A Hypothesis Testing-based Framework for Cyber Deception with Sludging

Abhinav kumar
Department of Computer Science
University of Cincinnati, OH
Email: kumar3a3@mail.uc.edu

Swastik Brahma
Department of Computer Science
University of Cincinnati, OH
Email: brahmask@ucmail.uc.edu

Baocheng Geng
Department of Computer Science
University of Alabama at Birmingham, AL
Email: geng@uab.edu

Charles A. Kamhoua
Network Security Branch
DEVCOM Army Research Laboratory, MD
E-mail: charles.a.kamhoua.civ@army.mil

Pramod K. Varshney
Dept. of Electrical Eng. and Computer Science
Syracuse University, NY
Email: varshney@syr.edu

2024 60th Annual Allerton Conference on Communication, Control, and Computing (Allerton) | 979-8-3315-4103-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/ALLERTON63246.2024.10735267

***Abstract*—In this paper, we present a novel hypothesis testing framework to model an attacker's decision-making during the reconnaissance phase and employ the framework to design strategic deception strategies that can provide the attacker with optimally falsified information structures. We characterize the criterion for such structures to "blind" the attacker, i.e., to completely confuse it, as well as the optimal falsification strategy when the attacker cannot be blinded. We also characterize optimal information acquisition costs that can be imposed on the attacker to maximally "sludge" its decision-making process during the reconnaissance phase. Numerical results have been presented to gain important insights into the developed strategies.**

***Index Terms*—Cyber Deception, Hypothesis Testing, Information Falsification, Sludging.**

I. Introduction

The use of deception techniques to defend a networked system against modern cyber threats has recently attracted attention [1]–[6]. Fundamentally, such techniques provide misleading information to deceive and/or confuse attackers into taking (or not taking) certain actions with the goal of enhancing network security [7]. Deception strategies can help to reduce the likelihood of an attacker's success, lessen the cost of defense from deception-less scenarios, and complement conventional defense measures for enhancing security.

Prior to launching an attack, an attacker typically gathers internal information about the target network during the reconnaissance phase [2], [6], [7] using various scanning techniques [8]–[13]. For example, an attacker may want to gather information about the types of operating systems (OS) that different devices in the network are using, ports that are open in various devices, and the intensity of traffic flows to/from different networked devices. Such exploitable information can aid the decision-making of attackers regarding their attacking strategies and enhance the success of attacks.

To counter such efforts of an attacker, the network's defender can employ deception tactics by strategically providing *falsified* information in the scan results [14]–[18]. However, the design of such deception strategies is still nascent, with the lack of a rigorous *mathematical* characterization of involved information structures, a comprehension of how an attacker would *optimally* process gathered information for its malicious decision-making, and a thorough study of how a defender can strategically employ an understanding of such factors to provide an attacker with *optimally* falsified information.

To address such issues, in this paper, we model an attacker's decision-making during the reconnaissance phase as a *hypothesis testing* problem [19], where the attacker gathers and processes information regarding the *features* of devices (servers) in a network to decide between the hypotheses of a server being "real" or being a "fake" that has been deployed by the defender as a decoy. Features of a server, for example, can correspond to the type of its OS and the intensity of its associated traffic flows. Our adoption of such a hypothesis testing perspective gives rise to a novel principled approach to the characterization of optimal deception strategies that can rigorously exploit the involved information structures.

Further, employing our hypothesis testing-based framework, in addition to studying optimal ways to fabricate falsified information structures, we investigate the novel possibility of *sludging* the attacker by imposing information acquisition costs with a goal to make it maximally harder for the attacker to decide and, subsequently, degrade the quality of its decisions. It can be noted that the notion of sludging has its roots in *Nudge Theory* [20], which, however, unlike the 'friction' that our sludges introduce in the attacker's decision-making process, aims to design decision environments to make it easier for people to decide in a manner that improves their welfare. In particular, the novel contributions of our paper are as follows:

- We present a novel *hypothesis testing* framework for modeling and analyzing an attacker's process of deciding the nature of servers (devices) in a network during the reconnaissance phase.
- We characterize the optimal *decision rule* that enables the

DISTRIBUTION A: Approved for Public Release. Distribution is Unlimited. This work was supported by the U.S. National Science Foundation (NSF) under Award Number CCF-2302197 and by the University of Cincinnati (UC).

attacker to minimize the probability of erroneously deciding the nature of the servers by strategically exploiting information that it acquires during reconnaissance.
- We characterize the criterion the defender must satisfy while feeding the attacker with falsified information to be able to *blind* the attacker, i.e., make it completely confused and incapable of making informed decisions about the nature of servers.
- We characterize the defender's optimal information falsification strategy when the attacker cannot be blinded.
- Under cost budgets of the defender and attacker, we characterize the optimal *information acquisition costs* that the defender should impose on the attacker to maximally sludge its decision-making process.

The rest of the paper is organized as follows. Section II discusses related research. Section III presents our developed hypothesis testing framework and characterizes the attacker's optimal decision rule that can best exploit its acquired information during reconnaissance. Section IV characterizes the defender's optimal blinding and non-blinding information falsification strategies. Section V presents game theoretic strategies for sludging the attacker in our hypothesis testing framework. Finally, Section VI concludes the paper.

II. Related Research

With cyber criminals increasingly launching more sophisticated attacks that can penetrate deeper into today's networked systems, it is imperative to design innovative cyber deception techniques that can complement traditional defense measures.

Before launching an attack, during the reconnaissance phase, an attacker typically performs scans (i.e., monitoring activities) to gather internal information about the target network and identify devices of interest in it [8], [12]. Such scans can be performed by sending *active probes* to various networked devices to understand their features, such as their OSs, open ports, versions of installed applications, and so on [13]. Again, such information can sometimes be *passively* gathered by attackers by scanning traffic (e.g., packet headers) that pass through network routers and switches [10]. To manage scanning activities, various *network fingerprinting* tools can be employed, such as Nmap [9] and P0f [11]. Information gathered by attackers during reconnaissance aids the decision-making of attackers and helps them optimize attacking strategies.

To defeat such reconnaissance efforts and mislead attackers, a defender can employ deception tactics by feeding attackers with *falsified* information in the scan results [14]–[18]. For example, to mislead attackers into incorrectly determining the type of OS of a device, [17] falsifies TCP/IP headers of packets sent by the device. To efficiently implement such packet structure manipulation schemes, [18] explores Software Defined Networking (SDN)-based Data-Plane Programming (DPP) techniques. While [17], [18] explore information falsification-based defenses from a system implementation perspective, the authors in [15] analytically characterize such defense strategies against an attacker who acquires probabilistic knowledge about networked devices' features by performing scans. The work in [14] complements [15] by developing game theoretic ways of providing falsified scan results when the attacker is strategic in nature. The work in [16] considers that attackers could observe multiple features of networked devices in the scan results to decide their attacking strategies, and devises algorithmic procedures to find falsified values of the features that should be presented to deceive attackers.

The use of information falsification-based defenses can be complemented by the use of *honeypots* [5], [21], which are fake entities that can be used as decoys to divert attackers away from valuable assets in a network. To enable such diversions, [1], [22] develop strategies to lure attackers towards honeypots in a strategic context, [3], [23] explore cost effective placement of honeypots in a network, and [24] studies evaluation of different honeypot technologies.

It should be noted that while past work has sought to devise strategies to deceive attackers by feeding them with falsified scan results, they fall short of mathematically characterizing and duly exploiting the involved information structures, leading to the developed information falsification strategies to be ad hoc and sub-optimal in nature– a problem that we aim to tackle by adopting a hypothesis testing perspective for modeling, analyzing, and characterizing the involved information processing and falsification methodologies.

We further leverage our developed hypothesis testing framework to characterize information acquisition costs that should be strategically imposed on the attacker to maximally sludge its decision-making process and degrade the quality of its decisions. Such sludging stems from Nudge Theory [20], which proposes to adaptively design the decision environment so as to influence the decision-making behavior of people in a predictable way. While nudges try to make it easier for people to decide in a way that improves their welfare [20], sludges make the decision-making process more frictional, leading people to decide in a manner that may not be aligned with their best interests [25]. The work in [26] is a preliminary work that discusses the benefits of using sludges in cyber defense, but does not develop an analytical approach for employing such a strategy. Our paper also fills such a void.

III. Cyber Deception Model and Attacker's Optimal Decision Rule

Consider the presence of two servers in a network, viz. Server G and Server H, with a defender ($\mathcal{D}$) using one of them as the real server and the other as the fake one, i.e., as a decoy (honeypot). Consider that $\mathcal{D}$ chooses G to be the real server with a probability P_G and H to be the real server with a probability $P_H = 1 - P_G$. Each server has M features, which an attacker ($\mathcal{A}$) inspects (by performing scans) during the reconnaissance phase to decide which server is real. Features describe characteristics of a server, such as information about the type of its OS, the intensity of traffic flows to/from the server (e.g., high or low), and so on.

We denote the *true* value of feature i of server X as f_i^X, $i \in \{1, \cdots, M\}$, $X \in \{G, H\}$, and consider that $f_i^X \in \{0, 1\}$,

i.e., every feature assumes binary values (e.g., the type of OS of a server being Windows or Linux). Further, we consider that for $i \in \{1, \cdots, M\}$ and $X \in \{G, H\}$,

$$P_i(f_i^X = 0|\text{X is real}) = 1 - P_i(f_i^X = 1|\text{X is real}) \quad (1)$$

is the probability of f_i^X assuming the value 0 when server X is the *real* server, and that

$$P_i(f_i^X = 0|\text{X is fake}) = 1 - P_i(f_i^X = 1|\text{X is fake}) \quad (2)$$

is the probability of f_i^X assuming the value 0 when server X is the *fake* server. For notational simplicity, we denote $P_i(f_i^X = a|\text{X is real})$ as $P_i^R(f_i^X = a)$ and $P_i(f_i^X = a|\text{X is fake})$ as $P_i^F(f_i^X = a)$, $a \in \{0, 1\}$, $i \in \{1, \cdots, M\}$, $X \in \{G, H\}$.

Note that, while our results could be extended to scenarios where $\mathcal{D}$ employs more than two servers whose features assume non-binary values, we consider the model described above in this paper for expositional simplicity of our novel approach.

A. Probabilistic Falsification (Flipping) of Feature Values

To decide which server is real before attacking, we consider that $\mathcal{A}$ inspects (i.e., gathers information about) the feature values of the two servers by performing scans. Further, we consider that $\mathcal{D}$, to strategically employ deception tactics, can provide probabilistically *flipped* (falsified) feature values in the results of the scan that are observed by $\mathcal{A}$. To model restrictions on flipping values of certain features from $\mathcal{D}$'s perspective, we consider that L out of the M features of each server are 'flippable' by $\mathcal{D}$ and that the ratio $L/M = \alpha$. For a flippable feature of a server, we consider $\mathcal{D}$ to send a flipped value of the feature with a probability p to $\mathcal{A}$. In other words, denoting the value of feature i of server X that is observed by $\mathcal{A}$ as u_i^X, we have

$$p = \text{Prob}(u_i^X = b|f_i^X = a) = 1 - \text{Prob}(u_i^X = a|f_i^X = a) \quad (3)$$

$a, b \in \{0, 1\}$, $a \neq b$, $i \in \{1, \cdots, M\}$, $X \in \{G, H\}$. To model $\mathcal{A}$'s uncertainty about which features are flippable, we consider $\mathcal{A}$ to view each feature of each server to be flippable with the probability α. Thus, for $X \in \{G, H\}$ and $i \in \{1, \cdots, M\}$, we have

$$P_i^R(u_i^X = 0) = 1 - P_i^R(u_i^X = 1) = (1-\alpha)P_i^R(f_i^X = 0) + \alpha\big\{pP_i^R(f_i^X = 1) + (1-p)P_i^R(f_i^X = 0)\big\} \quad (4a)$$

$$P_i^F(u_i^X = 0) = 1 - P_i^F(u_i^X = 1) = (1-\alpha)P_i^F(f_i^X = 0) + \alpha\big\{pP_i^F(f_i^X = 1) + (1-p)P_i^F(f_i^X = 0)\big\} \quad (4b)$$

where $P_i^R(u_i^X = a)$ and $P_i^F(u_i^X = a)$ are the probabilities of $\mathcal{A}$ observing the (potentially flipped) value of feature i of server X as a under X being the real and fake server, respectively, $a \in \{0, 1\}$.

B. Optimal Decision Rule of the Attacker

We consider $\mathcal{A}$ to decide which server is real based on the observed values of the M features of each of the two servers, which we consider to be described by an $M \times 2$ matrix, viz. $\big[[u_1^G, u_1^H], \cdots, [u_M^G, u_M^H]\big]^T$. Accordingly, $\mathcal{A}$'s probability of erroneously deciding that the fake server is the real one is

$$\begin{aligned} P_E^{\mathcal{A}} &= P_G P(\mathcal{A} \text{ decides } H \text{ is real}) + P_H P(\mathcal{A} \text{ decides } G \text{ is real}) \\ &= P_G \sum_{s \in S - S_R^G} \prod_{i \in \{1,\cdots,M\}} P_i^R\big(u_i^G = s_i^G\big) P_i^F\big(u_i^H = s_i^H\big) + \\ &\quad P_H \sum_{s \in S_R^G} \prod_{i \in \{1,\cdots,M\}} P_i^F\big(u_i^G = s_i^G\big) P_i^R\big(u_i^H = s_i^H\big) \end{aligned} \quad (5)$$

where set S contains all possible $M \times 2$ feature matrices that $\mathcal{A}$ can observe and S_R^G contains the set of all $M \times 2$ feature matrices observing which makes $\mathcal{A}$ to decide that G is the real server. In (5), s_i^G and s_i^H are the values of feature i of servers G and H, respectively, of the s^{th} $M \times 2$ feature matrix.

Next, we characterize the optimal decision rule that $\mathcal{A}$ should employ to decide the servers' natures based on the observed $M \times 2$ feature matrix.

THEOREM 1. *The optimal decision rule of $\mathcal{A}$ that minimizes its probability of erroneously deciding that the fake server is the real one, i.e., minimizes* (5)*, is:*

$$\sum_{i=1}^{M} \Big\{ \log \frac{P_i^F(u_i^G)}{P_i^R(u_i^G)} + \log \frac{P_i^R(u_i^H)}{P_i^F(u_i^H)} \Big\} \underset{G \text{ is real}}{\overset{H \text{ is real}}{\gtrless}} \log \frac{P_G}{P_H} \quad (6)$$

Proof. $\mathcal{A}$'s error probability (5) can be expressed as

$$\begin{aligned} P_E^{\mathcal{A}} &= P_G \sum_{s \in S} \prod_{i \in \{1,\cdots,M\}} P_i^R\big(u_i^G = s_i^G\big) P_i^F\big(u_i^H = s_i^H\big) + \\ &\quad \sum_{s \in S_R^G} \Big[P_H \Big\{ \prod_{i \in \{1,\cdots,M\}} P_i^F\big(u_i^G = s_i^G\big) P_i^R\big(u_i^H = s_i^H\big) \Big\} - \\ &\quad P_G \Big\{ \prod_{i \in \{1,\cdots,M\}} P_i^R\big(u_i^G = s_i^G\big) P_i^F\big(u_i^H = s_i^H\big) \Big\} \Big] \end{aligned} \quad (7)$$

Since $\sum_{s \in S} \prod_{i \in \{1,\cdots,M\}} P_i^R\big(u_i^G = s_i^G\big) P_i^F\big(u_i^H = s_i^H\big) = 1$, (7) reduces to

$$\begin{aligned} P_E^{\mathcal{A}} &= P_G + \sum_{s \in S_R^G} \Big[P_H \Big\{ \prod_{i \in \{1,\cdots,M\}} P_i^F\big(u_i^G = s_i^G\big) P_i^R\big(u_i^H = s_i^H\big) \Big\} \\ &\quad - P_G \Big\{ \prod_{i \in \{1,\cdots,M\}} P_i^R\big(u_i^G = s_i^G\big) P_i^F\big(u_i^H = s_i^H\big) \Big\} \Big] \end{aligned} \quad (8)$$

Clearly, to minimize (8), those $M \times 2$ feature matrices of S must be assigned to S_R^G that make the second term of (8) negative, i.e., that make

$$\begin{aligned} &P_H \Big\{ \prod_{i \in \{1,\cdots,M\}} P_i^F\big(u_i^G = s_i^G\big) P_i^R\big(u_i^H = s_i^H\big) \Big\} < \\ &\quad P_G \Big\{ \prod_{i \in \{1,\cdots,M\}} P_i^R\big(u_i^G = s_i^G\big) P_i^F\big(u_i^H = s_i^H\big) \Big\} \end{aligned} \quad (9)$$

for any $s \in S_R^G$. Thus, after observing an $M \times 2$ feature matrix, to minimize its error probability (5), which occurs when (9) is satisfied, $\mathcal{A}$ should use the following decision rule to decide which server is real:

$$\prod_{i \in \{1,\cdots,M\}} \frac{P_i^F(u_i^G) P_i^R(u_i^H)}{P_i^R(u_i^G) P_i^F(u_i^H)} \underset{\text{G is real}}{\overset{\text{H is real}}{\gtrless}} \frac{P_G}{P_H} \quad (10)$$

which makes $\mathcal{A}$ decide that Server G is 'real' if the quantity of the LHS of (10) is *less* than the threshold on the RHS

(i.e., P_G/P_H), and decide that Server H is 'real' otherwise. Now, taking the log of both sides of (10) and simplifying it yields (6). This proves the theorem. □

IV. Optimal Falsification-based Deception

In this section, we characterize the optimal strategies that should be employed by $\mathcal{D}$ to falsify (flip) the feature values of the servers in the scan results so as to maximally degrade $\mathcal{A}$'s capability of identifying the real server. To this end, we first present the falsification criterion that makes $\mathcal{A}$'s optimal decision rule in (6) to become completely ineffective in exploiting the observed scan results, in which case we say that $\mathcal{A}$ is *blind*, making $\mathcal{A}$ to experience the maximum possible decision-making error probability.

Lemma 1. *$\mathcal{A}$ becomes blind, i.e., incapable of making an informed decision based on the observed $M\times 2$ feature matrix, when $\alpha p = 1/2$.*

Proof. When $\mathcal{A}$ becomes blind, i.e., its decision rule (6) becomes completely ineffective, $\mathcal{A}$ would have to decide which server is real solely based on P_G and P_H (i.e., decide G is real if $P_G > P_H$, and decide H is real otherwise). From the LHS of (6), clearly, this happens when $P_i^R(u_i^X = a) = P_i^F(u_i^X = a)$, $i \in \{1, \cdots, M\}$, $X \in \{G, H\}$, $a \in \{0, 1\}$, which, using (4a) and (4b), yields

$$\begin{aligned}(1-\alpha)P_i^R(f_i^X = a) + \alpha\big[p\{1 - P_i^R(f_i^X = a)\} + (1-p)\\ P_i^R(f_i^X{=}a)\big]{=}(1{-}\alpha)P_i^F(f_i^X{=}a) + \alpha\big[p\{1 - P_i^F(f_i^X{=}a)\}\\ + (1-p)P_i^F(f_i^X = a)\big] \end{aligned} \tag{11}$$

Simplifying (11), we get

$$\big\{1 - 2\alpha p\big\}\{P_i^F(f_i^X = a) - P_i^R(f_i^X = a)\} = 0 \tag{12}$$

Clearly, (12) is satisfied when $\alpha p = 1/2$, which proves the lemma. □

Next, we characterize the minimum fraction of flippable features needed to blind $\mathcal{A}$.

Corollary 1. *The minimum fraction of flippable features needed to blind $\mathcal{A}$ is $\alpha_{blind} = 1/2$.*

Proof. From Lemma 1, in the criterion for blinding $\mathcal{A}$, viz. $\alpha = 1/(2p)$, clearly, α is minimized when p attains its maximum value of 1, which leads to $\alpha_{blind} = 1/2$. □

A. Optimal Falsification Strategy when $\mathcal{A}$ cannot be Blinded

We now characterize the optimal flipping probability p that should be employed by $\mathcal{D}$ to maximally degrade $\mathcal{A}$'s decision-making capability under use of the optimal decision rule in (6) when $\alpha < 1/2$, i.e., when $\mathcal{A}$ cannot be blinded. Analytical characterization of the error probability of $\mathcal{A}$'s decision rule in (6), however, to perform such an analysis is mathematically intractable. Hence, we find p that optimizes a surrogate function in lieu of the error probability of the decision rule in (6). Specifically, to define our surrogate function, let us first define

$$\Delta_i^X = \{\mathbb{E}^R[u_i^X] - \mathbb{E}^F[u_i^X]\}^2 \tag{13}$$

where

$$\mathbb{E}^R[u_i^X] = 0 \cdot P_i^R(u_i^X = 0) + 1 \cdot P_i^R(u_i^X = 1) \tag{14a}$$

$$\mathbb{E}^F[u_i^X] = 0 \cdot P_i^F(u_i^X = 0) + 1 \cdot P_i^F(u_i^X = 1) \tag{14b}$$

are the expectations of u_i^X under X being a real and fake server, respectively, $i \in \{1, \cdots, M\}$, $X \in \{G, H\}$. Using (13), we define our surrogate function as

$$\Delta = \sum_{X\in\{G,H\}} \sum_{i\in\{1,\cdots,M\}} \Delta_i^X \tag{15}$$

Using approaches similar to ones in [19], [27], it can be shown that the error probability of the decision rule in (6) monotonically increases as Δ (15) decreases. Thus, $\mathcal{D}$ would want to employ p that solves the following optimization problem:

$$\underset{p}{\text{minimize}} \qquad \Delta \tag{16a}$$

$$\text{Subj. to} \quad 0 \le p \le 1 \tag{16b}$$

Next, we characterize the optimal p that solves (16).

Lemma 2. *The optimal flipping probability that $\mathcal{D}$ should use to maximally degrade $\mathcal{A}$'s decision-making capability when $\alpha < 1/2$ is $p = 1$.*

Proof. Substituting (4a) and (4b) into (14), and subsequently substituting the simplified expressions of $\mathbb{E}^R[u_i^X]$ and $\mathbb{E}^F[u_i^X]$ that are yielded into (13), we get

$$\Delta_i^X = \big\{1 - 2\alpha p\big\}^2\{P_i^F(f_i^X = 1) - P_i^R(f_i^X = 1)\}^2 \tag{17}$$

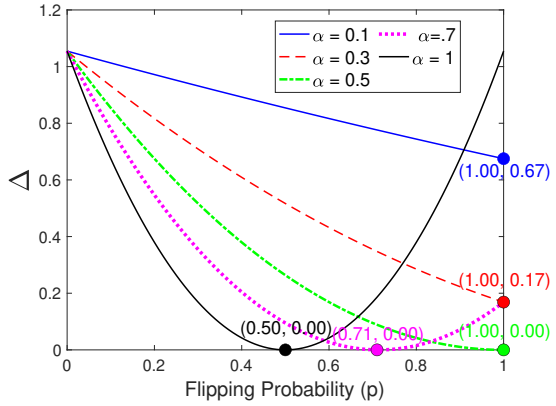
Substituting (17) into (15), we get

$$\Delta = \big\{1{-}2\alpha p\big\}^2 \sum_{X\in\{G,H\}} \sum_{i=1}^{M} \{P_i^F(f_i^X{=}1) - P_i^R(f_i^X{=}1)\}^2 \tag{18}$$

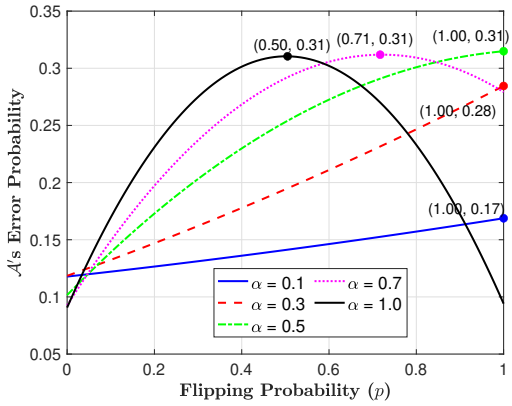
Denote $\gamma(p) = \{1 - 2\alpha p\}^2$, which is the only term that is a function of p in (18). Since $\frac{d^2}{dp^2}\gamma(p) = 8\alpha^2 \ge 0$, $\gamma(p)$ is a convex function of p. Further, since $\frac{d}{dp}\gamma(p) = 4\alpha(2\alpha p - 1)$, we have $\frac{d}{dp}\gamma(p) < 0$ when $p < \frac{1}{2\alpha}$. Thus, when $\alpha < 1/2$, in which case $\frac{d}{dp}\gamma(p) < 0$ for $0 \le p \le 1$, $p = 1$ minimizes (18)[1]. This proves the lemma. □

Next, in Fig. 1(a) and Fig. 1(b), we plot Δ (18) and the error probability of (6), respectively, versus p for different values of α. For the figures, we consider servers G and H to have $M = 3$ features, $P_G = 0.31$, $P_H = 1 - P_G = 0.69$, $[P_1^R(f_1^G = 1), P_2^R(f_2^G = 1), P_3^R(f_3^G = 1)] = [0.8, 0.58, 0.55]$, $[P_1^F(f_1^G = 1), P_2^F(f_2^G = 1), P_3^F(f_3^G = 1)] = [0.3, 0.25, 0.2]$, $[P_1^R(f_1^H = 1), P_2^R(f_2^H = 1), P_3^R(f_3^H = 1)] = [0.4, 0.67, 0.31]$, and $[P_1^F(f_1^H = 1), P_2^F(f_2^H = 1), P_3^F(f_3^H = 1)] = [0.9, 0.23, 0.67]$. The error probability of (6) was found via Monte Carlo simulations. As can be seen from Fig. 1(a), when $\alpha \ge 0.5$, there always exists a value of p that makes $\Delta = 0$, i.e., *blinds* $\mathcal{A}$, with optimal p in such a scenario ranging from 0.5 (when

[1]Consistent with Lemma 1, note that $\frac{d}{dp}\gamma(p) = 0$ when $p = 1/(2\alpha)$, a condition which makes $\Delta = 0$, thereby blinding $\mathcal{A}$ and resulting in the maximum possible error probability of (6) to occur, but attaining which requires $\alpha \ge 1/2$.



(a) Nature of Δ (18) w.r.t flipping probability (p) for different α.



(b) Nature of the error probability of (6) w.r.t flipping probability (p) for different α.

Fig. 1. Nature of Δ (18) and the error probability of (6) w.r.t flipping probability (p) for different α.

$\alpha = 1$) to 1 (when $\alpha = 0.5$) following the blinding criterion in Lemma 1. In accordance, as can be seen from Fig. 1(b), when $\alpha \geq 0.5$, the error probability of (6) maximizes at $p = \frac{1}{2\alpha}$ as prescribed by Lemma 1 (specifically at $p = \frac{1}{2 \cdot 0.5} = 1$ when $\alpha = 0.5$, at $p = \frac{1}{2 \cdot 0.7} = 0.71$ when $\alpha = 0.7$, and at $p = \frac{1}{2 \cdot 1} = 0.5$ when $\alpha = 1$).

Further, when $\alpha < 1/2$ in Fig. 1(a), note that $\Delta > 0$ for any p (i.e., $\mathcal{A}$ cannot be blinded), and that Δ monotonically decreases with p attaining its minimum value at $p = 1$. In accordance, when $\alpha < 1/2$, it can be noted from Fig. 1(b) that the error probability of (6) monotonically increases with p and maximizes at $p = 1$. The above observations corroborate Lemma 1 and Lemma 2 while depicting the relationship between Δ (18) and the error probability of (6).

V. Game Theoretic Sludging for Cyber Deception

In this section, in addition to being capable of feeding $\mathcal{A}$ with potentially falsified feature values, we consider that $\mathcal{D}$ invests certain costs for defending the servers' features to make it harder for $\mathcal{A}$ to inspect (scan) them, thereby sludging the decision-making process of $\mathcal{A}$ (and degrading the quality of its decision regarding the servers' natures). Employment of such cost structures can correspond to $\mathcal{D}$'s use of a sophisticated scheme to encrypt packets of the server(s) (e.g., to hinder $\mathcal{A}$'s efforts of understanding OS type(s) by scanning and analyzing traffic [10]), and, again, to the use of specialized firewall packages (e.g., to impede $\mathcal{A}$'s efforts of identifying open ports in the server(s) [28]). In such a scenario, under cost budgets of $\mathcal{A}$ and $\mathcal{D}$, we investigate how $\mathcal{A}$ can optimally determine which features to inspect and how $\mathcal{D}$ can choose a cost structure to defend the servers' features so as to maximally sludge $\mathcal{A}$.

For notational simplicity, in this section, w.l.o.g, we drop the superscript 'X' that we have been using to associate features with servers and label the $2M$ features of the two servers taken together from 1 to N, where $N = 2M$. Now, consider that the strategy of $\mathcal{D}$ is defined by the vector $\mathbf{c} = [c_1, \cdots, c_N]$, where $c_i \geq 0$ is the cost that $\mathcal{D}$ invests to defend feature i such that $\sum_{i=1}^{N} c_i \leq C^{\mathcal{D}}$, where $C^{\mathcal{D}}$ is the cost budget of $\mathcal{D}$, with $\mathcal{A}$ subsequently incurring c_i if it were to inspect feature i. Further, consider that the strategy of $\mathcal{A}$ is defined by the vector $\mathbf{z} = [z_1, \cdots, z_N]$, where z_i is the probability with which $\mathcal{A}$ inspects feature i such that $\sum_{i=1}^{N} c_i z_i \leq C^{\mathcal{A}}$, with $C^{\mathcal{A}}$ being $\mathcal{A}$'s cost budget. Given $\mathbf{z}$, following (15), the quality of $\mathcal{A}$'s decision can be described by

$$\Delta(\mathbf{z}) = \sum_{i=1}^{N} \Delta_i z_i \tag{19}$$

where Δ_i, $i \in \{1, \cdots, N\}$, following (17), becomes

$$\Delta_i = \{1 - 2\alpha p\}^2 \{P_i^F(f_i = 1) - P_i^R(f_i = 1)\}^2 \tag{20}$$

where $P_i^R(f_i = 1)$ and $P_i^F(f_i = 1)$ are the probabilities of f_i being 1 under the server that feature i belongs to being real and fake, respectively. Now, to analyze sludging in a strategic context, we model the problem as a *leader-follower* game, where $\mathcal{D}$ acts as the *leader* by choosing $\mathbf{c}$ with a goal to maximally sludge the decision-making process of $\mathcal{A}$ so as to minimize (19) while knowing that $\mathcal{A}$, acting as the *follower*, would choose $\mathbf{z}$ to inspect those features that maximize (19) against the set cost structure. We model the associated optimizations from $\mathcal{D}$'s and $\mathcal{A}$'s perspectives as the following *bilevel optimization* problem.

$$\min_{\mathbf{c}} \quad \sum_{i=1}^{N} \Delta_i z_i^* \tag{21a}$$

$$\text{s.t.} \quad \sum_{i=1}^{N} c_i \leq C^{\mathcal{D}} \quad (\mathcal{D}\text{'s budget constr.}) \tag{21b}$$

$$\mathbf{z}^* = \underset{\mathbf{z}}{\text{argmax}} \sum_{i=1}^{N} \Delta_i z_i \tag{21c}$$

$$\text{s.t.} \sum_{i=1}^{N} c_i z_i \leq C^{\mathcal{A}} \ (\mathcal{A}\text{'s budget constr.}) \tag{21d}$$

Note that the bilevel optimization in (21) consists of an *upper-level* optimization in (21a)-(21b), which models $\mathcal{D}$'s optimization task, and a *lower-level* optimization in (21c)-(21d), which models $\mathcal{A}$'s optimization task. Clearly, if $C^{\mathcal{D}} \leq C^{\mathcal{A}}$, then the optimal solution to the lower-level problem in (21c)-(21d) would be $z_i^* = 1$, $\forall i \in \{1, \cdots, N\}$, regardless of

how $\mathcal{D}$ chooses $\mathbf{c}$. Thus, in the following, we consider the more challenging case of $C^{\mathcal{D}} > C^{\mathcal{A}}$. Also, note that it would be straightforward to show that in the optimal solution of the upper-level problem in (21a)-(21b), we would have $\sum_{i=1}^{N} c_i = C^{\mathcal{D}}$.

Now, it can be noted that $\mathcal{A}$'s optimization task, which corresponds to the *lower-level* problem, can be treated as a continuous Knapsack problem [29], whose optimal solution, for a given $\mathbf{c}$, can be found using the following theorem.

THEOREM 2 ([29]). *Suppose that the N features are labeled such that*

$$\frac{\Delta_1}{c_1} \geq \frac{\Delta_2}{c_2} \geq \cdots \geq \frac{\Delta_{N-1}}{c_{N-1}} \geq \frac{\Delta_N}{c_N} \tag{22}$$

Further, suppose that feature k is such that $k = \min\{n : \sum_{i=1}^{n} c_i > C^{\mathcal{A}}\}$. Then, optimal $\mathbf{z}^*$ that solves the continuous Knapsack problem in (21c)-(21d)*, for a fixed $\mathbf{c}$, is given by*

$$z_i^* = \begin{cases} 1 & \text{if } 1 \leq i < k \\ (C^{\mathcal{A}} - \sum_{i=1}^{k-1} c_i)\frac{1}{c_k} & \text{if } i = k \\ 0 & \text{if } k < i \leq N \end{cases} \tag{23}$$

Next, we characterize the optimal $\mathbf{c}$ that solves the *upper-level* problem in (21a)-(21b) from $\mathcal{D}$'s side. We first present an important characteristic that must hold for $\mathbf{c}$ to be optimal.

LEMMA 3. *For $\mathbf{c}^* = [c_1^*, \cdots, c_N^*]$ to form the optimal cost structure that solves the upper-level problem in* (21a)-(21b) *from $\mathcal{D}$'s perspective against $\mathcal{A}$ strategically solving the lower-level problem using Theorem 2, we must have*

$$\frac{\Delta_1}{c_1^*} = \frac{\Delta_2}{c_2^*} = \cdots = \frac{\Delta_{N-1}}{c_{N-1}^*} = \frac{\Delta_N}{c_N^*} \tag{24}$$

Proof. We prove the lemma by showing that any deviation from $\mathbf{c}^*$ prescribed by (24) against $\mathcal{A}$ optimally selecting features for inspection using Theorem 2 is detrimental for $\mathcal{D}$. First, note that $\Delta_i/c_i^* = \Delta_j/c_j^*$, $i, j \in \{1, \cdots, N\}$, $i \neq j$, implies that

$$c_i^* = \frac{\Delta_i}{\Delta_j} c_j^* \tag{25}$$

Now, since $C^{\mathcal{D}} > C^{\mathcal{A}}$, Theorem 2 suggests that for $\mathcal{A}$'s strategy $\mathbf{z}^* = [z_1^*, \cdots, z_N^*]$ to be optimal against $\mathbf{c}^*$, we must have $\sum_{i=1}^{N} c_i^* z_i^* = C^{\mathcal{A}}$, which implies that

$$\begin{aligned} \sum_{i=1}^{N} \Big(\frac{\Delta_i}{\Delta_j} c_j^*\Big) z_i^* &= C^{\mathcal{A}} \quad \text{(using (25))} \\ \implies \Delta_{\mathbf{c}^*} &= \frac{\Delta_j}{c_j^*} C^{\mathcal{A}} \end{aligned} \tag{26}$$

where $\Delta_{\mathbf{c}^*}$ $(= \sum_{i=1}^{N} \Delta_i z_i^*)$ is the maximum value of the objective function in (21c) yielded by optimal $\mathbf{z}^*$ that solves the lower-level problem against $\mathbf{c}^*$, and $j \in \{1, \cdots, N\}$.

Now, consider an arbitrary cost structure $\mathbf{c}'$ that has V features that have *lesser* costs, W features that have *equal* costs, and $N - V - W$ features that have *greater* costs, than their corresponding costs in $\mathbf{c}^*$. Also, w.l.o.g, consider that the features of the two servers are labeled such that $\mathbf{c}' = [(c_1^* - \delta_1), \cdots, (c_V^* - \delta_V), (c_{V+1}^*), \cdots, (c_{V+W}^*), (c_{V+W+1}^* + \delta_{V+W+1}), \cdots, (c_N^* + \delta_N)]$, which implies that

$$\begin{aligned} \frac{\Delta_1}{c_1^* - \delta_1} \geq \cdots \geq \frac{\Delta_V}{c_V^* - \delta_V} > \frac{\Delta_{V+1}}{c_{V+1}^*} = \cdots = \frac{\Delta_{V+W}}{c_{V+W}^*} \\ > \frac{\Delta_{V+W+1}}{c_{V+W+1}^* + \delta_{V+W+1}} \geq \cdots \geq \frac{\Delta_N}{c_N^* + \delta_N} \end{aligned} \tag{27}$$

where $\delta_i > 0$, $i \in \{1, \cdots, N\}$, is the amount of change of feature i's cost in $\mathbf{c}'$ from the one in $\mathbf{c}^*$. Note, we must have

$$\sum_{i=1}^{V} \delta_i = \sum_{i=V+W+1}^{N} \delta_i \tag{28}$$

to ensure preservation of $\mathcal{D}$'s cost budget. Now, for $\mathbf{z}' = [z_1', \cdots, z_N']$ to form $\mathcal{A}$'s optimal strategy against $\mathbf{c}'$ following Theorem 2, clearly, $\mathbf{z}'$ must satisfy

$$\sum_{i=1}^{V} z_i'(c_i^* - \delta_i) + \sum_{i=V+1}^{V+W} z_i' c_i^* + \sum_{i=V+W+1}^{N} z_i'(c_i^* + \delta_i) = C^{\mathcal{A}} \tag{29}$$

Using (25), for $j \in \{1, \cdots, N\}$, we can express (29) as

$$\sum_{i=1}^{V} z_i' \Big[\frac{\Delta_i}{\Delta_j} c_j^* - \delta_i\Big] + \sum_{i=V+1}^{V+W} z_i' \frac{\Delta_i}{\Delta_j} c_j^* + \sum_{i=V+W+1}^{N} z_i' \Big[\frac{\Delta_i}{\Delta_j} c_j^* + \delta_i\Big] = C^{\mathcal{A}}$$

which, after some simplifications, yields

$$\sum_{i=1}^{N} z_i' \Delta_i = \frac{\Delta_j}{c_j^*} C^{\mathcal{A}} + \frac{\Delta_j}{c_j^*} \Big[\sum_{i=1}^{V} z_i' \delta_i - \sum_{i=V+W+1}^{N} z_i' \delta_i\Big] \tag{30}$$

Using (26), (30) can be expressed as

$$\Delta_{\mathbf{c}'} = \Delta_{\mathbf{c}^*} + \frac{\Delta_j}{c_j^*} \Big[\sum_{i=1}^{V} z_i' \delta_i - \sum_{i=V+W+1}^{N} z_i' \delta_i\Big] \tag{31}$$

where $\Delta_{\mathbf{c}'}$ $(= \sum_{i=1}^{N} z_i' \Delta_i)$ is the maximum value of the objective function in (21c) yielded by optimal $\mathbf{z}'$ that solves the lower-level problem against $\mathbf{c}'$. Now, note that from (27), we have $\frac{\Delta_v}{c_v^* - \delta_v} > \frac{\Delta_x}{c_x^* + \delta_x}$, for any $v \in \{1, \cdots, V\}, x \in \{V + W + 1, \cdots, N\}$. This implies that, using Theorem 2, for $z_x' > 0$, $x \in \{V + W + 1, \cdots, N\}$, we must have $z_v = 1$, $\forall v \in \{1, \cdots, V\}$. Thus, in (31), the least value of the term $\sum_{i=1}^{V} z_i' \delta_i - \sum_{i=V+W+1}^{N} z_i' \delta_i$ is $\sum_{i=1}^{V} \delta_i - \sum_{i=V+W+1}^{N} \delta_i$, which, using (28), equals 0, implying that $\Delta_{\mathbf{c}'} \geq \Delta_{\mathbf{c}^*}$. In other words, any deviation from $\mathbf{c}^*$ can enhance the decision-making performance of $\mathcal{A}$. This proves the lemma. □

Next, we characterize the optimal cost structure that satisfies Lemma 3's condition to solve the upper-level problem.

THEOREM 3. *The optimal cost structure $\mathbf{c}^* = [c_1^*, \cdots, c_N^*]$ that solves the upper-level problem in* (21a)-(21b) *from $\mathcal{D}$'s perspective corresponds to, for any chosen $i \in \{1, \cdots, N\}$, $c_i^* = \frac{\Delta_i}{\sum_{j=1}^{N} \Delta_j} C^{\mathcal{D}}$ and $c_j^* = \frac{\Delta_j}{\Delta_i} c_i^*$, $\forall j \in \{1, \cdots, N\}$, $j \neq i$.*

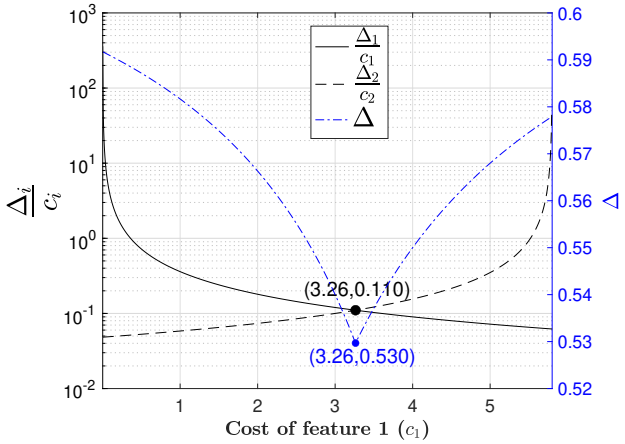


Fig. 2. Nature of Δ_i/c_i and Δ with varying costs of features for $N = 2$.

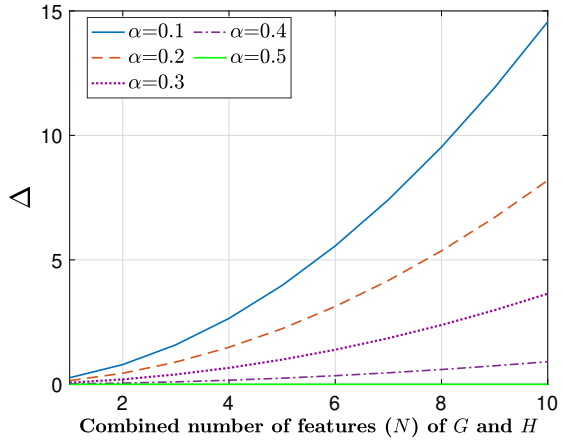


Fig. 3. Nature of Δ with increasing number of features.

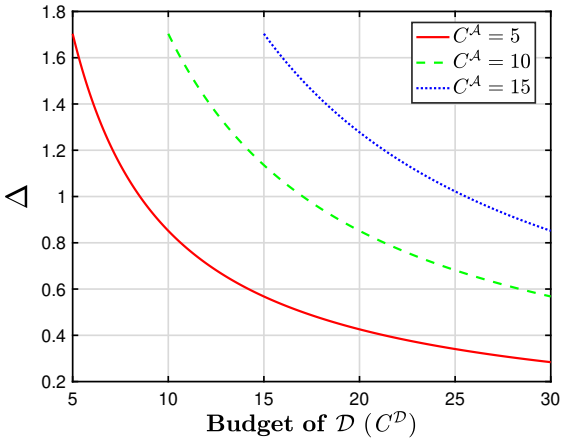


Fig. 4. Nature of Δ with varying cost budget ($C^{\mathcal{D}}$) of $\mathcal{D}$.

Proof. For optimal $\mathbf{c}^*$ to satisfy $\mathcal{D}$'s budget, for any chosen $i \in \{1, \cdots, N\}$, we must have $c_i^* + \sum_{j=1, j\neq i}^{N} c_j^* = C^{\mathcal{D}}$, in which, $\forall j \in \{1, \cdots, N\}, j \neq i$, setting $c_j^* = \frac{\Delta_j}{\Delta_i} c_i^*$ using (25) to make (24) hold true, we get

$$c_i^* + \sum_{j=1, j\neq i}^{N} \frac{\Delta_j}{\Delta_i} c_i^* = C^{\mathcal{D}} \tag{32}$$

Simplification of (32) yields $c_i^* = \frac{\Delta_i}{\sum_{j=1}^{N} \Delta_j} C^{\mathcal{D}}$, $i \in \{1, \cdots, N\}$, with $c_j^* = \frac{\Delta_j}{\Delta_i} c_i^*$ following (25), $\forall j \in \{1, \cdots, N\}$, $j \neq i$. This proves the theorem. □

Fig. 2 provides numerical results corroborating Lemma 3 and Theorem 3. For the figure, we consider $N = 2$, i.e., $M = 1$, with the feature of server G labeled as Feature 1 and that of H labeled as Feature 2 under $P_1^R(f_1 = 1) = 0.94$, $P_1^F(f_1 = 1) = 0.19$, $P_2^R(f_2 = 1) = 0.832$, $P_2^F(f_2 = 1) = 0.17$, $C^{\mathcal{D}} = 5.8$, $C^{\mathcal{A}} = 4.8$, $\alpha = 0.1$, and $p = 1$ (which is the optimal value of p since $\alpha < 0.5$ in the figure). The figure plots $\frac{\Delta_i}{c_i}$, $i \in \{1, 2\}$, as well as the objective function in (21c), viz. $\Delta = \sum_{i=1}^{N} \Delta_i z_i^*$, with z_i^* chosen using Theorem 2 to solve the lower-level problem in (21c)-(21d), versus c_1 for $0 \leq c_1 \leq C^{\mathcal{D}}$ (with $c_2 = C^{\mathcal{D}} - c_1$). It can be noted from the figure that the minimum value of Δ (21c) is achieved when $\Delta_1/c_1 = \Delta_2/c_2$, which corroborates Lemma 3, and that this occurs at $c_1 = 3.26$ (with $c_2 = C^{\mathcal{D}} - c_1 = 2.54$), which can be shown to tally with the solution prescribed by Theorem 3.

In Fig. 3, we plot (21a), i.e., $\Delta = \sum_{i=1}^{N} \Delta_i z_i^*$, corresponding to $\mathcal{D}$ and $\mathcal{A}$ optimally choosing $\mathbf{c}^*$ and $\mathbf{z}^*$ using Theorem 3 and Theorem 2, respectively, with varying N ($= 2M$), i.e., the combined number of features of servers G and H. For the figure, we consider, $\forall i \in \{1, \cdots, M\}$, $P_i^R(f_i^G = 1) = 0.8$, $P_i^F(f_i^G = 1) = 0.3$, $P_i^R(f_i^H = 1) = 0.4$, and $P_i^F(f_i^H = 1) = 0.9$, with $C^{\mathcal{D}} = 5.8$, $C^{\mathcal{A}} = 4.8$, and $p = 1$ (which is optimal since $\alpha \leq 0.5$ in the figure). As can be seen, Δ increases (i.e., $\mathcal{A}$'s decision-making quality gets enhanced) with N since, following Theorem 3, for a given $C^{\mathcal{D}}$, increase of N makes $\mathcal{D}$ to invest lesser cost in defending each feature which makes it easier for $\mathcal{A}$ to inspect more features to make more informed decisions. Further, as expected, for any given N, increase of α is beneficial for $\mathcal{D}$ as it enables it to perform more extensive flipping to deceive $\mathcal{A}$. Note that Δ always equals 0 when $\alpha = 0.5$ (under $p = 1$) since $\mathcal{A}$ is blind.

In Fig. 4, we plot (21a) under $\mathcal{D}$ and $\mathcal{A}$ optimally choosing $\mathbf{c}^*$ and $\mathbf{z}^*$ using Theorem 3 and Theorem 2, respectively, with varying cost budget ($C^{\mathcal{D}}$) of $\mathcal{D}$. For the figure, we consider $N = 10$, i.e., servers G and H to each have $M = 5$ features, with $[P_1^R(f_1^G = 1), P_2^R(f_2^G = 1), P_3^R(f_3^G = 1), P_4^R(f_4^G = 1), P_5^R(f_5^G = 1)] = [0.1, 0.3, 0.5, 0.56, 0.8]$, $[P_1^F(f_1^G = 1), P_2^F(f_2^G = 1), P_3^F(f_3^G = 1), P_4^F(f_4^G = 1), P_5^F(f_5^G = 1)] = [0.88, 0.86, 0.79, 0.12, 0.3]$, $[P_1^R(f_1^H = 1), P_2^R(f_2^H = 1), P_3^R(f_3^H = 1), P_4^R(f_4^H = 1), P_5^R(f_5^H = 1)] = [0.18, 0.89, 0.55, 0.4, 0.78]$, $[P_1^F(f_1^H = 1), P_2^F(f_2^H = 1), P_3^F(f_3^H = 1), P_4^F(f_4^H = 1), P_5^F(f_5^H = 1)] = [0.8, 0.28, 0.66, 0.9, 0.34]$, $\alpha = 0.1$, and $p = 1$ (which is optimal since $\alpha < 0.5$ in the figure). As can be seen, for any given $C^{\mathcal{A}}$, Δ decreases (i.e., $\mathcal{A}$'s decision-making quality degrades) with $C^{\mathcal{D}}$ since, following Theorem 3, for a given N, $\mathcal{D}$ can invest a higher cost in defending each feature as $C^{\mathcal{D}}$

increases. This makes it harder for $\mathcal{A}$ to inspect the features and increasingly sludges its decision-making process, resulting in $\mathcal{A}$ to make poorer quality decisions. Further, as expected, for any given $C^{\mathcal{D}}$, Δ increases with $C^{\mathcal{A}}$ since, following Theorem 2, $\mathcal{A}$ can inspect more features as $C^{\mathcal{A}}$ increases, enabling it to make more informed decisions.

Before concluding, for completeness, we make a remark regarding $\mathcal{A}$'s optimal decision rule when it inspects a subset of G's and H's features, as was considered in this section.

REMARK 1. *Following a similar procedure as used to derive* (6)*, it can be shown that $\mathcal{A}$'s optimal decision rule when it inspects a subset of G's and H's features is given by*

$$\sum_{i=1}^{M} \left\{ I_i^G \log \frac{P_i^F(u_i^G)}{P_i^R(u_i^G)} + I_i^H \log \frac{P_i^R(u_i^H)}{P_i^F(u_i^H)} \right\} \underset{G \text{ is real}}{\overset{H \text{ is real}}{\gtrless}} \log \frac{P_G}{P_H} \tag{33}$$

where I_i^X is a Boolean random variable such that $I_i^X = 1$ (denoting that $\mathcal{A}$ chooses to inspect feature i of server X) and $I_i^X = 0$ (denoting otherwise), with the value assumed by I_i^X governed by the probabilities in $\mathbf{z}$, $i \in \{1, \cdots, M\}$, $X \in \{G, H\}$.

VI. CONCLUSION

This paper presented a novel hypothesis testing framework that models an attacker's process of deciding the nature of servers in a network based on information that it gathers regarding their features during reconnaissance. The paper characterized the optimal decision rule that the attacker should use to process its gathered information for deciding the servers' natures. The paper also characterized the optimal information falsification strategy that the defender should use to minimize performance of the attacker's optimal decision rule, including characterization of the criterion that must be satisfied to blind the attacker. Further, under cost budgets, the paper characterized the optimal information acquisition costs that the defender can impose on the attacker to strategically sludge its decision-making process for maximally degrading the quality of its taken decisions.

In the future, we plan to build on our hypothesis testing-based cyber deception framework to make it adapt with possible cognitive biases of the defender and attacker.

REFERENCES

[1] S. Nan and S. Brahma, "Cyber deception under strategic and irrationality considerations," in *2023 IEEE 34th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2023, pp. 1–6.

[2] 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 Communications Surveys Tutorials*, vol. 23, no. 4, pp. 2460–2493, 2021.

[3] M. A. Sayed, A. H. Anwar, C. Kiekintveld, and C. Kamhoua, "Honeypot allocation for cyber deception in dynamic tactical networks: A game theoretic approach," in *Decision and Game Theory for Security*. Cham: Springer Nature Switzerland, 2023, pp. 195–214.

[4] S. Nan, S. Brahma, C. A. Kamhoua, and N. O. Leslie, "Mitigation of jamming attacks via deception," in *2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, 2020, pp. 1–6.

[5] A. Javadpour, F. Ja'fari, T. Taleb, M. Shojafar, and C. Benzaïd, "A comprehensive survey on cyber deception techniques to improve honeypot performance," *Computers Security*, vol. 140, p. 103792, 2024.

[6] T. Bao, M. Tambe, and C. Wang, *Cyber Deception: Techniques, Strategies, and Human Aspects*. Springer, 2023.

[7] C. Wang and Z. Lu, "Cyber deception: Overview and the road ahead," *IEEE Security Privacy*, vol. 16, no. 2, pp. 80–85, 2018.

[8] M. D. E. Bou-Harb and C. Assi, "Cyber scanning: A comprehensive survey," *IEEE Communications Surveys Tutorials*, vol. 16, no. 3, pp. 1496–1519, 2014.

[9] G. Fyodor Lyon, "Nmap network scanning: The official nmap project guide to network discovery and security scanning," *Insecure.com*, 2009.

[10] M. Laštovička, M. Husák, P. Velan, T. Jirsík, and P. Čeleda, "Passive operating system fingerprinting revisited: Evaluation and current challenges," *Computer Networks*, vol. 229, 2023.

[11] M. Zalewski, "p0f v3," https://lcamtuf.coredump.cx/p0f3/.

[12] D. Stuttard and M. Pinto, *The Web Application Hacker's Handbook: Discovering and Exploiting Security Flaws*. Wiley, 2011.

[13] J. P. S. Medeiros, A. M. Brito, and P. S. M. Pires, "A data mining based analysis of nmap operating system fingerprint database," in *Computational Intelligence in Security for Information Systems*, Á. Herrero, P. Gastaldo, R. Zunino, and E. Corchado, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 1–8.

[14] A. Schlenker, O. Thakoor, H. Xu, F. Fang, M. Tambe, L. Tran-Thanh, P. Vayanos, and Y. Vorobeychik, "Deceiving cyber adversaries: A game theoretic approach," in *AAMAS '18*, p. 892–900.

[15] S. Jajodia, N. Park, F. Pierazzi, A. Pugliese, E. Serra, G. I. Simari, and V. S. Subrahmanian, "A probabilistic logic of cyber deception," *IEEE Trans. Inf. Forensics Secur.*, vol. 12, no. 11, pp. 2532–2544, 2017.

[16] Z. R. Shi, A. D. Procaccia, K. S. Chan, S. Venkatesan, N. Ben-Asher, N. O. Leslie, C. Kamhoua, and F. Fang, "Learning and planning in the feature deception problem," in *Decision and Game Theory for Security*. Cham: Springer International Publishing, 2020, pp. 23–44.

[17] M. Albanese, E. Battista, and S. Jajodia, "A deception based approach for defeating os and service fingerprinting," in *2015 IEEE Conference on Communications and Network Security (CNS)*, 2015, pp. 317–325.

[18] S. Ha, G. Smith, and R. Starr, "Thwarting adversarial network reconnaissance through vulnerability scan denial and deception with data plane programming and p4," in *MILCOM 2023 - 2023 IEEE Military Communications Conference (MILCOM)*, 2023, pp. 793–798.

[19] P. K. Varshney, *Distributed Detection and Data Fusion*. NewYork: Springer-Verlag, 1997.

[20] R. Thaler and C. Sunstein, *Nudge : Improving decisions about health, wealth, and happiness*. Yale University Press, 2008.

[21] J. Franco, A. Aris, B. Canberk, and A. S. Uluagac, "A survey of honeypots and honeynets for internet of things, industrial internet of things, and cyber-physical systems," *IEEE Communications Surveys Tutorials*, vol. 23, no. 4, pp. 2351–2383, 2021.

[22] Q. D. La, T. Q. S. Quek, J. Lee, S. Jin, and H. Zhu, "Deceptive attack and defense game in honeypot-enabled networks for the internet of things," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 1025–1035, 2016.

[23] M. Osman, T. Nadeem, A. Hemida, and C. Kamhoua, "Optimizing honeypot placement strategies with graph neural networks for enhanced resilience via cyber deception," in *Proceedings of the 2nd on Graph Neural Networking Workshop 2023*, ser. GNNet '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 37–43. [Online]. Available: https://doi.org/10.1145/3630049.3630169

[24] O. Eriksson, "An evaluation of honeypots with compliant kubernetes," *Dissertation*, 2023.

[25] C. R. Sunstein, *Sludge: What Stops Us from Getting Things Done and What to Do about It*. The MIT Press, 09 2021. [Online]. Available: https://doi.org/10.7551/mitpress/13859.001.0001

[26] J. Dykstra, K. Shortridge, J. Met, and D. Hough, "Sludge for good: Slowing and imposing costs on cyber attackers," 2022. [Online]. Available: https://arxiv.org/abs/2211.16626

[27] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory*. Prentice Hall PTR, 1998.

[28] I. Pali and R. Amin, "Portsec: Securing port knocking system using sequence mechanism in sdn environment," in *2022 International Wireless Communications and Mobile Computing (IWCMC)*, 2022, pp. 1009–1014.

[29] G. B. Dantzig, "Discrete-variable extremum problems," *Operations Research*, vol. 5, no. 2, pp. 266–277, 1957.