and coherence penalties $(\lambda, \nu_k)$ appear only in $M^D_k$: they are the defender's internal constraints and do not affect the attacker's genuine incentives. Consequently $M^D_k[i,j] + M^A_k[i,j] \neq 0$ in general, which is precisely why a non-zero-sum LCP is required rather than a single minimax LP. The defender's best response (Step 2) is computed in expectation over the belief, whereas the attacker's best response (Step 3) uses its true utility, never the Lagrangian.

**Remark 2** (Exact termination)**.** *Bayes-DO is a combinatorial algorithm with exact termination: the expansion check at Step 6 returns true precisely when no profitable deviation exists outside the restricted sets, which is the definition of a Nash equilibrium of the full game. No approximation tolerance is required, since equilibria are typically supported on a small subset of actions.*

Tie-breaking. Both oracles maximize over a finite set, and ties are frequent rather than pathological, symmetric nodes within a layer often yielding identical marginal utility. Left unresolved they would prevent termination, since the expansion check compares actions rather than values and would keep inserting equivalent alternatives. We therefore return a maximizer already in the restricted set when one exists, which alone guarantees termination, then prefer minimum deployment cost, then the canonical node ordering. The rule is neutral in value, and varying it changes neither the equilibrium value nor the reported metrics.

### C. Main Algorithm

The outer algorithm R-BNE-DO (Algorithm 2) operates over a horizon $T$ in four phases per step. Phase A propagates the $C_1$ constraint, checks resilience feasibility, and calls Bayes-DO per layer in parallel; if the global resilience would fall below $\Omega_{\min}$, a fallback policy prioritizing node isolation is activated to restore resilience. Phase B samples actions and transitions the state. Phase C updates the per-layer belief. Phase D updates the Lagrange multipliers.

**Algorithm 2** R-BNE-DO: R-BNE via Bayesian Double Oracle

**Require:** MD-CDBG, $B$, $\Omega_{\min}$, $\delta_{\min}$, $\eta_0$, $\varepsilon$, $T$
Initialize $\lambda^0 \leftarrow 0$, $\nu_k^0 \leftarrow 0$, $\mu_k^0(\theta_A) \leftarrow p(\theta_A)$
**for** $t = 0, \ldots, T-1$ **do**
  **[A]** Propagate $C_1$; compute $\Omega(t)$
  **for** $k = 1, 2, 3$ **in parallel do**
    **if** $\Omega(t) < \Omega_{\min}$ **then**
      Activate fallback (isolation) policy
    **else**
      $(\sigma_k^{D,t}, \sigma_k^{A,t}) \leftarrow$ Bayes-DO$(\mathcal{G}_k, \lambda^t, \nu_k^t, \mu_k^t)$
    **end if**
  **end for**
  **[B]** Sample $a^A(t), a^D(t)$; transition $\mathbf{s}(t{+}1)$
  **[C]** Update belief $\mu_k^{t+1}$ via (3)
  **[D]** Update $\lambda^{t+1}, \nu_k^{t+1}$ by dual ascent
  **if** $\max(\|\Delta\sigma^D\|, |\Delta\lambda|, \max_k |\Delta\nu_k|) < \varepsilon$ **then**
    **return** $(\sigma^{D,t+1}, \sigma^{A,t+1})$
  **end if**
**end for**

### D. Convergence Analysis

**Theorem 1** (Convergence)**.** *Under the conditions of Proposition 1, with Robbins–Monro steps and assuming (A1) uniqueness of the restricted Nash equilibrium at each Bayes-DO iteration (or a fixed selection rule), and (A2) local concavity of $\mathcal{L}_k$ in $\sigma_k^D$, the R-BNE-DO algorithm converges to an R-BNE.*

*Proof sketch.* Convergence combines three arguments operating at the three nested levels. (i) Inner level: for fixed multipliers, Bayes-DO terminates in finitely many iterations at the exact restricted Nash equilibrium, computed by the LCP [28]; assumption (A1) prevents cycling. (ii) Outer level: dual gradient ascent with diminishing steps converges to the saddle point of the Lagrangian (26) under Slater's strong-duality condition [29], which holds because $\Pi_\Omega^D \neq \emptyset$; at convergence the complementary slackness conditions hold, $\lambda^* > 0$ only if $\sum_k f_4^k = B$, and $\nu_k^* > 0$ only if $\mathrm{Coh}_k = \delta_{\min}$. (iii) Belief level: by Lemma 1, $\mu_k^t(\theta_A^*) \to 1$ almost surely, so the defender's best response converges to the one under perfect type knowledge, tightening the bound on the iterations needed. □

**Proposition 2** (Per-call complexity)**.** *Let $r_D = |R_D|$ and $r_A = |R_A|$ be the restricted-set sizes at termination. The LCP solved by support enumeration [30] costs $\mathcal{O}(2^{r_D} 2^{r_A} (r_D + r_A)^3)$. In the MD-CDBG, $r_D \leq 5$ and $r_A \leq 6$, so each call costs at most $\approx 2.7 \times 10^6$ elementary operations, and the number of outer iterations to reach precision $\varepsilon$ is $\mathcal{O}(1/\varepsilon^2)$.*

Measured runtime. Over ten seeded replications on a single core, R-BNE-DO completes a $T = 50$ campaign in $930 \pm 30$ ms against $8348 \pm 111$ ms for Asghar-2025, a factor of nine; the naive baselines run in under 700 ms since they solve no game. The gap is structural: the greedy oracle of [5] re-evaluates every candidate at each expansion, whereas the LCP operates on restricted games of at most $r_D \leq 5$, $r_A \leq 6$ actions.

TABLE II
PROVENANCE OF THE PARAMETER VALUES.

| | |
|---|---|
| **Calibrated** | $\psi_k(\theta_A)$ from APT detection rates [2]; decoy ratios |
| **Swept** | $\Omega_{\min}$, $\delta_{\min}$, $B$ (Section VII-F) |
| **Standard** | $\gamma = 0.95$, $\eta_0 = 0.10$, $T = 50$, 10 runs |
| **Choice** | three-tier topology, $\lvert\mathcal{N}_k\rvert = \{15, 8, 4\}$ |
| **Ordering only** | $\alpha$, $\beta_k^\Omega$, $w_j$: no source measures these |

## VII. EXPERIMENTAL EVALUATION

### A. Implementation

The MD-CDBG and the R-BNE-DO algorithm were implemented from scratch in Python 3 using `NumPy` for the numerical core and `SciPy` for linear programming. The custom simulator faithfully mirrors Algorithms 1 and 2: it maintains a per-layer state and belief, runs the four phases (feasibility check, Bayes-DO, transition, belief and multiplier updates) at each of the $T$ time steps, and records the per-step utilities, resilience, and beliefs. The non-zero-sum local equilibria are computed by a dedicated Linear Complementarity Program solver based on support enumeration [30], which returns the exact mixed Nash equilibrium of each restricted bimatrix game; the zero-sum baseline instead calls a minimax linear program through `SciPy`'s `linprog`. Each experiment runs on a standard CPU in a few minutes; no GPU is required.

### B. Setup

We simulate a three-layer IoT system with $|\mathcal{N}_k| = \{15, 8, 4\}$ real nodes and $|\hat{\mathcal{N}}_k| = \{10, 5, 3\}$ decoys, over a horizon $T = 50$ and 10 independent replications per scenario. The three scenarios correspond to the three attacker types: O (opportunistic, $\psi = 0.20$), P (persistent, $\psi = 0.50$), and E (state-sponsored, $\psi = 0.65$–$0.70$). We set $\Omega_{\min} = 0.55$, $\delta_{\min} = 0.60$, budget $B = 1.2$, discount $\gamma = 0.95$, and step $\eta_0 = 0.10$. Defender utility weights are $\alpha = (0.55, 0.32, 0.07, 0.06)$. Parameter ranges are calibrated from the IoT-security literature: detection efficiencies $\psi_k(\theta_A)$ follow the attacker-sophistication levels reported for APT honeypot detection [2], decoy ratios reflect typical honeypot-deployment densities, and the resilience threshold $\Omega_{\min}$ is set to a service-continuity level consistent with critical-IoT requirements. All results are reported with 95% confidence intervals.

Table II states the provenance of each value, separating what is calibrated on a published source, what is a modeling choice, and what is swept. No source publishes values for an adversary's internal trade-offs, so the utility weights are declared as choices whose ordering alone is anchored, their robustness established by the sweep of Section VII-F.

### C. Baselines

We compare R-BNE-DO+LCP against three baselines spanning the spectrum from naive to state-of-the-art, chosen so that each isolates one design element:

- No-Deception: the defender performs passive monitoring only and deploys no decoys; the attacker plays a rational best response. This is the lower bound that quantifies the value of deception itself.

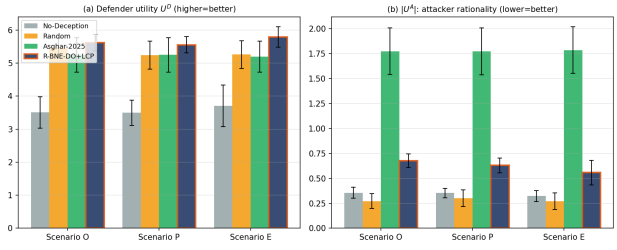



Fig. 3. Defender utility $U^D$ and absolute attacker utility $|U^A|$ across the three scenarios (10 runs, 95% CI). The MD-CDBG induces a far more rational attacker ($|U^A| \approx 0.6$) than the zero-sum proxy of Asghar-2025 ($|U^A| \approx 1.8$).

- Random: decoys are deployed uniformly at random across actions, with no strategic reasoning. Comparing against it isolates the contribution of optimized placement over mere presence of decoys.
- Asghar-2025 [5]: the state-of-the-art GameSec 2025 double-oracle deception model, ported to our IoT setting: a greedy submodular defender oracle with restricted games solved by a linear program under the authors' strategic zero-sum equivalence, with no Bayesian belief and no resilience constraint. It shares the double-oracle backbone with our method, so comparing against it isolates the two genuine differences: the non-zero-sum (LCP) treatment and the resilience constraint.

This design lets each comparison attribute the observed gain to a specific component of the MD-CDBG rather than to the framework as a whole.

### D. Metrics

We report three metrics. The first two are the discounted total utilities accumulated over the horizon, where the discount factor $\gamma \in (0, 1)$ weights early time steps more heavily, reflecting the greater value of preventing early compromise:

$$U^D = \sum_{t=0}^{T-1} \gamma^t U^D(\sigma^{D,t}, \mu^t), \quad U^A = \sum_{t=0}^{T-1} \gamma^t U^A(\sigma^{A,t} \mid \theta_A^*). \tag{32}$$

A higher $U^D$ is better for the defender; for $U^A$, a value closer to zero indicates a more rational, less self-defeating attacker model (a strongly negative $U^A$ means the attacker is forced into moves that harm itself, which signals an unrealistic adversary model rather than a defender success). The third metric is the violation rate

$$\mathrm{VC} = \frac{1}{T} \sum_{t=0}^{T-1} \mathbf{1}[\Omega(t) < \Omega_{\min}], \tag{33}$$

the fraction of time steps during which the resilience constraint is breached. A lower VC means the system maintains service continuity more consistently and is the primary indicator of the defender's true objective. When comparing two methods we report the relative violation reduction $\Delta_{\mathrm{VC}} = (\mathrm{VC}_{\mathrm{base}} - \mathrm{VC}_{\mathrm{ours}})/\,\mathrm{VC}_{\mathrm{base}}$, i.e., the fractional decrease in violation rate of our method with respect to a baseline; all reported metrics are averaged over 10 independent replications.

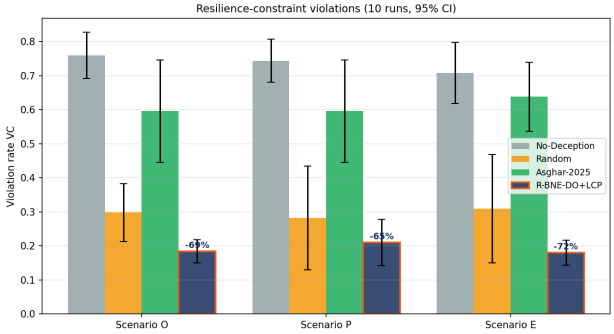



Fig. 4. Resilience-constraint violation rate VC (10 runs, 95% CI). R-BNE-DO+LCP reduces violations by 65–72% relative to Asghar-2025 across all scenarios.

TABLE III
FOUR-METHOD COMPARISON (MEAN OVER 10 RUNS). LOWER VC IS BETTER; LESS NEGATIVE $U^A$ INDICATES A MORE RATIONAL ATTACKER.

| Sc | Method | $U^D$ | $U^A$ | VC |
|---|---|---|---|---|
| O | No-Deception | 3.51 | −0.36 | 0.76 |
| | Random | 5.47 | −0.27 | 0.30 |
| | Asghar-2025 | 5.25 | −1.77 | 0.60 |
| | **R-BNE-DO+LCP** | 5.62 | −0.68 | **0.18** |
| P | No-Deception | 3.50 | −0.35 | 0.74 |
| | Random | 5.24 | −0.30 | 0.28 |
| | Asghar-2025 | 5.25 | −1.77 | 0.60 |
| | **R-BNE-DO+LCP** | **5.56** | −0.63 | **0.21** |
| E | No-Deception | 3.71 | −0.32 | 0.71 |
| | Random | 5.26 | −0.27 | 0.31 |
| | Asghar-2025 | 5.19 | −1.78 | 0.64 |
| | **R-BNE-DO+LCP** | **5.79** | −0.56 | **0.18** |

### E. Results

Table III summarizes the four-method comparison, and Figures 3–5 present the same data graphically.

Resilience (Fig. 4). This is the headline result. R-BNE-DO+LCP achieves the lowest violation rate in every scenario (VC = 0.18–0.21). Applying the relative violation reduction $\Delta_{\mathrm{VC}}$ defined above, this is a 69%, 65%, and 72% decrease relative to Asghar-2025 in scenarios O, P, and E respectively (e.g., $(0.60 - 0.18)/0.60 \approx 0.69$ for O), and up to 76% relative to No-Deception, with the ordering R-BNE-DO < Random < Asghar-2025 < No-Deception. That Random outperforms the more sophisticated Asghar-2025 on resilience is revealing: methods that optimize a non-resilience objective can steer the system into low-resilience states, whereas random placement at least spreads risk.

Defender utility (Fig. 3a). R-BNE-DO+LCP dominates in scenarios P (5.56 vs 5.25) and E (5.79 vs 5.19), and is statistically indistinguishable from the best baseline in O. Even against the weakest attacker (O), where raw utilities are comparable, our method still secures the system far better (VC 0.18 vs 0.60): the defender obtains equal or higher utility and substantially better resilience, without trading one for the other.

Attacker rationality (Fig. 3b). The zero-sum proxy of Asghar-2025 drives the attacker utility to $\approx -1.8$ (an adver-

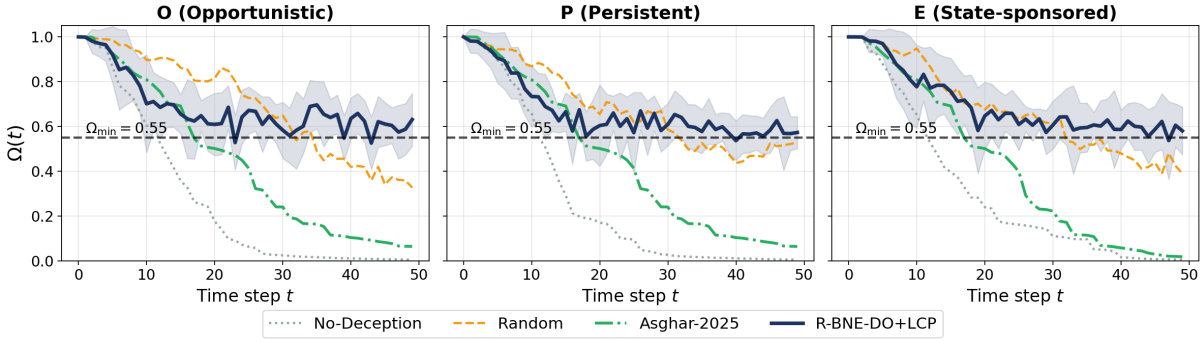



Fig. 5. Global resilience $\Omega(t)$ over time (10 runs, shaded $\pm$std). R-BNE-DO+LCP maintains $\Omega(t)$ above $\Omega_{\min}$ far more consistently than the baselines.

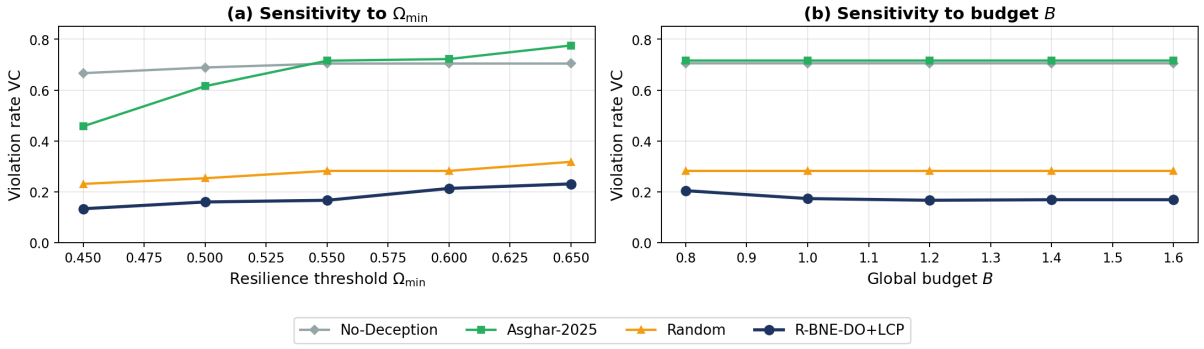



Fig. 6. Sensitivity of the violation rate to (a) the resilience threshold $\Omega_{\min}$ and (b) the global budget $B$. R-BNE-DO+LCP remains the lowest-violation method across the entire range, with a 70–77% reduction relative to Asghar-2025 throughout.

sary forced into self-harming moves), whereas our non-zero-sum treatment yields $U^A \approx -0.6$, a rational attacker playing to its genuine incentives. This gap signals that the zero-sum model misrepresents the threat, which can lead to defenses tuned against an unrealistic adversary.

Temporal behavior (Fig. 5). R-BNE-DO+LCP stays above $\Omega_{\min}$ most of the time and rebounds quickly after compromise spikes, a direct effect of the fallback isolation policy, while the baselines drift below the threshold and never recover.

### F. Sensitivity Analysis

A natural concern is whether the resilience advantage is an artifact of the particular parameter values chosen in Section VII-B. To address this, we vary the two parameters most directly tied to our contribution, the resilience threshold $\Omega_{\min}$ and the global budget $B$, over wide ranges, holding all else fixed, and re-measure the violation rate of every method. Figure 6 reports the results.

Across $\Omega_{\min} \in [0.45, 0.65]$ (Fig. 6a), R-BNE-DO+LCP remains the lowest-violation method at every operating point, and the relative reduction with respect to Asghar-2025 stays in the 70%–77% band throughout. As the threshold tightens, all methods incur more violations, but the gap to the baselines does not close: the constraint-aware fallback keeps R-BNE close to the threshold even when it is demanding. Varying the budget $B \in [0.8, 1.6]$ (Fig. 6b) yields the same picture: the reduction with respect to Asghar-2025 stays in the 71%–77% band, and the ordering R-BNE-DO $<$ Random $<$ Asghar-2025 $<$ No-Deception never changes. Notably, R-BNE already secures most of its benefit at a small budget, indicating that the advantage stems from how decoys are allocated under the equilibrium and resilience constraint, not merely from how many are deployed. These sweeps confirm that the reported advantage is structural and robust to reasonable reparameterization rather than tuned to a single configuration.

### G. Discussion

Two higher-level lessons emerge. First, the comparison design (Section VII) attributes the resilience gain to specific causes: the gap to Asghar-2025 isolates the value of the non-zero-sum (LCP) treatment and the resilience constraint, while the gap to Random isolates the value of equilibrium-based placement over the mere presence of decoys; both gaps are large, so neither component alone suffices. Second, the consistent ordering across three very different attacker profiles indicates the advantage is structural rather than tuned to a particular adversary.

## VIII. Conclusion

We presented the MD-CDBG, a multi-domain cyber-deception Bayesian game that couples the three IoT layers under formal resilience constraints, together with the R-BNE solution concept and the R-BNE-DO algorithm. By correctly modeling the non-zero-sum structure via an LCP oracle and embedding resilience as a hard constraint, the framework

reduces resilience violations by 65–72% relative to the state-of-the-art GameSec 2025 approach while inducing rational attacker behavior.

Several directions extend this work; while the present results are established in simulation, the framework is designed to carry over to deployed settings. (i) Learning-based oracle: the greedy and equilibrium oracles could be complemented by a trained deep detector (e.g., a graph attention network or transfer autoencoder) acting as an additional best-response oracle within the double-oracle loop, combining the forward-looking guarantees of game theory with the pattern-recognition strength of machine learning. (ii) Symmetric information: a symmetric extension would model the attacker's own belief over the defender type, turning the game into a two-sided incomplete-information interaction. (iii) Richer deception techniques: the defender's action repertoire can be extended with additional network-layer mechanisms such as software-defined network reconfiguration and deceptive routing or address-shuffling, each adding new defender actions while preserving the multi-domain coupling. (iv) Scalability: approximate equilibrium solvers would let the model scale to IoT deployments with thousands of heterogeneous nodes. Together, these directions point toward resilient-by-design cyber-deception, where security and service continuity are optimized jointly rather than traded off against each other.

## ACKNOWLEDGMENT

This work was supported by the French Government Scholarship (BGF) and conducted within the cotutelle program between Université Paris-Saclay (Laboratoire DAVID) and University of Ngaoundere.

## REFERENCES

[1] J. Pawlick, E. Colbert, and Q. Zhu, "A game-theoretic taxonomy and survey of defensive deception for cybersecurity and privacy," *ACM Computing Surveys*, vol. 52, no. 4, pp. 1–28, 2019.

[2] Z. Wan, J.-H. Cho, M. Zhu, A. H. Anwar, C. A. Kamhoua, and M. P. Singh, "Foureye: Defensive deception against advanced persistent threats via hypergame theory," *IEEE Transactions on Network and Service Management*, vol. 19, no. 1, pp. 112–129, mar 2022.

[3] A. H. Anwar, C. Kamhoua, and N. Leslie, "A game-theoretic framework for dynamic cyber deception in internet of battlefield things," in *Proc. 16th EAI Int. Conf. Mobile and Ubiquitous Systems (MobiQuitous)*, Houston, TX, USA, 2019.

[4] M. A. Sayed, A. H. Anwar, C. Kiekintveld, B. Bosansky, and C. Kamhoua, "Cyber deception against zero-day attacks: A game theoretic approach," in *Decision and Game Theory for Security (GameSec)*, 2023.

[5] A. B. Asghar, A. Hemida, C. Kamhoua, and J. Kleinberg, "Strategic defense allocation in cyber-physical sensor systems under dual-domain attacks," in *Game Theory and AI for Security (GameSec)*, ser. LNCS, vol. 16224. Cham: Springer, 2025, pp. 81–100.

[6] L. Huang and Q. Zhu, "Adaptive honeypot engagement through reinforcement learning of semi-markov decision processes," in *Decision and Game Theory for Security (GameSec)*, ser. LNCS, vol. 11836. Springer, 2019, pp. 196–216.

[7] W. He, J. Tan, R. Wang, Z. Liu, X. Luo, H. Hu, and H. Zhang, "A deep reinforcement learning approach to time delay differential game deception resource deployment," *IEEE Transactions on Dependable and Secure Computing*, vol. 23, no. 1, pp. 1655–1670, 2026.

[8] M. Tambe, *Security and Game Theory: Algorithms, Deployed Systems, Lessons Learned*. Cambridge, U.K.: Cambridge University Press, 2011.

[9] S. Jajodia, V. S. Subrahmanian, V. Swarup, and C. Wang, *Cyber Deception: Building the Scientific Foundation*. Cham, Switzerland: Springer, 2016.

[10] C. A. Kamhoua, C. D. Kiekintveld, F. Fang, and Q. Zhu, Eds., *Game Theory and Machine Learning for Cyber Security*. Hoboken, NJ, USA: Wiley-IEEE Press, 2021.

[11] P. Chassem Kamdem, A. B. Zemkoho, L. Njilla, M. Nkenlifack, and C. A. Kamhoua, "Two-layer deception model based on signaling games against cyber attacks on cyber-physical systems," *IEEE Access*, vol. 12, pp. 171 559–171 570, 2024.

[12] H. B. McMahan, G. J. Gordon, and A. Blum, "Planning in the presence of cost functions controlled by an adversary," in *Proc. 20th Int. Conf. Machine Learning (ICML)*, Washington, DC, 2003, pp. 536–543.

[13] A. L. Nguemkam, A. H. Anwar, V. K. Tchendji, D. K. Tosh, and C. Kamhoua, "Optimal honeypot allocation using core attack graph in partially observable stochastic games," *IEEE Access*, vol. 12, pp. 187 444–187 455, 2024.

[14] H. Q. Ngo, M. Guo, and H. Nguyen, "Catch me if you can: Effective honeypot placement in dynamic ad attack graphs," in *Proc. IEEE Conf. Computer Communications (INFOCOM)*, 2024, pp. 451–460.

[15] K. Durkota, V. Lisý, B. Bošanský, and C. Kiekintveld, "Optimal network security hardening using attack graph games," in *Proc. 24th Int. Joint Conf. Artificial Intelligence (IJCAI)*, 2015, pp. 526–532.

[16] A. B. Asghar, A. Hemida, and C. Kamhoua, "A scalable double-oracle algorithm for multi-domain deception game," in *Proc. IEEE Conf. Communications and Network Security (CNS)*. IEEE, 2024, pp. 1–9.

[17] H. Moulin and J.-P. Vial, "Strategically zero-sum games: The class of games whose completely mixed equilibria cannot be improved upon," *International Journal of Game Theory*, vol. 7, no. 3–4, pp. 201–221, 1978.

[18] J. C. Harsanyi, "Games with incomplete information played by "bayesian" players, part i. the basic model," *Management Science*, vol. 14, no. 3, pp. 159–182, 1967.

[19] I. C. Kombou Sihomnou, A. Benslimane, A. H. Anwar, G. Deugoue, and C. Kamhoua, "Defending internet of things against energy depletion attack using bayesian game," *IEEE Internet of Things Journal*, 2025, early access.

[20] I. C. Kombou Sihomnou, A. Benslimane, A. H. Anwar, G. Deugoue, and C. A. Kamhoua, "Cyber deception against battery drain dos attacks in wireless sensor networks using signaling game," *IEEE Access*, vol. 13, pp. 5219–5235, 2025.

[21] M. A. Musa, A. Y. Gital, K. I. Musa, H. Chiroma, and E. N. Ramson, "A multi-domain base cyber intrusion detection using deep autoencoder model," *Int. J. Advances in Engineering and Management (IJAEM)*, vol. 5, no. 2, pp. 530–535, 2023.

[22] A. Njanko, C. Kamhoua, A. H. Anwar, and D. B. Rawat, "An unsupervised gnn approach to deceive network reconnaissance attack," in *Proc. IEEE Military Communications Conf. (MILCOM)*, 2023.

[23] D. Henry and J. E. Ramirez-Marquez, "Generic metrics and quantitative approaches for system resilience as a function of time," *Reliability Engineering & System Safety*, vol. 99, pp. 114–122, 2012.

[24] I. Linkov, T. Bridges, F. Creutzig, J. Decker, C. Fox-Lent, W. Kröger, J. H. Lambert, A. Levermann, B. Montreuil, J. Nathwani, R. Nyer, O. Renn, B. Scharte, A. Scheffler, M. Schreurs, and T. Thiel-Clemen, "Changing the resilience paradigm," *Nature Climate Change*, vol. 4, no. 6, pp. 407–409, 2014.

[25] Q. Zhu and T. Başar, "Game-theoretic methods for robustness, security, and resilience of cyberphysical control systems: Games-in-games principle for optimal cross-layer resilient control systems," *IEEE Control Systems Magazine*, vol. 35, no. 1, pp. 46–65, 2015.

[26] J. Nash, "Non-cooperative games," *Annals of Mathematics*, vol. 54, no. 2, pp. 286–295, 1951.

[27] H. Robbins and S. Monro, "A stochastic approximation method," *The Annals of Mathematical Statistics*, vol. 22, no. 3, pp. 400–407, 1951.

[28] C. E. Lemke and J. T. Howson, Jr., "Equilibrium points of bimatrix games," *Journal of the Society for Industrial and Applied Mathematics*, vol. 12, no. 2, pp. 413–423, 1964.

[29] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[30] R. Porter, E. Nudelman, and Y. Shoham, "Simple search methods for finding a nash equilibrium," *Games and Economic Behavior*, vol. 63, no. 2, pp. 642–662, 2008.