Honeypot Allocation for Cyber Deception Under Uncertainty

Ahmed H. Anwar, *Member, IEEE*, Charles A. Kamhoua, *Senior Member, IEEE*, Nandi O. Leslie, and Christopher Kiekintveld, *Senior Member, IEEE*

***Abstract*—Cyber deception aims to misrepresent the state of the network to mislead the attackers, falsify their reconnaissance conclusions, and deflect them away from their goals. Honeypots serve as decoy devices inside networks that can capture adversaries for monitoring purposes. We propose a two-phase deception approach based on honeypot allocation. In the first phase, we develop a proactive deceptive honeypot allocation policy, the second phase proposes a reactive deception approach that dynamically allocates honeypots according to IDS updates. Considering a practical scenario, the defender partially monitors the adversary's activities. To this end, we develop our deception approach using a combination of game-theoretic and reinforcement learning models. We cast the problem of reactive deception as a partially observable Markov decision process (POMDP) based on a game-theoretic dynamic model to accommodate the imperfect monitoring of the actions taken by the attacker. We solve this combined partially observable game model using Monte-Carlo tree search to overcome the game model complexity. We give a game-theoretic analysis to explain the attack-defense policies at equilibrium. Finally, we present numerical results to validate the effectiveness of the proposed deception approach.**

***Index Terms*—Cyber deception, multi-agent reinforcement learning, POMDP, game theory.**

I. Introduction

MODERN computer networks are highly connected and heterogeneous in nature to provide necessary complex services with rapidly increasing demands. A heterogeneous network connects computers with different operating systems and protocols, wireless networks accommodate different access technologies and also provide services through a Wi-Fi LAN or a cellular network. The number of devices added to today's networks every day is more than ever [1]. The deployment of wireless-enabled devices [e.g., Internet of Things (IoT), robots, sensors] has made networks larger and denser. These networks are prone to higher levels of interference and attacks which make them more vulnerable. The diversity of devices also makes maintaining them (e.g., patching vulnerabilities, reconfiguration, updating) a more challenging problem.

The IoT network structure is also being deployed in battlefield contexts, where it is known as the Internet of Battlefield Things (IoBT) [2], [3]. In a broader sense, the IoBT also refers to devices useful for military battles that may communicate over tactical networks other than the Internet. Therefore, protecting the resilience and robustness of critical nodes of such a network in a battlefield is crucial. We propose an approach for optimizing cyber deception to prevent possible attackers from characterizing effective attack strategies. Attackers gather information about the targeted systems/networks before launching their attacks [4], collecting inside information about the targeted network using a set of tools and scanning techniques [5]. This can include using software scanning tools like Nmap [6] or traffic analysis techniques to map the network. The network administrator can attempt to protect the network at this early stage of reconnaissance by deceiving the attacker to disguise the true state of the network. We investigate an attack scenario in which the defender protects critical nodes and most important system components, to the defender, via introducing false information (i.e., deception) to disrupt the attacker's decision-making. The applications of these sorts of game theoretic models can shape the understanding of robustness for connected devices in a tactical network (e.g., IoBT). Over the past several decades, there has been a growing concern with network resilience for industries and government, given that security breaches are pervasive, and cyber deception can be an effective proactive network resilience approach to this end. For example, in 2019, the U.S. Department of the Army published a document on its strategy for military deception for multi-domain operations [7], including network security, that addresses deception–this includes the presentation of false weakness in information systems–as a means to dissuade network attacks.

The security games literature studies strategic decision-making problems in a game-theoretic framework between two players, specifically, the network defender and the attacker [8], [9]. Security games applications includes the protection of critical infrastructures, [8], [10], [11], computer networks [12], [13], [14], [15], [16]. The success of strategic deception hinges on the information observed by both players. The defender may strategically leak or let the attacker

Manuscript received 20 September 2021; revised 6 February 2022 and 10 April 2022; accepted 12 April 2022. Date of publication 2 June 2022; date of current version 12 October 2022. The associate editor coordinating the review of this article and approving it for publication was Q. Li. *(Corresponding author: Ahmed H. Anwar.)*

Ahmed H. Anwar and Charles A. Kamhoua are with the Network Security Branch, U.S. Army Research Lab, Adelphi, MD 20783 USA (e-mail: a.h.anwar@knights.ucf.edu; charles.a.kamhoua.civ@army.mil).

Nandi O. Leslie is with the Engineering & Mission Assurance, Raytheon Intelligence & Space, Arlington, VA 22209 USA (e-mail: nandi.o.leslie@rtx.com).

Christopher Kiekintveld is with the Computer Science Department, University of Texas at El Paso, El Paso, TX 79968 USA (e-mail: cdkiekintveld@utep.edu).

Digital Object Identifier 10.1109/TNSM.2022.3179965

access manipulated information about the network that lures the attacker to behave in a certain way that may or may not be observed by the defender. In this work, we focus on Cyber Deception Games (CDG). This class of game models that considers scenarios in which the defender protects the network by manipulating/misrepresenting its state via decoy devices or deceptive signals [17], [18], [19], [20], [21].

Although attackers have used difference aspects and techniques of cyber deception such as phishing, botnets and social engineering, cyber deception could be used to proactively defend against reconnaissance activities taking place in early stages of cyber attacks. The practical usage of honeynets requires comprehensive and configurable honeynets that can be deployed on small-scale devices for commercial and military applications. It is also essential for IoT security enhancement to have devices that are capable of deploying low and hi-fidelity honeypots in a dynamic, flexible and selective manner, [22]. To this end, our proposed approach develops critical mathematical foundations in computational game and decision theory for the strategic allocation of honeypots. Due to the dynamic nature of practical attack defense situations, we propose a reactive deception solution for reactive honeypot allocation given the available network observations and action cost that abstracts the network modification requirement, node criticality, and deployment burden.

The type of information available determines the class of the game to be played between the attacker and the defender. In a complete information game each player fully observes the game reward function, possible actions, as well as the actions taken by the other player. However, if players don't know all the information about their opponent then we have a game with incomplete information. This is different from games with imperfect monitoring: in such games, a player knows all the game information, but does not know what specific action is played by his/her opponent. A framework that combines an evolving game with imperfect monitoring is the partially-observable stochastic game (POSG). In a POSG the game information is known to both players but each player observes the game evolution and/or his opponent's actions partially.

A single-agent dynamic game can be expressed as a Markov decision process (MDP). The optimal policy of an MDP can be obtained efficiently in polynomial time as shown in [23]. Conitzer and Sandholm have shown that hardness of determining whether a pure-strategy Nash equilibrium (NE) exists in a Markov (stochastic) game is $\mathcal{PSPACE}$-hard [24]. Adding uncertainty to an MDP results in a partially-observable MDP (POMDP), where an optimal policy can no longer be obtained in polynomial time. The complexity of solving a POMDP is known to be $\mathcal{PSPACE}$-complete [23]. A POMDP is considered a single-agent POSG. In fact, Goldsmith and Mundhenk [25] have shown that extending a POMDP to a noncooperative multi-agent scenario (i.e., POSG) results in a NEXP^{NP}-complete problem to determine whether there is a "good" strategy for this game and optimal policy existence problem associated with POSGs remains an open problem [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37]. To overcome the prohibitive complexities for the general POSG formulation in this paper, we present an alternative formalism that casts the problem as a POMDP which embeds a game-theoretic component to reason about the actions of the adversary. The proposed formulation not only avoids the complexity of POSG, but also leverages the existing literature of POMDP solvers for efficient solutions that account for the asymmetric information available to the defender and adversary.

Our main contributions are summarized as follows:

- Model a cyber deception approach using honeypot allocation over attack graphs as a two-player zero-sum game under uncertainty between a network defender and an adversary combining game theory and reinforcement learning techniques.
- Develop a two-phase defensive deception: Proactive deception via characterizing the optimal initial honeypot allocation policy against reconnaissance attacks. Second, a reactive deception via dynamic allocation of honeypots in response to IDS security alerts and partially observable malicious activities.
- Unlike existing work, we consider a practical threat model in which the defender can only partially observe the activity of the adversary. We characterize dynamic defensive deception based on a novel POMDP approach that embeds a game-theoretic component to account for the adversary's imperfect monitoring and her rationale.
- We solve the POMDP + GT model via leveraging the Monte-Carlo tree search algorithm (i.e., POMCP) to enhance the scalability of our deception approach and overcome the model complexities due to partial observation and competition aspects.
- We present a thorough numerical analysis to validate the effectiveness of the proposed deception policy in allocating honeypots in terms of defender reward compared with standard allocation policies.

The rest of the paper is organized as follows: First, we discuss other relevant work on cyber deception in Section II. After that in Section III we present our system model and discuss the adversarial mode. In Section IV we formulate the proposed game model and theoretically analyze its static version. Finally, we present our numerical results in Section V and conclude our work in Section VI.

II. Related Work

This work intersects with the literature on POMDPs, game theory, and IRS systems (specifically dynamic IRS via stochastic control).

IRS: Intrusion prevention systems (IPS) and intrusion detection systems (IDS) are both activated before intrusion response systems (IRS) are activated [38] to protect a network or host. An IPS is a proactive defense system and the IDS is a passive defense system. An IRS as a reactive defense system that allows the defender to activate one or more defense mechanisms after receiving alerts from the IDS. For example, this could be to quarantine the threat and regain control and recover the compromised parts of the system [39]. IRS is defined as the set of security countermeasures that are activated upon detecting an attack [40]. Responding to emergent threats could be

manually performed by the system administrator or automated. However, deciding the best countermeasure to take relies heavily on the accuracy of the information reported by the IDS. IRS actions could be costly as it may affect the system performance to mitigate the impact of an attack. Designing a deception defensive mechanism to secure network resources is part of the designing an IRS.

POMDP: In [41] a novel dynamic IRS mechanism was developed which can be triggered via alerts from an anomaly sensor in order to calculate the probability that the host is in an attack state. The approach uses a POMDP to select countermeasures in order to mitigate the attack progressive impact. The authors in [42] developed a POMDP-based defense scheme via modeling the attacker presence in the network using Bayesian attack graphs. Using noisy security alert information, the defender maintains a belief over the current progression of the attacker. In [43] authors proposed a Markov Decision Process based IRS that characterize a defense strategy to enable the network administrator gain control of the attacked network. The approach developed in [42], [44] is related to our model in the sense that they use attack graphs that captures the dependencies between network vulnerabilities/nodes and casted their problem as a POMDP. One of the main features that distinguishes our work is that it takes into account the rational behaviour of the attacker using an embedded game theoretic model within the POMDP formalism.

Game Theory: Cyber Deception Games (CDG) are a growing topic in the game theory literature, modeling scenarios in which the defender protects the network by manipulating its state using decoys or deceptive signals [17], [18], [19], [20]. The defender may camouflage nodes to hide critical networked devices. Recent work formulated honeypot deception games as a one-shot game [45]. Schlenker *et al.* The authors in [46] proposed a defense approach based on a Stackelberg game model to increase attacker uncertainty about the system parameters to misguide the network scanning tools. The attacker gets false information about what operating system is running on what machine, what port is assigned to which service, and the names of subnetworks and active users. In [47], the authors presented a Stackelberg game model in which the defender chooses optimal mitigations that reduce the capability of the attacker to achieve his goals. Graph theory is adopted extensively to study cybersecurity problems because of the value of representing the network topology, vulnerabilities, and exploits [32], [33], [48], [49], [50], [51]. For example, in a social network a malicious user attacks the network by influencing a node. One way to counter his negative influence is by blocking a subset of edges (i.e., a subgraph) [33]. An attack graph captures the dependencies between security conditions and exploits. We adopt a modified attack graph model of that used by Ammann *et al.* [52] for modeling an attack, where each node is a networked device and each link represents a connection between any two devices at some point in time. We believe that our graph model provide clearer guide for the defender to the allocation policy and it also augments the vulnerabilities that leads to node being compromised. Although stochastic games were introduced in 1953 by Shapley [53], the special class of games with partial observation has been understudied in the literature due to its complexity and computational intractability [54]. Special cases of the general POSG model has been addressed in the literature like stochastic games with complete observation parity games [55]. Another class that modeled the partial observability as games with signals was studied in [56]. In [57] the authors studied a series of smaller, related Bayesian games using heuristics such as Q-MDP to provide the future discounted value of actions.

To overcome the POSG complexities in cyber deception and security problems Horák *et al.* [58], proposed a compact representation of the value function in such games. The proposed technique reduced the high dimensionality of the game by projecting the belief space into a lower-dimensional vector space of marginal probabilities. This abstraction dramatically increased the scalability of the POSG algorithm. Moreover, in [59], the authors leveraged their scalable POSG algorithm to optimize honeypot allocation strategies on attack graphs. In their model, the game is played between an attacker who targets a specific node and the network defender whose goal is to discover the target through optimal honeypot allocation. To overcome the game complexity, authors adopted the compact representation abstraction technique in addition to the method of incremental strategy-generation which expands the action space throughout the game. However, solving POSG in its most general setting is still an open research problem. We believe that an attacker in the reconnaissance stage of the cyber kill chain (CKC) does not have all the information about the network connectivity, node types, vulnerabilities, etc. to launch an attack on a large scale. This assumption allows us to solve a reduced-POSG game model leveraging efficient POMDP solvers, while the game model still captures the attackers local information and rationale.

Our approach accounts for the attacker best response strategies through an embedded game theoretic model within a POMDP formulation. Recent work by Al Amin *et al.* [60], [61] proposed a related system model with decoy allocation for deception based on POMCP. However, unlike our two-phase deception model, in their model the attacker preferences are not captured. Although authors in [61] use a POMCP algorithm to model partial observability, both the transition probability matrix and observation matrix ignore the attacker model, action preference, and attack cost. Conversely, we adopted a game theoretic model to study this cyber deception problem taking into account the strategic actions for both players and the associated costs. We also assume no prior information available for the defender. Nevertheless, the developed game model allowed the defender to develop a proactive deception policy as shown in phase I, and to estimate the attacker's dynamics in the reactive deception phase. Moreover, unlike [61] and [62] our proposed formulation assumes that the attacker actions are affecting the state transitions (and hence the belief update rule) of the underlying partially observable Markov decision problem (POMDP).

III. System Model

We now present our network model and the game model for cyber deception.

Let $\mathcal{G}(\mathcal{N}, \mathcal{E})$, denote a directed graph to represent the network attack graph, where $\mathcal{N}$ is the set of nodes where each node may represent a true host in the network or a vulnerability associated with this node. The defender categorizes the nodes into three categories of nodes, *'entry'*, *'intermediate'*, and *'target'* nodes. An entry node represents the Web servers, IoT devices, and nodes that are accessible from outside the network. The attack starts when the attacker successfully compromise one or more entry nodes. The subset of *'target'* nodes contain the most valuable nodes in the network, such as database servers or nodes that are directly linked to the chain of command in a military tactical network. The remaining nodes that represent the stepping stone nodes that the adversary needs to compromise in order to reach a target are called *'intermediate'* nodes. $\mathcal{E}$ is a set of directed edges $\mathcal{E}$ constituting the network connectivity such that $e_{u,v} \in \mathcal{E}$, indicates that a destination node v is reachable from a source node u, for $u, v \in \mathcal{N}$. Suppose that the attacker has infiltrated node u, and the defender wants to lure the attacker from node u to node v. The defender may choose to protect the true node v by placing a honeypot on the edge $e_{u,v}$. The attacker now interacts with a honeypot if she tried to reach node v through an edge $e_{u,v}$ after the network defender has placed a honeypot on the edge connecting u and v. This abstraction allows us to deal with a fixed-size network graph. The deception strategy changes the status of each edge as either 'false' or 'true' to indicate whether a honeypot is placed over this edge or not. A false edge implies that all the data flowing through this edge will be directed to a honeypot instead. By luring an attacker into a honeypot, the defender deceives the attacker into wasting time as opposed to infiltrating an actual networked device. Similar modeling of honeypots as edges of the network has been proposed in [59], [63].

Based on a risk assessment each node in the attack graph is assigned a value. The node value represents its importance to the network. Although node values are assigned from the defender's perspective, we assume that the attacker can detect the importance of the neighbors of any node that has been compromised, so they can locally evaluate the most valuable nodes to attack next (e.g., high-valued targets).

The objective of the defender is to find the optimal honeypot allocation policy to protect the network; especially but not only the target nodes. This objective is achieved via a two-phase deception approach: one is proactive and places honeypots before the network is being attacked, and the second is reactive using a dynamic deception policy in response to the IDS alerts if the attacker is detected by the honeypots already in place. The defender is interested in revealing the progress of the attacker, and therefore wants to maximize the chance that the attacker interacts with the honeypots.

The system health is affected by the interaction between the network administrator and the attacker malicious activities. The defender is interested in placing honeypots along a path that intersects the attacker route from network entry point to a target. The attacker is interested in avoiding the honeypots, while attacking the most valuable nodes with every action or movement. We next describe the defender and the attacker model and possible actions in more detail.

A. The Defender Model

The defender's goal is to protect the target nodes, $\mathcal{T}$ via selecting the location(s) of the honeypot(s). Let B denote the total number of honeypots to be allocated (i.e., the budget constraint). To preserve the network topology, we assume that the honeypots will be placed as interfaces to each node. To protect a node, the original node address will be hidden and the IP address will be switched to a honeypot that mimics the location of the true node. A honeypot is modeled as a connecting edge that captures any attempts to reach a node if the defender decided to protect it via a honeypot. To compromise any node, the attacker has to compromise an immediate parent node first. The defender needs to take into account the node values, the total budget, and the possible paths that the attacker can follow before making his decision.

The defender action space for a single honeypot contains all the set of edges. Let $\mathcal{E}$ be the set of edges and let $\mathcal{A}_d$ denote the defender actions space. Considering B honeypots, $\mathcal{A}_d = \{\mathbf{e} \in 2^{\mathcal{E}} \quad s.t \quad \mathbf{1}^T\mathbf{e} \leq B\}$. Where $\mathbf{1}^T\mathbf{e}$ is the inner product of the vectors. $\mathbf{e} \in \mathcal{E}$ is a binary vector of length $|\mathcal{E}|$, with an entry is set to 1 if a honeypot is allocated on this edge, and is set to 0, otherwise. The inner product condition between $\mathbf{e}$ and a vector of all ones, $\mathbf{1}^T$, ensures that the total number of allocated honeypots does not exceed the defender's budget. Therefore, a feasible deception action $a_d \in \mathcal{A}_d$ is binary vector of length $|E|$ such that, $\|a_d\|_1 \leq B$ To avoid trivial deception scenarios, we assume a limited budget that does not allow for covering the entire network via decoys, so a strategic allocation is required to detect the adversary and protect the high-value nodes.

The payoff function for the defender is designed to account for the number of successful and wasteful honeypots. If the attacker goes through an edge where a honeypot has been placed, this represents a successful allocation for the defender. Otherwise it is a cost since the attacker is able to avoid the honeypot. Each successful event for protecting a specific node is multiplied by the node value. Node values are measures of the node importance which can be quantified using methods such as node centrality, functionality, and the sensitivity of the data [64]. The defender assigns a value for each node, which is also assumed to capture the attacker preference inside the network. A loss is penalized by the value of the node that is successfully compromised.

The network administrator plans the allocation of honeypots in a sequential decision-making process that maximizes his expected reward. The defender decides which edges to allocate honeypots to protect. As the system evolves, the set of compromised nodes changes and the defender allocates new honeypots based on any new observations of the attacker.

B. The Attacker Model

We assume that the attacker is able to learn the network topology using passive monitoring and active probing techniques that can map out the network. However, the attacker cannot distinguish between a true edge (an existing unpatched vulnerability) and fake edges (honeypots). The attacker moves from one compromised node to another. If the attacker

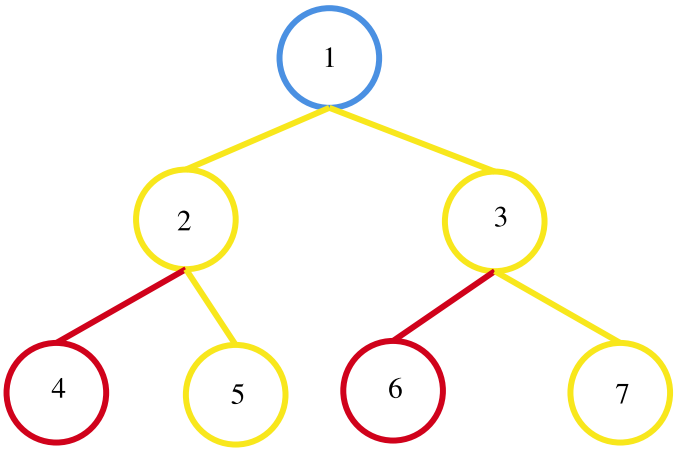

Fig. 1. A 7-node tree network topology with single point of entry and two target nodes (4, 6).

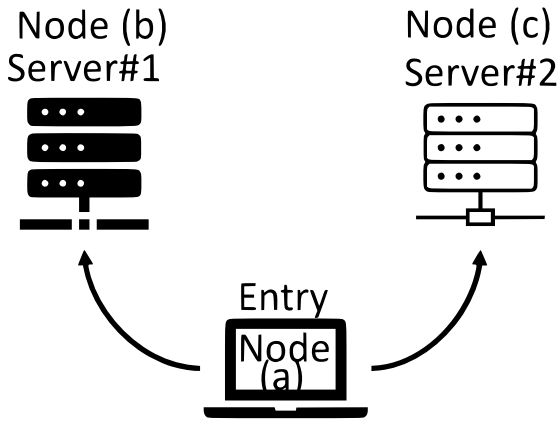

Fig. 2. A 3-node network diagram.

interacts with a honeypot, this allows the defender to gain an accurate belief regarding one or more of the compromised nodes. Based on the network topology, the defender can estimate the set of compromised nodes using the honeypot data. We consider an attacker who expects defensive deception and that some nodes are fake, but does not know the locations of honeypots. Given the local information of neighboring nodes' values the attacker rationally chooses which path to take to compromise the following node. The attacker incurs a cost attack action that represents the potential risk of hitting a honeypot.

The attacker's goal is to reach a target node. The attacker tries to select a path from the network entry node to the target node, avoiding paths that are likely to have honeypots. Let $\mathcal{P}$ be a set that contains all feasible paths between every pair of nodes in the network. For example, path between nodes 1 and 6 for the network in Fig. 1 is $P_{\{1,6\}} = \{1, 3, 6\}$, containing the 2 edges of (1, 3) and (3, 6). Other network topologies may allow for multiple paths between the same two nodes in $\mathcal{P}$. Starting from an entry node v_0, the attacker's action space $\mathcal{A}_a$ is a subset of $\mathcal{P}$ where v_0 is the starting node. An action played by the attacker $a_a \in \mathcal{A}_a$ is a vector containing all the nodes along the attack path leading to one of the target nodes starting from v_0.

C. Game Model for Proactive Deception

The simplest version of this problem can be formulated as a one-shot game. Let $\Gamma_1(\mathcal{N}, \mathcal{A}, \mathcal{R})$ be a tuple, where $\mathcal{N} = 1, 2$ is the set of players, the defender and the attacker. Let $\mathcal{A} = \mathcal{A}_d \times \mathcal{A}_a$ denote the action spaces for both players, and $\mathcal{R}$ is the reward space. Let R_d be defender reward, $R_a = -R_d$ for a zerosum game. The defender action space $\mathcal{A}_d$ is the set of possible honeypot locations, i.e., the set of edges on the network graph. On the attacker side, $\mathcal{A}_a$ represents the set of nodes to be attacked forming a path to a target. Next, we describe the game reward for the defender associated with each action profile $(a_d, a_a) \in \mathcal{A}$.

1) Reward Function: The network defender incurs a fixed cost for placing a new honeypot on an edge in the network. This cost is an average cost per honeypot assuming the cost per honeypot is fixed and denoted by, P_c. This cost prevents the defender from placing honeypots everywhere in the network. In addition, P_c represents the monetary cost of the honeypot and the operational cost including overhead on system performance. If the defender places a honeypot on the same edge that the attacker exploits, the defender gains a reward. Hence, the tradeoff that the defender faces is to reduce the cost of placing honeypots while increasing the chance of capturing the attacker by placing larger number of honeypots. The game formulation can also accommodate an additional constraint on the maximum number of allocated honeypots.

On the attacker side, there is a cost per attack denoted by A_c that represents the risk taken by the attacker. If the attacker exploits a safe edge, the attacker gains a successful attack reward. Let *Cap* and *Esc* denote the defender capture reward and the attacker success reward, respectively.

We adopt a reward function that takes into account the importance of the network nodes. Both the capturing reward and the successful attack reward are weighted by the value of the secured or attacked node, w_v, where $v \in \mathcal{N}$. We start by expressing the reward matrix for the game illustrated in Fig. 2 and present the general reward matrix later.

$$R_d = \begin{bmatrix} -P_c + A_c + Cap * w_b & -P_c + A_c - Esc * w_c & -P_c \\ -P_c + A_c - Esc * w_b & -P_c + A_c + Cap * w_c & -P_c \\ +A_c - Esc * w_b & A_c - Esc * w_c & 0 \end{bmatrix}. \tag{1}$$

The attacker reward matrix, $R_a = -R_d$.

The reward function can be generalized to an arbitrary number of possible edges as follows:

$$R_d(a_d, a_a) = \begin{cases} -P_c + A_c + Cap * w_v; & a_1 = e_{a,v}, a_2 = v \quad \forall v \in \mathcal{N} \\ -P_c + A_c - Esc * w_u; & a_1 = e_{a,v}, a_2 = u \forall u \neq v \in \mathcal{N} \\ -P_c; & a_1 = e_{a,v}, a_2 = 0 \quad \forall v \in \mathcal{N} \\ 0; & a_1 = 0, a_2 = 0 \end{cases} \tag{2}$$

where $a_d = 0$ denotes that the defender is not allocating any new honeypots. Similarly, $a_a = 0$ denotes the attacker decided not to attack.

2) Mixed Strategy: We now define the actions available to each player in the game, i.e., $\mathcal{A}_d$ and $\mathcal{A}_a$. $\mathcal{A}_d$ is the set that contains all the possible honeypot placement locations and $\mathcal{A}_a$

contains all possible attacking paths for the attacker. A pure strategy is a strategy that selects one of these actions. For the defender a pure strategy means to place *B* honeypots over a set of edges in a deterministic way. A pure strategy for the attacker is a specific attack path starting from an entry node to a target node. Alternatively, players may choose to adopt a mixed (randomized) strategy defined through a probability distribution over the set of pure actions (i.e., $\mathcal{A}_d$ and $\mathcal{A}_a$). As we show below in Theorem 1, the game has a solution in terms of mixed strategies. This is consistent with the purpose of the game to deceive the attacker, so the defender exact action is unknown and cannot be computed by his opponent even under perfect information conditions. Given the set of actions of the defender, $\mathcal{A}_d$, let $\Pi(\mathcal{A}_d)$ denote the set of all probability distributions over $\mathcal{A}_d$. Then, the set of mixed strategies for the defender is $\Pi(\mathcal{A}_d)$, denoted by $\mathcal{X}$. Therefore, in a mixed strategy $\mathbf{X} \in \mathcal{X}$, action a_d^i is played with probability x_i such that,

$$\mathbf{X} = \left[x_1, x_2, \ldots, x_n\right]^T, \tag{3}$$

where $n = |\mathcal{A}_d|$. Similarly, the attacker may also play a randomized strategy, $\mathbf{Y} = \left[y_1, y_2, \ldots, y_m\right]^T$, where $m = |\mathcal{A}_a|$.

Hence, the expected defender reward, denoted U_d, can be expressed as

$$U_d = \mathbf{X}^T \mathbf{R_d} \mathbf{Y} \tag{4}$$

Each player aims to maximize his own reward. In a zero-sum game, this implies minimizing the other player's reward. The expected utility for the defender in equilibrium, is $U_d = -U_a$. The minimax theorem implies that U_d holds constant in all equilibria and is the same value that the defender achieves under a minimax strategy by attacker. Using this result, we can construct the optimization problem of the defender as a linear program (LP) as follows,

$$\begin{aligned} \underset{\mathbf{X}}{\text{maximize}} \quad & U_d \\ \text{subject to} \quad & \sum_{a_d \in \mathcal{A}_d} R_d(a_d^i, a_a) x_i \geq U_d, \qquad \forall a_a \in \mathcal{A}_a. \\ & \sum_{i=1}^{|\mathcal{A}_d|} x_i = 1, \quad x_i \geq 0, \end{aligned} \tag{5}$$

The first constraint follows from the definition of Nash equilibrium. Therefore, the expected reward to be maximized needs to be greater than the value of the game. Since the value of the game depends on the mixed strategy played by the attacker ,**Y**, the defender should ensure that his response (i.e., **X**) to that strategy is the best response as shown in the first constraint. The remaining two constraints ensure that **X** is a valid probability distribution. The attacker solves a corresponding LP that can be characterized along the same lines to satisfy that the optimal mixed strategy **Y** is a best response for every possible action played by the defender. The resulting mixed strategy forms a NE for the two players.

3) Nash Equilibrium Analysis: Our model has four different parameters (A_c, P_c, *Cap*, and *Esc*) that affect the attacker and/or defender actions. To better understand the role of each of these parameters, we analyze the game NE theoretically.

For sake of simplicity and to avoid distractions we consider a toy graph. W assume two nodes, *u*, *v* connected through a single edge. The attacker enters the network through node *u*. The attacker has to decide whether to attack node *v* or to back off. The defender decides whether to allocate a honeypot to the existing edge or to save the cost of honeypot allocation. To derive the game equilibrium with a mixed strategy, let the attacker compromise node *v* of value w_v with probability y and the defender allocate a honeypot with probability x. The defender expected reward can be expressed as:

$$\begin{aligned} u_d(x, y) = (-P_c - A_c + Cap * w_v)xy - P_c(1-y)x \\ + (A_c - Esc * w_v)y(1-x). \end{aligned} \tag{6}$$

For every attack strategy y adopted by the attacker, the defender reward:

$$\begin{aligned} u_d(x) = (-2A_c y + (Cap + Esc) * w_v y - P_c)x \\ + (A_c - Esc * w_v)y; \qquad \forall y \in [0,1]. \end{aligned} \tag{7}$$

As shown in (7), the defender expected reward forms a linear equation. If the slope of this line is negative, the defender's best response for the attacker strategy y is to play $x^* = 0$. On the other hand, if $m > 0$, the defender best response is $x^* = 1$, i.e., allocate the honeypot through edge $e_{u,v}$. If the slope $m = 0$, this makes the defender action indifferent to the played strategy y. Hence, it satisfies the equilibrium point of the game.

Lemma 1: For the formulated game, the defender surely allocates a honeypot if the allocation cost A_c satisfies the following condition:

$$A_c < \frac{(Cap + Esc)w_v y - P_c}{2y}; \qquad \forall y \in [0,1], \tag{8}$$

Lemma 2: Similarly, the attacker decides to back off if the cost upon being captured *Cap* satisfies the following condition:

$$Cap > \frac{Esc(1-x) - A_c(1-2x)}{w_v}; \qquad \forall x \in [0,1]. \tag{9}$$

Theorem 1: The formulated game has a NE in a mixed strategy at

$$x^* = \frac{Esc w_v - A_c}{(Cap + Esc)w_v - 2A_c}$$

and

$$y^* = \frac{P_c}{(Cap + Esc)w_v - 2A_c}.$$

Moreover, the game admits a NE in a pure strategy such that the defender allocates a honeypot and the attacker backs off if Lemma 1 and $Cap > A_c / w_v$.

Proof: The proof of the theorem follows directly from the definition of NE. The mixed strategy x^* stated in the theorem ensures that the defender reward, presented in (7), is indifferent to the attacker action, so the defender has no incentive to deviate from this point. Similarly, y^* makes the attacker

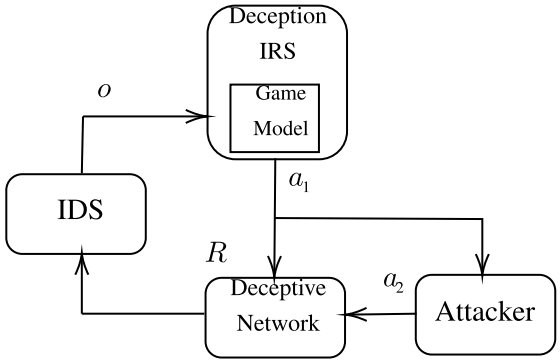

Fig. 3. System model schematic.

reward indifferent to the defender strategy. In the second part of the theorem, the equilibrium characterized in pure strategies follows directly from Lemma 1 and Lemma 2 and under the specified conditions. ■

The obtained mixed strategy at equilibrium represents the optimal initial honeypot allocation policy over the attacker graph as a proactive defensive deception policy to be placed before any attack takes place. Next, we introduce the game dynamics in order to develop the second phase of the proposed deception approach (i.e., the dynamic deception) in response to partially observable malicious activities to be detected by the IDS.

IV. Reactive Deception

We start with a complete, perfect information dynamic game model assuming that both players can perfectly monitor each other's actions to show that some constraints may not hold in practice. Then we introducee a practical POMDP-based dynamic deception algorithm for phase-II of the proposed deception approach as depicted in Fig. 3.

A. Observable Game Model

First, assume that the game is played sequentially with complete information, where both players can perfectly monitor each other's actions. We define the state of the game, $s \in \mathcal{S}$, as a combination of the location of honeypots and the subset of successfully compromised nodes. Hence, the current state of the game captures the progress of the attacker. We assume this information is perfectly known by both players at each state. After each player takes an action, that action is revealed to his/her opponent. The game admits a stationary mixed strategy at every state, s [65]. The optimal randomized stationary policies for the defender and the attacker ($\mathbf{x}^*(s)$ and $\mathbf{y}^*(s)$, respectively) are the solutions to the following $\max\min$ problem that maximizes the discounted expected payoff for each player. We can solve this using dynamic programming value iteration using tabular form for each state. Let $V(s)$ denote the value function at any given state s. Since the game evolves from one state to the other, the value of each state depends on the value if the future state s' as expressed below.

$$V^*(s) = \max_{\mathbf{x}(s)} \min_{\mathbf{y}(s)} \mathbb{E}\Big[R(s,s') + \gamma V^*(s') \Big| \mathbf{x}(s), \mathbf{y}(s)\Big] \quad (10)$$

Here γ is a discount factor, and $\gamma < 1$ ensures that the expected reward is finite for infinite (or indefinite) horizon problems. It also implies that early states yield higher rewards. The immediate reward term is defined as

$$\mathbb{E}\Big[R(s,s') \Big| \mathbf{x}(s), \mathbf{y}(s)\Big] = \sum_{s' \in \mathcal{S}} \sum_{i=1}^{|\mathcal{A}^d|} \sum_{j=1}^{|\mathcal{A}^a|} R_d\Big(s, {a_d}^i, {a_a}^j, s'\Big) \times x^i(s) y^j(s) p\Big(s' | s, {a_d}^i, {a_a}^j\Big). \quad (11)$$

Therefore, the optimal stationary policies are $\pi^* = \{\mathbf{x}^*(\mathsf{s}^1), \ldots, \mathbf{x}^*(\mathsf{s}^n)\}$, *and* $\theta^* = \{\mathbf{y}^*(\mathsf{s}^1), \ldots, \mathbf{y}^*(\mathsf{s}^n)\}$.

To solve (10) for the optimal policy, we adopt the Q-minmax value-iteration based algorithm of Littman [65]. However, this dynamic game model does not hold in practice for the following reasons. The attacker does not know the locations of the honeypots in the network. We assume the attacker is in the reconnaissance stage, and the defender is introducing deception to disrupt the reconnaissance process. Similarly, the defender may not be able to perfectly observe all of the attacker activities and compromised nodes. Hence, assuming complete observations for both players' actions is not practical. Instead, we extend the game into a partially observable game model.

B. Partially Observable Game Model

The network state, $s \in \mathcal{S}$, is a combination of the set of compromised nodes, $\mathcal{N}_c$, and the set of honeypot edges, $\mathcal{E}_h$, where $\mathcal{N}_c \subseteq \mathcal{N}$ and $\mathcal{E}_h \subseteq \mathcal{E}$. Hence, a state denoted by s can be expressed as $s = [\mathcal{E}_h, \mathcal{N}_c]$. The state stores the history of actions taken by both players to model honeypot locations and compromised nodes. However, each player only observes part of the state. The defender knows $\mathcal{E}_h$ and the attacker observes $\mathcal{N}_c$. The defender may not be fully aware of all the compromised nodes, $\mathcal{N}_c$, given that his network defense solutions (e.g., intrusion detection systems) have non-zero error rates. Hence, the defender has a limited view of the network state, s. This belief state is updated whenever an attack action is detected. The attacker does not observe the locations of honeypots before acting in the network. Our main focus is the defender deceptive strategies under this partial observation, so we consider a reduced attacker problem that allows the defender to model the attacker's actions based on a game-theoretic model. The attacker game is limited to the local information that the attacker gain sequentially in the network.

Additionally, the network state transition function $T(s, s', a_1, a_2)$ is deterministic with no exogenous random variable that affects the game evolution. In other words, $Pr(s'|s, a_1, a_2) = 1$ since the game evolution depends solely on the action profile of the two players. However, the defender experiences a non-stationary environment due to the partially observable actions of his opponent. Since the attacker cannot observe $\mathcal{E}_h$ at any time, the attacker can only decide his next action given the available information. A POMDP provides a suitable model for solving for the strategy of the defender, taking into account the asymmetric partial information for both the attacker and defender. The defender assumes that the

attacker cannot observe the network deceptions induced by the defender. Therefore, the attacker's best effort is to reason about his actions using a standard Nash equilibrium at each state.

C. Attacker Problem and Solution Approach

We assume thee attacker has no means to distinguish between real and honeypot nodes. In practice, the attacker may eventually realize that she is being hosted in a honeypot after spending some time interacting with a host. Moreover, the attacker cannot observe the full network topology. However, she can map out the local network topology that contains the immediate neighbors and the associated value, w, of each of the neighboring nodes. The limited view of the network state prevents the attacker from perfectly planning its future actions. Instead, the attacker's best effort is to maximize the immediate reward at every step and state. This decouples the interdependence in the attacker's problem between the current actions and the value of future states, and reduces it to a family of static games that are played sequentially. In other words, the attacker expected reward function can be formulated as,

$$J_a(s) = \min_{a_d|s} \max_{a_a|s} \left[R(S_0, a_d, a_a) + \gamma \sum_{s' \in \mathcal{S}} J_2^*(s')\, T(s, s', a) \right] \tag{12}$$

Since the attacker is only aware of its local neighborhood, the attacker's expected reward function can be reduced to:

$$J_a(s) = \min_{a_d|s} \max_{a_a|s} [R(S_0, a_d, a_a)] \forall s \in \mathcal{S} \tag{13}$$

This reduction of the attacker problem yields a collection of strategic-form games. In each game the attacker considers all of the possible actions taken by the defender, given the locally available network state information. The attacker will account for the possibility that any of the neighboring nodes could be a honeypot. The attacker's action is to select which node to attack next. Since the action spaces for both players are discrete, this game can be represented as a matrix game as seen by the attacker.

D. Defender Problem and Solution Approach

The defender problem is reduced to a single agent acting against an opponent who affects the state transition of the network and therefore the reward of the defender. We formulate the defender problem as a POMDP to account for the partial observation of the state which includes the set of compromised nodes as defined in Section IV-B. The optimal policy of the formulated POMDP is a mapping function between the belief and action spaces $\mathcal{B} \rightarrow \mathcal{A}_d$. Since the defender knows the locations of the honeypots, $\mathcal{E}_h$, we can reduce the state to contain only the set of compromised nodes by the attacker, which represents the hidden part.

$$V^*(b) = \max_{a_d \in \mathcal{A}_d} \left[R(b, a_d) + \gamma \sum_{b' \in \mathcal{B}} \tau(b', a_d, b)\, V(b') \right] \tag{14}$$

For the sake of clarity, we now fully define the POMDP and show how we construct each component. First, we define a belief update rule, $\tau(b', a_d, b)$.

The defender updates its belief state according to the following model:

- Transition probability function, $T(s', a, s) = Pr(s'|s, a)$, where $a = (a_d, a_a)$
- Observations are governed by the associated function, $O(o, s', a) = Pr(o|s', a)$ where s' is the future state at the subsequent timestep.

The transition probability model, $T(s', a, s)$, is a function of a_a as it determines the new compromised node. The defender constructs $T(.)$ as a $|\mathcal{S}| \times |\mathcal{S}|$ matrix such that each entry is $\Pr(a_a|s)$. The defender estimates this probability based on the game model. Assuming he acts rationally, the attacker action would follow his mixed strategy at NE given the available information described earlier, like network topology and the perceived state of the immediate neighbors. The game output gives the probability with which an attacker takes a specific action, so the transition probabilities are estimated from the mixed strategy of the attacker at NE.

Defense observations are network security alerts, generated as a result of the attackers' exploit attempts – some of which may go undetected. These observations are represented by the observation function, $O(o, s', a) = Pr(o|s', a, b)$, where s' is the future state.

$$P(o|s', a_d, b) = \sum_{s \in \mathcal{S}} P(o|s', a_d, s)\, b(s) \tag{15}$$

Since s' includes a_d,

$$P(o|s', a_d, s) = P(o|s', s), \tag{16}$$

which can be represented using total probability law as,

$$P(o|s', s) = \sum_{a_a \in \mathcal{A}_a} P(o|a_a, s', s) P(a_a|s', s), \tag{17}$$

where $P(o|a_a, s', s)$ is an IDS-related model that relates every observation to the attacker's action that can generate it.

Once the defender constructs the observation function, $O(o, s', a)$, and the transition function, $T(s', a, s)$, the defender's belief update rule, $\tau(b', a_d, b)$, can be calculated as:

$$\tau(b', a_d, b) = \sum_{\{o \in \mathcal{O} | SE(b, a_d, 0) = b'\}} P(o|a_d, b) \tag{18}$$

where, $SE(b, a_d, 0)$, denotes the state estimate which is the probability $P(s'|a_d, 0, b)$, expressed as:

$$SE(b, a_d, 0) = \frac{O(s', a_d, o) \sum_{s \in \mathcal{S}} T(s, a_d, s')\, b(s)}{P(o|a_d, b)}. \tag{19}$$

To solve the formulated POMDP we use a standard Monte-carlo tree search algorithm [66]. This approach allows us to scale the proposed deception scheme to the approximate size of a realistic platoon network.

Observation Set: We assume that there is a finite set of observations (related to the alerts generated by IDS, for example). The authors in [67] introduced a lateral movement detection framework to monitor the incoming and outgoing connections between each host in the network. The authors

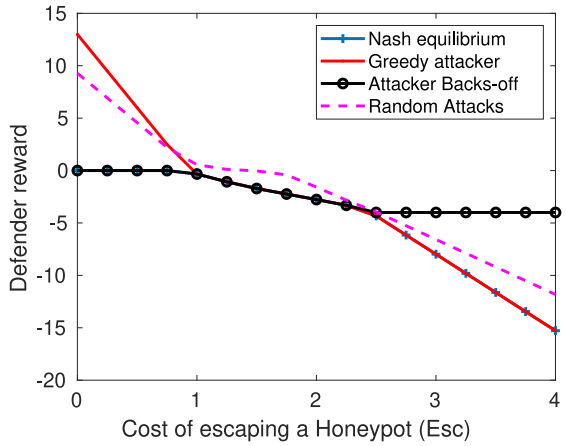

Fig. 4. Defender reward versus the cost of attacker escaping a honeypot.

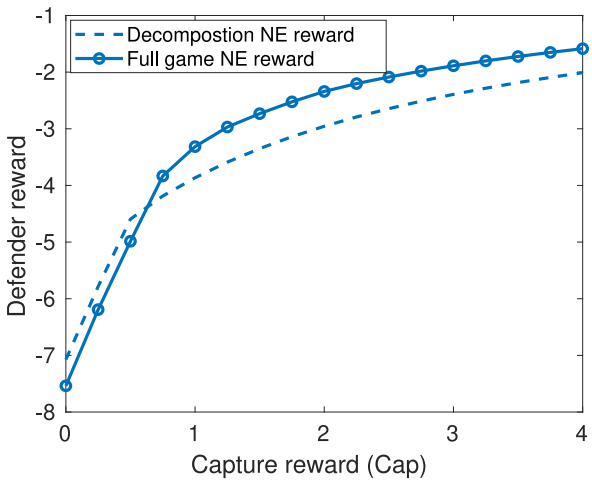

Fig. 5. Defender reward versus the reward parameter of capturing the attacker.

in [68] quantitatively analyzed system security under lateral movement attacks. Specifically, they investigated the transient analysis of a system compromised by these types of attacks to determine when the system will be at risk and how long such risk can last. Therefore, $\mathcal{O}$ represents a finite set of defense observations from the IDS. The defender uses the observation function, $P(o|a_a)$, to estimate the probability that the attacker plays a certain action a_a, for every $o \in \mathcal{O}$.

V. Numerical Results

We present numerical results to validate the proposed game-theoretic models. First, we present the proactive deception approach using the game model defined for one-shot (static) game as well as for the multi-stage (dynamic) game model. Second, we present the numerical results for the reactive game model under uncertainty. Finally, we discuss our findings, model limitations, and future directions of our ongoing research.

A. Proactive Deception: Phase-I

The one-shot game model provides the initial honeypot allocation strategy for the defender, and is part of the game that is embedded in the partially observable case for the reactive deception phase. In this case, we solve the zero-sum game defined in Section IV-C and find the Nash equilibrium in terms of the mixed strategies, x^* and y^*, for the defender and attacker strategies, respectively.

The 7-node network topology shown in Fig. 1 represents a path graph (tree) with single root node (i.e., the entry node $\mathcal{E} = \{1\}$). In this scenario we define the set of 'target' nodes as four nodes, $\mathcal{T} = \{4, 5, 6, 7\}$. Hence, the attacker has four possible routes to reach any of the four target nodes.

To show the effectiveness of the proposed cyber deception approach, we compare the Nash equilibrium strategy for placing honeypots against a class of existing attack polices. In Fig. 4 we plot the defender reward against several cost values for escaping a honeypot in the network. We compare the Nash equilibrium reward to a heuristic where the attacker is greedy and always chooses to attack the highest-valued nodes regardless of the costs. We also consider the other extreme where the attacker is very conservative to spend any resources and hence backs off immediately. In addition, we plot an attacker who does not have much surveillance information and attacks the network randomly. As expected of Nash equilibrium strategy, when the attacker deviates from his rational strategy $\mathbf{Y}^*$, the defender gains a higher reward. When *Esc* is low, the defender reward against a greedy attacker is much higher than the Nash equilibrium since it is better for the attacker in this case to back off rather than attack since the gain is minimal. On the other hand, when the reward of evading a honeypot is high, it is better to attack the network. In this case, if the attacker backs off the defender reward is higher than the NE reward. This short investigation illustrates the trade-off captured by the game and shows that naive policies far from optimal.

The dimension of the action space for both the attacker and the defender grows exponentially with the size of the network. When he number of possible actions for each player is 7 we have 49 possible outcomes for a one-shot game. To overcome this complexity, we propose a decomposition-based NE solution for this game under the *one-look ahead* assumption for both players. Under the assumption of one-step ahead, we can solve three subgames: one at each parent node (1, 2, 3). In Fig. 5 we plot the defender reward for both game solving approaches, i.e., the full game and the subgame decomposition approach. The two approaches are behave similarly in terms of the defender reward over a range of capturing reward values, $Cap \in [0, 4]$. Although characterizing a theoretical gap between the expected outcome for the two approaches is intractable and depends on the structure of the payoff function, our results show that the behavior is comparable. Therefore, we propose this approach as a feasible scalable solution to overcome the exponential explosion in the game size for large-scale networks. In Fig. 5, the game parameters were set as follows: the cost per attack, $A_c = 6.5$, honeypot placement cost, $P_c = 2$, and honeypot escaping cost, $Esc = 2$. While fixing other game parameters, the reward of the defender monotonically increases as *Cap* value increases.

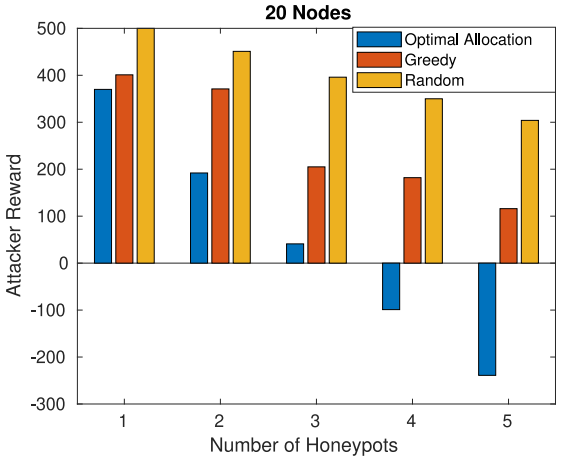

Fig. 6. Attacker reward versus the number of honeypots.

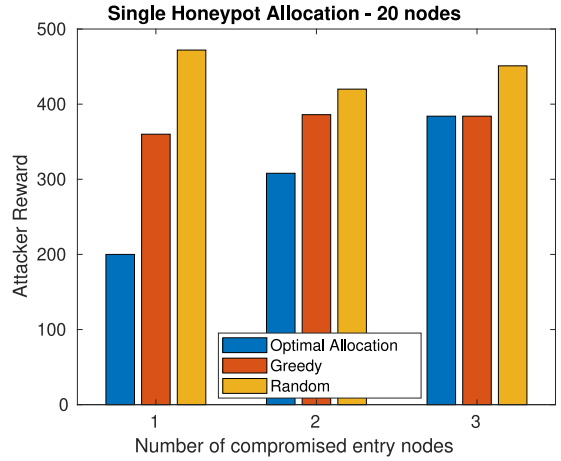

Fig. 7. Attacker reward versus the number of compromised entry nodes.

We consider a 20-node network topology with 3 potential entry nodes and 3 target nodes. The remaining subset of nodes are intermediate nodes. We evaluate the proposed deception approach in terms of the attacker's reward. In Fig. 6 we plot the average attack reward for different honeypot budgets. We compare the performance for our approach 'optimal allocation' to a 'greedy' allocation that always allocates honeypots to protect the nodes with higher values regardless of the network topology. We also compare against 'random' allocation policy. This analysis shows that the optimal honeypot budget for this network is to use three or more honeypots, which dramatically mitigates the effect of the attack. We also plot the attacker's performance (reward) for a different number of compromised nodes in the network as shown in Fig. 7. The three entry nodes are compromised at the start of the attack, so the attacker can attack using any of the existing paths in the network. In this case the optimal allocation and the greedy yield the same level of reward.

Moreover, we tested the proposed deception strategies against random and greedy strategies for large scale networks. We generated a random network graphs of ranging from 50 to 95 nodes. We generated small-world Watts-Strogatz graphs.

TABLE I
THE % INCREASE IN PERFORMANCE OF THE PROPOSED DECEPTION STRATEGIES COMPARED TO RANDOM AND GREEDY ALLOCATION STRATEGIES

Nodes	Random (%)	Greedy (%)
50	48.1	45.01
60	45	43.16
70	38.41	28.18
80	22.9	23.24
95	20.9	19.72

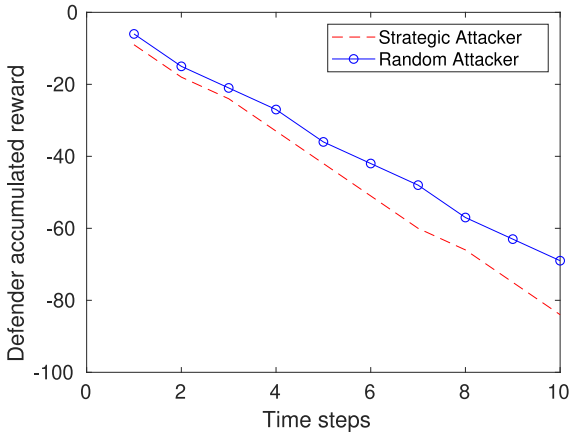

Fig. 8. Defender accumulated reward comparing NE reward and random attack strategy.

For two honeypots and two entry nodes, we compare the game-theoretic deception strategy to the random and greed strategies as depicted in Table I. The greedy algorithm that does not take the network topology into account and depends only on node values, while the random strategy represent a blind deception strategy that ignores both the node values and the connectivity. The proposed algorithm is 48% better than the random strategy and 45% better than the greedy algorithm, for a 50 node graph. As the network size increases, the three strategies converge to the same performance.

Multi-Stage Game: In Fig. 8, we compare the defender cumulative reward over time when both players adopt NE strategies to the case where the attacker is less informed and adopts a random attack strategy. The attack reward can never be better when deviating from the NE strategy, so for the zero-sum game the defender reward increases for this case. Figure 10. represents a scenario in which the game is played over multiple stages. The defender reward function follows as in eqn (2). Each time step in Figure 10 represents the game being at specific state as defined in Section IV-B, $s = [\mathcal{E}_h, \mathcal{N}_c]$. The game is played at this security state forward and the reward is collected as the game transitions from one state to the next. Since the network is under attack, the defender reward is likely to be non positive. The y-axis for Figure 10 is the cumulative reward over the 10 transitions which is decreasing. However, compared to the random policy, the slope of proposed allocation policy is less indicated a less successful attack. For the tree network plotted in Fig. 1, we show that the

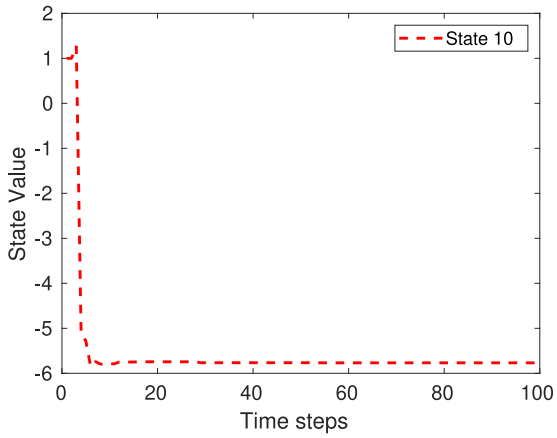

Fig. 9. State value function converges over time using Q-minmax algorithm.

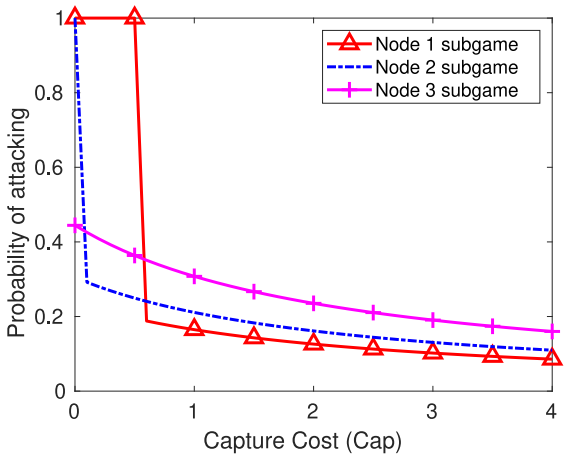

Fig. 11. Probability of attacking against capture cost when Esc = 2.25.

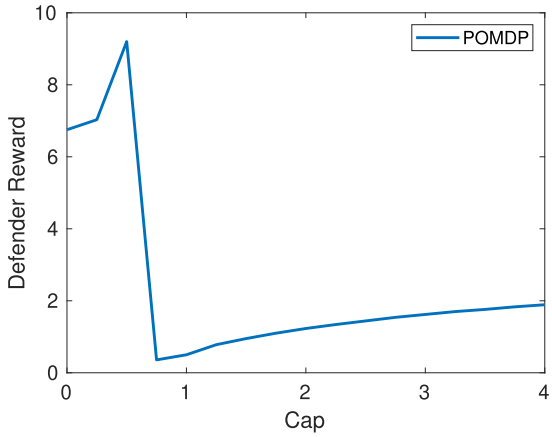

Fig. 10. Defender reward against capture cost for a 7 node network.

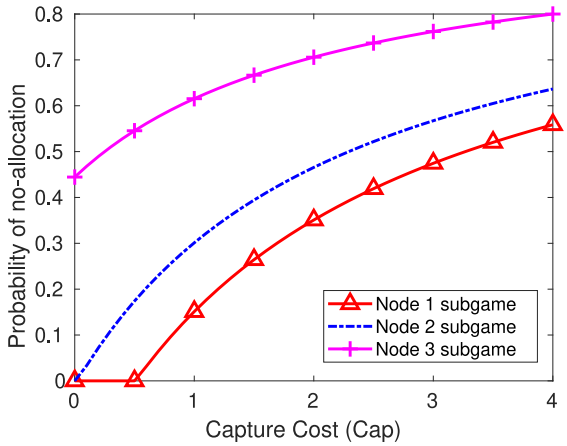

Fig. 12. Probability of no-allocation against capture cost when Esc = 2.25.

algorithm converges to NE policies for a particular state as in Fig. 9. We observe that the algorithm converges in less than 10 iterations to the NE policy for a particular state. We remark that for the dynamic game the future state value automatically captures the potential reward associated with any future state and accounts for it in the total expected reward function. Next, we plot the numerical results for reactive deception under uncertainty.

B. Reactive Deception: Phase-II

As detailed in Section III and IV-B, the proposed POMDP deception approach embeds a game model. The embedded game model is a predictive model of the possible actions of the attacker and provides guidance to the defender on the likely future transitions. Recall that the defender does not observe all of the attacker's actions. In Fig. 10, we plot the defender reward for different values of the capturing cost (Cap) for a 7-node network. To investigate the attacker and defender behaviors at each of the possible stages of the game (i.e., Node 1, Node 2, Node 3 subgames). The first stage represents the game starting at the root node. The other two possible stages, are Node 2 subgame, and Node 3 subgame. In Fig. 11, the probability of attacking the network at the entry point is plotted as well as at the other two stages, as the cost associated with successful deception increases, the attacker tends to back-off to avoid detection. On the defender side, in Fig. 12 shows the probability that the defender does not allocate honeypots as Cap increases, as the defender best respond to a no-attacking strategy is to avoid the cost of placing honeypots. However, as the loss in case of successful attack increase (i.e., Esc = 4.25), the defender is more willing to allocate honeypots to avoid costly successful attacks, Fig. 14. The defender reward achieves its maximum at the point when the defender backs off the network at the entry point. in Fig. 11.

In Fig. 16, the defender is plotted for the 13-node network as plotted in Fig. 18.

To analyze a more interactive attack scenario, we set Esc to be 4.25 instead of 2.25. The increase in the defender's reward parameter encourages the defender to place honeypots with higher probability as shown in Fig. 14. The defender probability of no allocation at these 3 nodes went down compared to Fig. 12. This behavior results in the attacker best responding by attacking the network less as shown in Fig. 13.

Moreover, we evaluate the proposed reactive deception approach on a larger 30-node network. In Fig. 19 we plot the

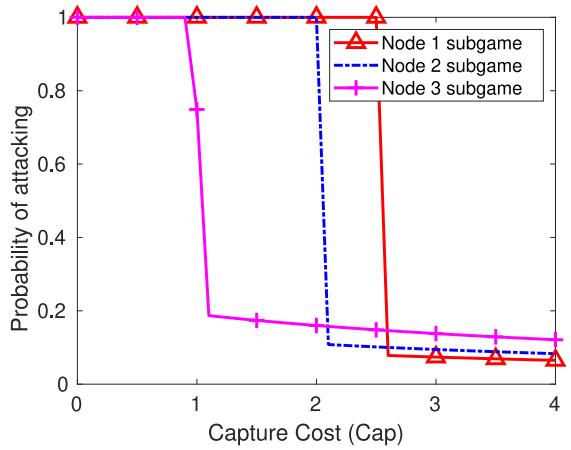

Fig. 13. Probability of attacking against capture cost when Esc = 4.25.

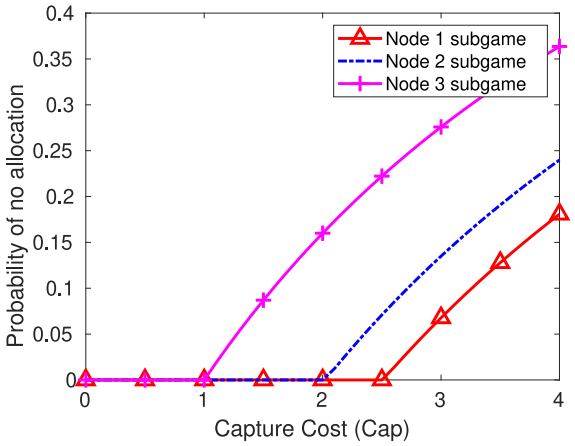

Fig. 14. Probability of no-allocation against capture cost when Esc = 4.25.

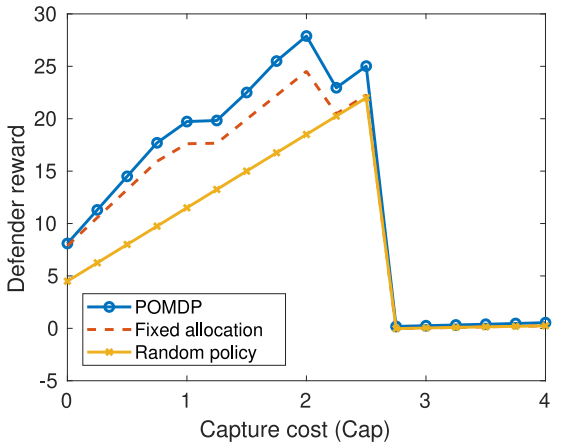

Fig. 15. Defender reward against capture cost for a 7 node network.

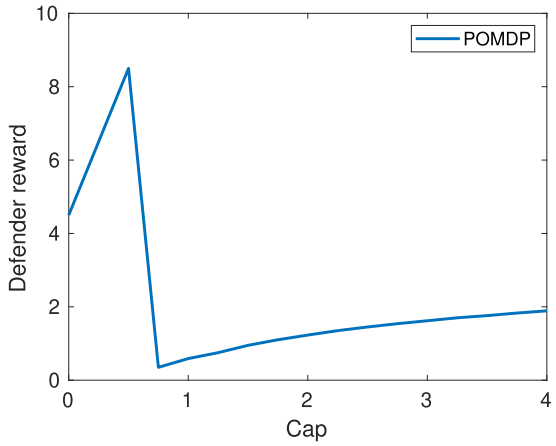

Fig. 16. Defender reward against capture cost for a 13 node network.

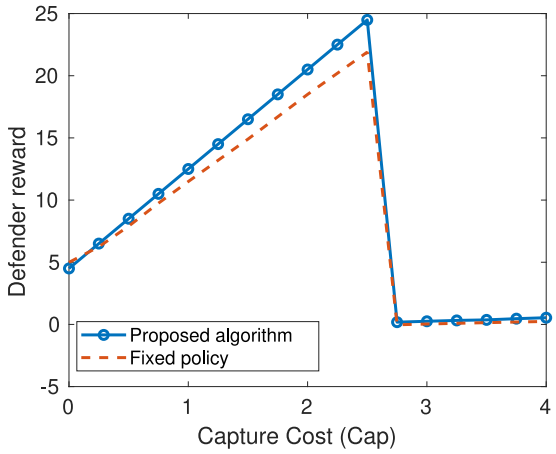

Fig. 17. Defender reward against capture cost for a 13 node network at Esc = 4.25.

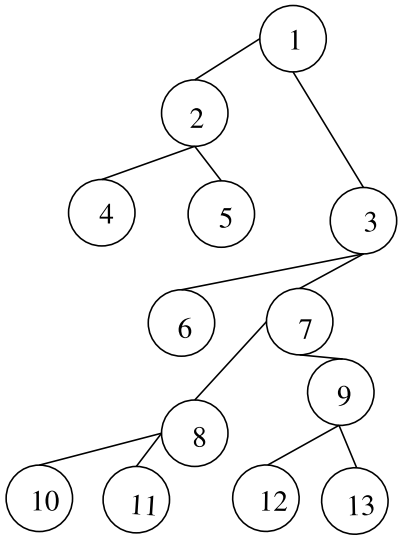

Fig. 18. A 13-node network topology.

defender's reward in comparison to a random allocation policy. Based on the analysis, the defender reactive policy tends to suggest placing a new honeypot along the way to the nearest target node such that it enhance the probability of keep tracking the attacker.

C. Discussion and Future Work

We have provided a framework to characterize a honeypot allocation strategy for cyber deception. In our model, we used

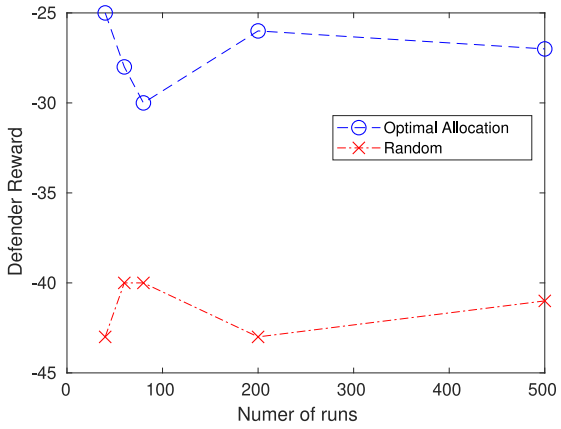

Fig. 19. Defender reward of reactive deception applied to a 30-node network versus number of runs per simulation, with 5 steps for each run.

a game-theoretic approach. We also leveraged Reinforcement learning techniques for decision-making under uncertainty. Both game-theoretic models and RL techniques are known to suffer computational complexities when applied to large-scale networks in addition to the complexities of the associated action space and state-space. Nevertheless, we have provided complexity reduction techniques to deal with solving larger size games using the subgames and decomposition approaches. We also leveraged the state-of-the-art RL technique (POMCP) with game theory to enhance the efficiency of our approach. Moreover, in our ongoing work, we are investigating other solutions to enhance the scalability of our approach. For example, for sparse graphs of large scale, one can solve the honeypot allocation problem per cluster to exploit the structure of the graph. We can also leverage parallel computing to protect attacks from different entry points in parallel.

Future research directions also include node mobility and zero-day attacks. We are considering the effects of possible zero-day attacks on the proposed deception strategies. Zero-day attacks exploit unknown vulnerabilities, hence the graph may have unknown edges to the defender. Additionally, in the future, we plan to consider dynamic network topologies with mobile nodes that can get connected and disconnected to the network graph. Hence, the defender needs to update the defense strategy. Due to the time and efforts needed for this update, the defender needs to find the optimal update policy including partial and periodic updates.

VI. Conclusion

Cyber deception is being applied more extensively to overcome the unknowns of security in tactical and enterprise networks alike. We propose a novel defensive deception approach based on honeypot allocation approach under uncertainty. We presented both a proactive deception scheme and a reactive deception scheme. To overcome uncertainty, we combined a game-theoretic and reinforcement learning solution approach using a POMDP framework. The proposed approach leverages existing efficient POMDP solvers to obtain the defender best reward, while still accounting for the adversarial responses of attackers. We conducted numerical analysis to validate the effectiveness of the proposed deception policy in allocating honeypots over synthesized network graphs compared with plausible baselines. The results show that the proposed defense approach outperforms fixed, random, and less sophisticated honeypot allocation policies across a range of conditions. Our results also show that the proposed algorithmic approach is scalable enough to find strong strategies for realistic network sizes, despite the complications of adversarial interactions and uncertainty for both the attacker and defender.

Acknowledgment

Research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-19-2-0150. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

[1] "Global mobile data traffic forecast update, 2016–2021 white paper," Cisco, San Jose, CA, USA, Rep. C11-738429-00, 2017.
[2] A. Kott, A. Swami, and B. J. West, "The Internet of battle things," *Computer*, vol. 49, no. 12, pp. 70–75, Dec. 2016.
[3] C. A. Kamhoua, "Game theoretic modeling of cyber deception in the Internet of battlefield things," in *Proc. 56th Annu. Allerton Conf. Commun. Control Comput. (Allerton)*, 2018, p. 862.
[4] N. C. Rowe and H. C. Goh, "Thwarting cyber-attack reconnaissance with inconsistency and deception," in *Proc. Inf. Assur. Security Workshop*, 2007, pp. 151–158.
[5] P. Engebretson, *The Basics of Hacking and Penetration Testing: Ethical Hacking and Penetration Testing Made Easy*. San Diego, CA, USA: Elsevier, 2013.
[6] G. F. Lyon, *Nmap Network Scanning: The Official Nmap Project Guide to Network Discovery and Security Scanning*. Sunnyvale, CA, USA: Insecure, 2009.
[7] (Dept. Army, Arlington, VA, USA). *Army Support to Military Deception (FM 3-13.4)*. (2019). [Online]. Available: https://fas.org/irp/doddir/army/fm3-13-4.pdf
[8] C. Kiekintveld, M. Jain, J. Tsai, J. Pita, F. Ordóñez, and M. Tambe, "Computing optimal randomized resource allocations for massive security games," in *Proc. 8th Int. Conf. Auton. Agents Multiagent Syst. Vol. 1*, 2009, pp. 689–696.
[9] A. H. Anwar, G. Atia, and M. Guirguis, "Game theoretic defense approach to wireless networks against stealthy decoy attacks," in *Proc. 54th Annu. Allerton Conf. Commun. Control Comput. (Allerton)*, 2016, pp. 816–821.
[10] J. Pita *et al.*, "Deployed ARMOR protection: The application of a game theoretic model for security at the Los Angeles international airport," in *Proc. 7th Int. Joint Conf. Auton. Agents Multiagent Syst. Ind. Track*, 2008, pp. 125–132.
[11] E. Shieh *et al.*, "Protect: A deployed game theoretic system to protect the ports of the United States," in *Proc. 11th Int. Conf. Auton. Agents Multiagent Syst. Vol. 1*, 2012, pp. 13–20.
[12] O. Vanek, Z. Yin, M. Jain, B. Bosanskỳ, M. Tambe, and M. Pechoucek, "Game-theoretic resource allocation for malicious packet detection in computer networks," in *Proc. AAMAS*, 2012, pp. 905–912.
[13] K. Durkota, V. Lisỳ, B. Bošanskỳ, and C. Kiekintveld, "Optimal network security hardening using attack graph games," in *Proc. 24th Int. Joint Conf. Artif. Intell.*, 2015, pp. 526–532.
[14] C. Kiekintveld, V. Lisỳ, and R. Píbil, "Game-theoretic foundations for the strategic use of honeypots in network security," in *Cyber Warfare*. Cham, Switzerland: Springer, 2015, pp. 81–101.

[15] W. Cai *et al.*, "A survey on cloud gaming: Future of computer games," *IEEE Access*, vol. 4, pp. 7605–7620, 2016.

[16] T. Nguyen, M. P. Wellman, and S. Singh, "A Stackelberg game model for botnet data exfiltration," in *Proc. Int. Conf. Decis. Game Theory Security*, 2017, pp. 151–170.

[17] T. E. Carroll and D. Grosu, "A game theoretic investigation of deception in network security," *Security Commun. Netw.*, vol. 4, no. 10, pp. 1162–1172, 2011.

[18] Y. Li, Y. Xiao, Y. Li, and J. Wu, "Which targets to protect in critical infrastructures—A game-theoretic solution from a network science perspective," *IEEE Access*, vol. 6, pp. 56214–56221, 2018.

[19] W. Xu, W. Trappe, Y. Zhang, and T. Wood, "The feasibility of launching and detecting jamming attacks in wireless networks," in *Proc. 6th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2005, pp. 46–57.

[20] A. Clark, Q. Zhu, R. Poovendran, and T. Başar, "Deceptive routing in relay networks," in *Proc. Int. Conf. Decis. Game Theory Security*, 2012, pp. 171–185.

[21] S. Jajodia, P. Shakarian, V. S. Subrahmanian, V. Swarup, and C. Wang, *Cyber Warfare: Building the Scientific Foundation*, vol. 56. Cham, Switzerland: Springer, 2015.

[22] J. C. Acosta, A. Basak, C. Kiekintveld, and C. A. Kamhoua, "Lightweight on-demand honeypot deployment for cyber deception," in *Proc. 12th EAI Int. Conf. Digit. Forensics Cyber Crime (EAI ICDF2C)*, Singapore, 2021, pp. 1–19.

[23] C. H. Papadimitriou and J. N. Tsitsiklis, "The complexity of Markov decision processes," *Math. Oper. Res.*, vol. 12, no. 3, pp. 441–450, 1987.

[24] V. Conitzer and T. Sandholm, "New complexity results about Nash equilibria," *Games Econ. Behav.*, vol. 63, no. 2, pp. 621–641, 2008.

[25] J. Goldsmith and M. Mundhenk, "Competition adds complexity," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, 2008, pp. 561–568.

[26] O. Madani, S. Hanks, and A. Condon, "On the undecidability of probabilistic planning and infinite-horizon partially observable Markov decision problems," in *Proc. AAAI/IAAI*, 1999, pp. 541–548.

[27] G. Shani, J. Pineau, and R. Kaplow, "A survey of point-based POMDP solvers," *Auton. Agents Multi-Agent Syst.*, vol. 27, no. 1, pp. 1–51, 2013.

[28] A. Neyman, S. Sorin, and S. Sorin, *Stochastic Games and Applications*. vol. 570. Dordrecht, The Netherlands: Springer, 2003.

[29] M. Jain, V. Conitzer, and M. Tambe, "Security scheduling for real-world networks," in *Proc. Int. Conf. Auton. Agents Multi-Agent Syst.*, 2013, pp. 215–222.

[30] G. Kamdem, C. Kamhoua, Y. Lu, S. Shetty, and L. Njilla, "A Markov game theoritic approach for power grid security," in *Proc. IEEE 37th Int. Conf. Distrib. Comput. Syst. Workshops (ICDCSW)*, 2017, pp. 139–144.

[31] L. Wang, T. Islam, T. Long, A. Singhal, and S. Jajodia, "An attack graph-based probabilistic security metric," in *Proc. IFIP Annu. Conf. Data Appl. Security Privacy*, 2008, pp. 283–296.

[32] A. H. Anwar, C. Kamhoua, N. O. Leslie, and C. Kiekintveld, "A game theoretic framework for software diversity for network security," in *Proc. Int. Conf. Decis. Game Theory Security*, 2020, pp. 297–311.

[33] F. Jia, K. Zhou, C. Kamhoua, and Y. Vorobeychik, "Blocking adversarial influence in social networks," in *Proc. Int. Conf. Decis. Game Theory Security*, 2020, pp. 257–276.

[34] B. Xi and C. A. Kamhoua, "A hypergame-based defense strategy toward cyber deception in Internet of battlefield things (IoBT)," in *Modeling and Design of Secure Internet of Things*. Piscataway, NJ, USA: IEEE Press, 2020, pp. 59–77.

[35] 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 Ubiquitous Syst. Comput. Netw. Serv.*, 2019, pp. 522–526.

[36] C. T. Do *et al.*, "Game theory for cyber security and privacy," *ACM Comput. Surv.*, vol. 50, no. 2, pp. 1–37, 2017.

[37] Z. R. Shi *et al.*, "Learning and planning in the feature deception problem," in *Proc. Int. Conf. Decis. Game Theory Security*, 2020, pp. 23–44.

[38] Z. Inayat, A. Gani, N. B. Anuar, M. K. Khan, and S. Anwar, "Intrusion response systems: Foundations, design, and challenges," *J. Netw. Comput. Appl.*, vol. 62, pp. 53–74, Feb. 2016.

[39] B. Foo, M. W. Glause, G. M. Howard, Y. S. Wu, S. Bagchi, and E. H. Spafford, "Intrusion response systems: A survey," in *Information Assurance: Dependability and Security in Networked Systems*. Boston, MA, USA: Elsevier, 2008, pp. 377–412.

[40] Y.-M. Chen and Y. Yang, "Policy management for network-based intrusion detection and prevention," in *Proc. IEEE/IFIP Netw. Oper. Manag. Symp.*, vol. 2, 2004, pp. 219–232.

[41] O. P. Kreidl and T. M. Frazier, "Feedback control applied to survivability: A host-based autonomic defense system," *IEEE Trans. Rel.*, vol. 53, no. 1, pp. 148–166, Mar. 2004.

[42] E. Miehling, M. Rasouli, and D. Teneketzis, "Optimal defense policies for partially observable spreading processes on Bayesian attack graphs," in *Proc. 2nd ACM Workshop Moving Target Defense*, 2015, pp. 67–76.

[43] S. Iannucci and S. Abdelwahed, "A probabilistic approach to autonomic security management," in *Proc. IEEE Int. Conf. Auton. Comput. (ICAC)*, 2016, pp. 157–166.

[44] E. Miehling, M. Rasouli, and D. Teneketzis, "A POMDP approach to the dynamic defense of large-scale cyber networks," *IEEE Trans. Inf. Forensics Security*, vol. 13, pp. 2490–2505, 2018.

[45] T. Zhang and Q. Zhu, "Hypothesis testing game for cyber deception," in *Proc. Int. Conf. Decis. Game Theory Security*, 2018, pp. 540–555.

[46] A. Schlenker *et al.*, "Deceiving cyber adversaries: A game theoretic approach," in *Proc. 17th Int. Conf. Auton. Agents MultiAgent Syst.*, 2018, pp. 892–900.

[47] J. Letchford and Y. Vorobeychik, "Optimal interdiction of attack plans," in *Proc. AAMAS*, 2013, pp. 199–206.

[48] A. H. Anwar, C. Kamhoua, and N. Leslie, "Honeypot allocation over attack graphs in cyber deception games," in *Proc. Int. Conf. Comput. Netw. Commun. (ICNC)*, 2020, pp. 502–506.

[49] S. Milani *et al.*, "Harnessing the power of deception in attack graph games," in *Proc. Int. Conf. Decis. Game Theory Security*, 2020, pp. 502–506.

[50] A. N. Kulkarni, J. F. H. Luo, C. A. Kamhoua, and N. N. Leslie, "Decoy placement games on graphs with temporal logic objectives," in *Proc. Int. Conf. Decis. Game Theory Security*, 2020, pp. 168–187.

[51] O. Tsemogne, Y. Hayel, C. A. Kamhoua, and G. Deugoue, "Partially observable stochastic games for cyber deception against network epidemic," in *Proc. Int. Conf. Decis. Game Theory Security*, 2020, pp. 312–325.

[52] P. Ammann, D. Wijesekera, and S. Kaushik, "Scalable, graph-based network vulnerability analysis," in *Proc. 9th ACM Conf. Comput. Commun. Security*, 2002, pp. 217–224.

[53] L. S. Shapley, "Stochastic games," *Proc. Nat. Acad. Sci.*, vol. 39, no. 10, pp. 1095–1100, 1953.

[54] V. Kovařík, M. Schmid, N. Burch, M. Bowling, and V. Lisý, "Rethinking formal models of partially observable multiagent decision making," 2019, *arXiv:1906.11110*.

[55] N. Bertrand, B. Genest, and H. Gimbert, "Qualitative determinacy and decidability of stochastic games with signals," *J. ACM*, vol. 64, no. 5, pp. 1–48, 2017.

[56] D. Rosenberg, E. Solan, and N. Vieille, "Stochastic games with imperfect monitoring," in *Advances in Dynamic Games*. Boston, MA, USA: Birkhäuser, 2006, pp. 3–22.

[57] R. Emery-Montemerlo, G. Gordon, J. Schneider, and S. Thrun, "Approximate solutions for partially observable stochastic games with common payoffs," in *Proc. 3rd Int. Joint Conf. Auton. Agents Multiagent Syst.*, 2004, pp. 136–143.

[58] K. Horák, B. Bošanskỳ, C. Kiekintveld, and C. Kamhoua, "Compact representation of value function in partially observable stochastic games," 2019, *arXiv:1903.05511*.

[59] K. Horák, B. Bošanskỳ, P. Tomášek, C. Kiekintveld, and C. Kamhoua, "Optimizing honeypot strategies against dynamic lateral movement using partially observable stochastic games," *Comput. Security*, vol. 87, Nov. 2019, Art. no. 101579.

[60] M. A. R. Al Amin, S. Shetty, L. Njilla, D. K. Tosh, and C. Kamhoua, "Online cyber deception system using partially observable Monte-Carlo planning framework," in *Proc. Int. Conf. Security Privacy Commun. Syst.*, 2019, pp. 205–223.

[61] M. A. R. Al Amin, S. Shetty, L. L. Njilla, D. K. Tosh, and C. A. Kamhoua, "Dynamic cyber deception using partially observable Monte-Carlo planning framework," in *Modeling and Design of Secure Internet of Things*. Piscataway, NJ, USA: IEEE, 2020, pp. 331–355.

[62] M. Zhu, A. H. Anwar, Z. Wan, J.-H. Cho, C. A. Kamhoua, and M. P. Singh, "A survey of defensive deception: Approaches using game theory and machine learning," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 4, pp. 2460–2493, 4th Quart., 2021.

[63] K. Horák, Q. Zhu, and B. Bošanskỳ, "Manipulating adversary's belief: A dynamic game approach to deception by design for proactive network security," in *Proc. Int. Conf. Decis. Game Theory Security*, 2017, pp. 273–294.

[64] Z. Wan, J.-H. Cho, M. Zhu, A. H. Anwar, C. Kamhoua, and M. P. Singh, "Foureye: Defensive deception against advanced persistent threats via hypergame theory," *IEEE Trans. Netw. Service Manag.*, vol. 19, no. 1, pp. 112–129, Mar. 2022.

[65] M. L. Littman, "Markov games as a framework for multi-agent reinforcement learning," in *Proc. 11th Int. Conf. Mach. Learn.*, vol. 157, 1994, pp. 157–163.
[66] D. Silver and J. Veness, "Monte-Carlo planning in large POMDPs," in *Neural Information Processing Systems*. Red Hook, NY, USA: Curran, 2010.
[67] A. Fawaz, A. Bohara, C. Cheh, and W. H. Sanders, "Lateral movement detection using distributed data fusion," in *Proc. IEEE 35th Symp. Rel. Distrib. Syst. (SRDS)*, 2016, pp. 21–30.
[68] Y. Shi, X. Chang, R. J. Rodríguez, Z. Zhang, and K. S. Trivedi, "Quantitative security analysis of a dynamic network system under lateral movement-based attacks," *Rel. Eng. Syst. Saf.*, vol. 183, pp. 213–225, Mar. 2019.

Ahmed H. Anwar (Member, IEEE) received the B.Sc. degree (Highest Hons.) in electrical engineering from Alexandria University, Alexandria, Egypt, in 2011, the M.Sc. degree in wireless Information Technology from Nile University, Egypt, in 2013, and the Ph.D. degree in electrical engineering from the University of Central Florida in 2019. He has been a Postdoctoral Research Scientist with the U.S. Army Research Lab, Adelphi, MD, USA, since 2019. Before that he worked as a Research Assistant with Nile University and Qatar University from 2011 to 2013. His research interests include network security, algorithmic game theory, and machine learning.

Charles A. Kamhoua (Senior Member, IEEE) received the B.S. degree in electronics from the University of Douala (ENSET), Cameroon, in 1999, the M.S. degree in telecommunication and networking and the Ph.D. degree in electrical engineering from Florida International University (FIU) in 2008 and 2011, respectively. He is a Senior Electronics Engineer with the Network Security Branch, U.S. Army Research Laboratory (ARL), Adelphi, MD, USA, where he is responsible for conducting and directing basic research in the area of game theory applied to cyber security. Prior to joining the Army Research Laboratory, he was a Researcher with the U.S. Air Force Research Laboratory (AFRL), Rome, NY, USA, for six years and an Educator in different academic institutions for more than ten years. He has held visiting research positions with the University of Oxford and Harvard University. He has coauthored more than 250 peer-reviewed journal and conference papers that include five best paper awards. He is a co-inventor of 3 patents and 6 patent applications. He has been at the forefront of several new technologies, co-editing four books at Wiley–IEEE Press titled *Game Theory and Machine Learning for Cyber Security*, *Modeling and Design of Secure Internet of Things*, *Blockchain for Distributed System Security*, and *Assured Cloud Computing*. He has been recognized for his scholarship and leadership with numerous prestigious awards, including the 2021 IEEE-USA Harry Diamond Memorial Award "For contribution and leadership in the area of blockchain and game theory for cybersecurity while in U.S. government service," the 2020 Sigma Xi Young Investigator Award "for outstanding leadership and contribution to game theory applied to cyber security," the 2019 US Army Civilian Service Commendation Medal, the 2019 Federal 100-FCW annual awards for individuals that have had an exceptional impact on federal IT, the 2019 IEEE ComSoc Technical Committee on Big Data Best Journal Paper Award, the 2018 Fulbright Senior Specialist Fellowship, two ARL Achievement Award (2019 and 2018), the 2017 AFRL Information Directorate Basic Research Award "For Outstanding Achievements in Basic Research," 40 Air Force Notable Achievement Awards, the 2016 FIU Charles E. Perry Young Alumni Visionary Award, the 2015 Black Engineer of the Year Award, the 2015 NSBE Golden Torch Award—Pioneer of the Year, and selection to the 2015 Heidelberg Laureate Forum. He is the Chair of ARL Africa, an initiative to increase research collaboration between the U.S. Army research Laboratory and research institution in Africa. He is an Associate Editor of the IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT. He has been congratulated by the White House, the U.S. Congress and the Pentagon for those achievements. He is currently an Advisor for the National Research Council Postdoctoral Program, a member of the FIU Alumni Association and Sigma Xi, and a Senior Member of ACM.

Nandi O. Leslie received the B.S. degree *(magna cum laude)* in mathematics from Howard University in 1999, and the M.A. and Ph.D. degrees in applied and computational mathematics from Princeton University in 2002 and 2005, respectively. She serves as a Senior Engineering Fellow with Raytheon Technologies with over 22 years of experience as an applied mathematician. In this position, she currently also serves as a Chief Data Scientist with research and development interests focused on the intersectional fields of data science, machine learning, statistics, stochastic processes, cybersecurity, and sensor performance. She has written over 55 publications in journals, conference proceedings, and chapters in edited books. In addition, she currently serves on over five scientific advisory boards, international society committees, and conference program committees, including as an 1) Advisory Board Member for the Howard University, Center of Excellence in Artificial Intelligence and Machine Learning; 2) Committee Member, Society of Industrial and Applied Mathematics Industry Committee, Education Subcommittee; 3) Committee Member, American Mathematical Society, Short Course Subcommittee; 4) Conference Session Co-chair and Program Committee for the SPIE Defense + Commercial Sensing Conference, Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications; and 5) Founding Partner of the Women of Color in Engineering Collaborative, an organization funded by the Society of Women Engineers. Since 2020, she has served as a Lecturer and Research Advisor for Master's Degree Theses and research in the Applied and Computational Mathematics and Data Science Programs with Johns Hopkins University. Before joining Raytheon, from 2007 to 2015, she led and contributed to government projects on using modeling and simulation (M&S) for sensor (e.g., acoustic, Global Positioning System) performance and cybersecurity for the Navy Submarine Security and Technology Programs, Office of the Under Secretary of Defense for Research and Engineering, and other customers at Systems Planning and Analysis, Inc. From 2005 to 2007, she taught with the Department of Mathematics, University of Maryland, College Park, where she was also a Postdoctoral Researcher sponsored by the National Science Foundation Vertical Integration of Research and Education in the Mathematical Sciences conducting research on M&S of geographical information systems data to better understand and model land-use and cover change in the neotropics. Furthermore, she received the Black Engineer of the Year Award for Outstanding Technical Contribution in Industry in 2020.

Christopher Kiekintveld (Senior Member, IEEE) received the B.S.E. and M.S. degrees from the University of Michigan, and the Ph.D. degree in computer science and engineering from the University of Michigan in 2008. He is an Associate Professor of Computer Science with the University of Texas at El Paso, where he also serves as the Director of Graduate Programs for Computer Science. Before joining UTEP in 2010, he was a Postdoctoral Research Fellow with the University of Southern California for two years. He also works on applications of artificial intelligence methods to security (including both physical and cyber), resource allocation, trading agents, smart grids, and other areas with the potential to benefit society. He currently directs the Intelligent Agents and Strategic Reasoning Laboratory, UTEP, and has mentored more than thirty graduate and undergraduate students. He has coauthored more than 80 articles in peer-reviewed conferences and journals (e.g., AAMAS, IJCAI, AAAI, JAIR, JAAMAS, and ECRA), with an i10-index of more than 60. His primary research interests are in artificial intelligence, especially in the areas of multiagent systems, computational decision making and game theory, reasoning under uncertainty, and multiagent learning. He has received recognition including multiple best paper awards, the UTEP College of Engineering Dean's Award for Research, the David Rist Prize from the Military Operations Research Society, and an NSF CAREER Award. He has given many invited talks and keynote presentations at conferences, academic institutions, and research laboratories around the world, and has helped to organize numerous workshops and conferences including AAAI, AAMAS, and GameSec. He has also had a lead role in developing several deployed applications of game theory for security, including systems in use by the Federal Air Marshals Service and Transportation Security Administration.