Epistemic Signaling Games
for Cyber Deception with Asymmetric Recognition

Hampei Sasahara and Henrik Sandberg

Abstract— This study provides a model of cyber deception with asymmetric recognition represented by private beliefs. Signaling games, which are often used in existing works, are built on the implicit premise that the receiver's belief is public information. However, this assumption, which leads to symmetric recognition, is unrealistic in adversarial decision making. For a precise evaluation of risks arising from cognitive gaps, this paper proposes epistemic signaling games based on the Mertens-Zamir model, which explicitly quantifies players' asymmetric recognition. Equilibria of the games are analytically characterized with an interpretation.

I. INTRODUCTION

Cyber deception [1], [2], which can obscure important data such as customer information or system architecture, is an emerging defense technology. Examples of cyber deception include honeypots [3], moving target defense [4], and obfuscation [5]. Game theory offers mathematical models for strategic decision making [6]–[11]. In particular, signaling games are often used for representing asymmetric players' knowledge, which arises especially in cyber deception [12]–[14]. Signaling games are two-player games between a sender and a receiver, in which the sender's type is not known to the receiver. At the beginning of the game, the receiver forms her prior belief on the sender's type. Subsequently, the sender transmits her message, and the receiver updates her belief according to the message and chooses her action. Using signaling games, reasonable actions of the attacker and the defender can mathematically be represented as equilibria. The consequences of the game, such as the attack's impact and the deployed defense strategy's effectiveness, can be assessed in a quantitative manner by analyzing the equilibria.

An implicit assumption of traditional signaling games is that there exists a common prior, i.e., the receiver's prior belief is public information. In the context of cyber deception, this assumption means that the defender exactly knows what the attacker believes. Moreover, the attacker knows that the defender knows the attacker's belief. This process repeats indefinitely, and their mutual beliefs are shared. In this sense, the players' recognition is symmetric in traditional signaling games. However, this assumption is obviously unrealistic. As suggested by behavioral economics, recognition plays an important role in human's decision making [15], and we may
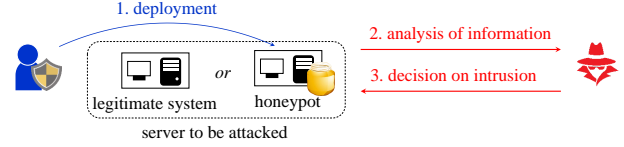
Fig. 1. Example of cyber deception: the defender may deploy a honeypot, spending cost for its disguise, while the attacker decides whether to execute an intrusion by analyzing the available information.

underestimate security risks without an adequate model that can describe *asymmetric recognition.*

This study proposes *epistemic signaling games* to resolve this issue. The problem of asymmetric recognition has been pointed out in the general context of economics, and the Mertens-Zamir model has been proposed to represent asymmetric recognition in epistemic game theory [16]–[18]. Using this model, we incorporate asymmetric recognition in signaling games. We characterize equilibria of the proposed epistemic signaling games with an interpretation.

This paper is organized as follows. In Sec. II, we briefly review traditional signaling games. Sec. III provides epistemic signaling games and analyze the equilibria. Finally, Sec. IV draws the conclusion. The appendix contains the proofs.

II. BRIEF REVIEW: TRADITIONAL SIGNALING GAMES

A. Example: Cyber Deception using a Honeypot

An example is unauthorized access to a workstation that may be a honeypot, which is a system placed on a network to attract the attention of attackers. A honeypot does not store any valuable data and collects information about the intruder's identity by alluring attackers [3]. Once an attacker compromises a honeypot, the defender analyzes the adversarial actions in detail and utilizes the information to improve the network protection. The adversarial decision making in the honeypot example is illustrated by Fig. 1. The defender may deploy a honeypot, while the attacker may be able to identify the system by analyzing information caused by unusual behavior [19], [20]. In this example, the sender is a model of the defender who decides whether to spend cost for disguising a honeypot, while the receiver is a model of the attacker who decides whether to execute an intrusion by analyzing the information.

B. Traditional Signaling Game Model

In signaling games, the sender's private information is referred to as its *type*, which is denoted by $\theta^{\mathrm{s}} \in \Theta^{\mathrm{s}}$. For simplicity, we assume the type to be binary, i.e., $\Theta^{\mathrm{s}} =$

$\{\theta_0^s, \theta_1^s\}$. For example, θ_0^s and θ_1^s represent the legitimate system and the honeypot, respectively. The true type is known to the sender but unknown to the receiver.

Given the type, the sender chooses a *message* $m \in \mathcal{M}$, which is assumed to be binary, i.e., $\mathcal{M} = \{m_0, m_1\}$. We refer to m_i as an honest message when the type is θ_i^s and the other as a deceptive message for any $i \in \{0, 1\}$. We consider *mixed strategies* and denote the sender's strategy by $\sigma^s \in \mathcal{S}^s$ such that $\sigma^s(m|\theta^s)$ gives the probability with which the sender sends the message m when her type is θ^s. We refer to $m \in \mathcal{M}$ as an *on-path message* when $\sigma^s(m|\theta^s) > 0$ for some $\theta^s \in \Theta^s$, and as an *off-path message* otherwise. In the honeypot example, m represents the service provided by the server. The defender's choice of m is a decision whether to disguise the system's behavior and to pay a cost, or not. An example of the disguise is to replace a cheap honeypot that provides only an open service port with a sophisticated one that provides full functional support [20].

After receiving the message, the receiver chooses an *action* $a \in \mathcal{A}$, which is assumed to be binary, i.e., $\mathcal{A} = \{a_0, a_1\}$. In the honeypot example, a_0 and a_1 represent execution of the intrusion and withdrawal, respectively. The receiver's mixed strategies are denoted by $\sigma^r \in \mathcal{S}^r$.

Traditional signaling games assume existence of a *common prior* on the type, i.e., the type is determined by nature according to a probability distribution over Θ^s, which is known to both players, at the beginning of the game. The determined type is informed to the sender and the receiver updates her belief based on the prior distribution and the transmitted message. The prior and posterior beliefs are denoted by $\pi^r(\theta^s)$ and $\pi^r(\theta^s|m)$, respectively. In the next section we will revisit and discuss this assumption on existence of a common prior, which is standard in Harsanyi's incomplete information games [18].

Let $u^s : \Theta^s \times \mathcal{M} \times \mathcal{A} \to \mathbb{R}$ denote a utility function for the sender. Similarly, let $u^r : \Theta^s \times \mathcal{A} \to \mathbb{R}$ denote a utility function for the receiver. Note that the receiver's utility is independent of the message because the messaging cost is not owned by the receiver. Throughout this paper, we assume that the utilities satisfy

$$\begin{cases} u^s(\theta_0^s, m, a_0) < u^s(\theta_0^s, m, a_1), \\ u^s(\theta_1^s, m, a_1) < u^s(\theta_1^s, m, a_0), \end{cases} \begin{cases} u^r(\theta_0^s, a_0) > u^r(\theta_0^s, a_1), \\ u^r(\theta_1^s, a_1) > u^r(\theta_1^s, a_0), \end{cases}$$

for any $m \in \mathcal{M}$. This assumption means that, the attacker prefers a_0 if $\theta^s = \theta_0^s$ and a_1 if $\theta^s = \theta_1^s$, and the defender prefers the opposite if the message is fixed. Moreover, we also assume that the sender's utility is symmetric with respect to her type for honest messaging, i.e., $u^s(\theta_0^s, m_0, a_0) = u^s(\theta_1^s, m_1, a_1)$ and $u^s(\theta_0^s, m_0, a_1) = u^s(\theta_1^s, m_1, a_0)$ to simplify the results. Without this assumption, similar results are available by dividing the cases.

We suppose that deceptive messaging requires a cost. The sender's utility function is assumed to be represented as

$$\begin{cases} u^s(\theta_0^s, m_1, a) = u^s(\theta_0^s, m_0, a) - C, \\ u^s(\theta_1^s, m_0, a) = u^s(\theta_1^s, m_1, a) - C, \end{cases} \forall a \in \mathcal{A}$$

where $C > 0$ is the deceptive messaging cost. In the honeypot example, C may represent the cost for disguising the hon-
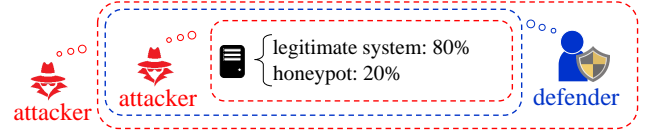
(a) Symmetric recognition in the honeypot example: The attacker believes that the system of interest is legitimate with probability 0.8 and it is a honeypot with probability 0.2. The defender knows the attacker's belief, and the attacker knows that the defender knows the attacker's belief. This is the model used in the traditional signaling games.
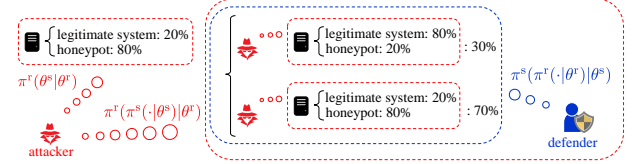
(b) Asymmetric recognition in the honeypot example: The attacker believes that the system of interest is legitimate with probability 0.2 and it is a honeypot with probability 0.8. On the other hand, the defender believes that the attacker believes that the system of interest is legitimate with probability 0.8 and it is a honeypot with probability 0.2 with probability 0.3 and that the attacker believes the system of interest is legitimate with probability 0.2 and it is a honeypot with probability 0.8 with probability 0.7. The attacker knows the defender's belief on the attacker's belief. This is the proposed model used in the epistemic signaling games.

Fig. 2. Symmetric and asymmetric recognitions in the honeypot example.

eypot as a legitimate system. This disguise can be achieved by replacing a cheap honeypot with a sophisticated, but expensive, one that can provide full functional support [20], for example. Signaling games without deceptive messaging cost are referred to as cheap-talk signaling games [12]. Our analysis can be extended to the cheap-talk case.

Reasonable strategies in signaling games are perfect Bayesian equilibria (PBE) where the players maximize their expected utilities and the receiver's belief is rationally updated according to Bayes' rule. By investigating resulting PBE, as studied in [12], [13], we can assess security risks in a quantitative manner.

III. EPISTEMIC SIGNALING GAMES WITH ASYMMETRIC RECOGNITION

A. Implicit Assumption on Recognition in Traditional Model

An important feature of traditional signaling games is the existence of a common prior, under which the players precisely know what the opponent believes. In other words, the existence of a common prior implicitly assumes *symmetric recognition*. However, in the security domain, the attacker's belief is not necessarily shared by the players, i.e., the attacker and the defender may possess *asymmetric recognition*. Figs. 2a and 2b illustrate symmetric and asymmetric recognitions in the honeypot example, respectively.

We briefly review the existing approach to resolving this issue used in the general context. The most fundamental notion is *belief hierarchy*, which has been introduced in epistemic game theory [16]–[18]. A belief hierarchy is formed as follows. Let $\Delta(\mathcal{X})$ denote the set of probability measures on $\mathcal{X}$. The first-order belief is given as $\pi_1 \in \Delta(\Theta^s)$, which describes the attacker's belief on the system architecture.

The second-order belief is given as $\pi_2 \in \Delta(\Delta(\Theta^{\mathrm{s}}))$, which describes the defender's belief on the attacker's first-order belief. In a similar manner, the belief at any level is given, and the tuple of the beliefs at all levels is referred to as a belief hierarchy.

To handle incomplete information games without common prior and the resulting belief hierarchy, the *Mertens-Zamir model* has been introduced [16]–[18]. The model considers *type structure*, in which a belief hierarchy is embedded. A type structure consists of players, sets of types, and beliefs. In particular, for signaling games, a type structure can be given by

$$\mathcal{T} = ((\mathrm{s},\mathrm{r}), (\Theta^{\mathrm{s}}, \Theta^{\mathrm{r}}), (\pi^{\mathrm{s}}, \pi^{\mathrm{r}})) \tag{1}$$

where (s,r) represents the sender and the receiver, Θ^{s} and Θ^{r} represent the sets of players' types, and $\pi^{\mathrm{s}} : \Theta^{\mathrm{r}} \times \Theta^{\mathrm{s}} \to [0,1]$ and $\pi^{\mathrm{r}} : \Theta^{\mathrm{s}} \times \Theta^{\mathrm{r}} \to [0,1]$ represent the beliefs. The value $\pi^{\mathrm{s}}(\theta^{\mathrm{r}}|\theta^{\mathrm{s}})$ denotes the sender's belief of the receiver's type θ^{r} when the sender's type is θ^{s}, and $\pi^{\mathrm{r}}(\theta^{\mathrm{s}}|\theta^{\mathrm{r}})$ denotes the corresponding receiver's belief. The first-order belief is given by $\pi_1(\theta^{\mathrm{s}}) = \pi^{\mathrm{r}}(\theta^{\mathrm{s}}|\theta^{\mathrm{r}})$ for the true receiver's type $\theta^{\mathrm{r}} \in \Theta^{\mathrm{r}}$, and the second-order belief is given by $\pi_2(\pi^{\mathrm{r}}(\cdot|\theta^{\mathrm{r}})|\theta^{\mathrm{s}}) = \pi^{\mathrm{s}}(\theta^{\mathrm{r}}|\theta^{\mathrm{s}})$ for the true sender's type $\theta^{\mathrm{s}} \in \Theta^{\mathrm{s}}$ as long as there is a one-to-one correspondence between the receiver's types and her beliefs. The higher-order beliefs are illustrated in Fig. 2b. In a similar manner, the belief at any level of the belief hierarchy can be derived from the type structure. An important fact is that, for any reasonable belief hierarchy there exists a type structure that can generate the belief hierarchy of interest [16]–[18]. In this sense, the Mertens-Zamir model has the sufficient capability of describing any situation with asymmetric recognition. For a formal discussion, see [17], [18].

B. Epistemic Signaling Games with Asymmetric Recognition

In this subsection, we propose *epistemic signaling games* using the Mertens-Zamir model for describing adversarial decision making with asymmetric recognition.

In the honeypot example, the sender's type set is given by $\Theta^{\mathrm{s}} = \Theta^{\mathrm{s}}_{\mathrm{sys}} \times \Theta^{\mathrm{s}}_{\mathrm{rec}}$ where $\Theta^{\mathrm{s}}_{\mathrm{sys}}$ and $\Theta^{\mathrm{s}}_{\mathrm{rec}}$ represent the sets of the system's attribute, namely, a legitimate system or a honeypot, and of the defender's recognition, respectively. For simplicity, we assume that $\Theta^{\mathrm{s}}_{\mathrm{rec}}$ is singleton, i.e., there is only one possible defender's recognition. This implies that the defender's recognition is public information. In this sense, this assumption corresponds to the worst case where the defender's belief is perfectly known to the attacker. Under this assumption, *the sender's belief is independent of her type*. We denote the sender's belief by $\pi^{\mathrm{s}}(\theta^{\mathrm{r}})$ instead of $\pi^{\mathrm{s}}(\theta^{\mathrm{r}}|\theta^{\mathrm{s}})$. We also assume Θ^{s} and Θ^{r} in (1) to be binary, i.e., $\Theta^{\mathrm{s}} := \{\theta^{\mathrm{s}}_0, \theta^{\mathrm{s}}_1\}$ and $\Theta^{\mathrm{r}} := \{\theta^{\mathrm{r}}_0, \theta^{\mathrm{r}}_1\}$.

The contrasting ingredients of the traditional and our epistemic signaling games are listed in Table I. Accordingly, the sender's expected utility becomes

$$\bar{u}^{\mathrm{s}}(\sigma^{\mathrm{s}}, \sigma^{\mathrm{r}}|\theta^{\mathrm{s}})$$
$$:= \sum_{a \in \mathcal{A}, m \in \mathcal{M}, \theta^{\mathrm{r}} \in \Theta^{\mathrm{r}}} \sigma^{\mathrm{r}}(a|m, \theta^{\mathrm{r}})\pi^{\mathrm{s}}(\theta^{\mathrm{r}})\sigma^{\mathrm{s}}(m|\theta^{\mathrm{s}})u^{\mathrm{s}}(\theta^{\mathrm{s}}, m, a),$$

	Traditional	Epistemic
receiver's strategy	$\sigma^{\mathrm{r}}(a|m)$	$\sigma^{\mathrm{r}}(a|m, \theta^{\mathrm{r}})$
sender's belief	none	$\pi^{\mathrm{s}}(\theta^{\mathrm{r}}|\theta^{\mathrm{s}})$
receiver's belief	$\pi^{\mathrm{r}}(\theta^{\mathrm{s}})$	$\pi^{\mathrm{r}}(\theta^{\mathrm{s}}|\theta^{\mathrm{r}})$

which depends on the sender's belief in contrast to the traditional one. Similarly, the receiver's expected utility is given by

$$\bar{u}^{\mathrm{r}}(\sigma^{\mathrm{r}}|m, \theta^{\mathrm{r}}) := \sum_{a \in \mathcal{A}, \theta^{\mathrm{s}} \in \Theta^{\mathrm{s}}} \sigma^{\mathrm{r}}(a|m, \theta^{\mathrm{r}})\pi^{\mathrm{r}}(\theta^{\mathrm{s}}|m, \theta^{\mathrm{r}})u^{\mathrm{r}}(\theta^{\mathrm{s}}, a),$$

where $\pi^{\mathrm{r}}(\theta^{\mathrm{s}}|m, \theta^{\mathrm{r}})$ denotes the posterior belief. For simplicity, we assume that the receivers with the type θ^{r}_0 and θ^{r}_1 prefer a_0 and a_1, respectively, when the sender's strategy is independent of her type, i.e., the message does not possess any information on sender's type. In a mathematical form, we assume $\bar{u}^{\mathrm{r}}(a_0|\theta^{\mathrm{r}}_0) > \bar{u}^{\mathrm{r}}(a_1|\theta^{\mathrm{r}}_0)$ and $\bar{u}^{\mathrm{r}}(a_0|\theta^{\mathrm{r}}_1) < \bar{u}^{\mathrm{r}}(a_1|\theta^{\mathrm{r}}_1)$ where $\bar{u}^{\mathrm{r}}(a|\theta^{\mathrm{r}}) := \sum_{\theta^{\mathrm{s}} \in \Theta^{\mathrm{s}}} \pi^{\mathrm{r}}(\theta^{\mathrm{s}}|\theta^{\mathrm{r}})u^{\mathrm{r}}(\theta^{\mathrm{s}}, a)$.

The solution concept is defined as follows.

Definition 1: A PBE of the epistemic signaling game is a strategy profile $(\sigma^{\mathrm{s}*}, \sigma^{\mathrm{r}*})$ and posterior belief $\pi^{\mathrm{r}}(\theta^{\mathrm{s}}|m, \theta^{\mathrm{r}})$ such that

$$\begin{cases} \sigma^{\mathrm{s}*} \in \arg\max_{\sigma^{\mathrm{s}} \in \mathcal{S}^{\mathrm{s}}} \bar{u}^{\mathrm{s}}(\sigma^{\mathrm{s}}, \sigma^{\mathrm{r}*}|\theta^{\mathrm{s}}), & \forall \theta^{\mathrm{s}} \in \Theta^{\mathrm{s}}, \\ \sigma^{\mathrm{r}*} \in \arg\max_{\sigma^{\mathrm{r}} \in \mathcal{S}^{\mathrm{r}}} \bar{u}^{\mathrm{r}}(\sigma^{\mathrm{r}}|m, \theta^{\mathrm{r}}), & \forall m \in \mathcal{M}, \forall \theta^{\mathrm{r}} \in \Theta^{\mathrm{r}} \end{cases}$$

and

$$\pi^{\mathrm{r}}(\theta^{\mathrm{s}}|m, \theta^{\mathrm{r}}) = \frac{\sigma^{\mathrm{s}*}(m|\theta^{\mathrm{s}})\pi^{\mathrm{r}}(\theta^{\mathrm{s}}|\theta^{\mathrm{r}})}{\sum_{\phi^{\mathrm{s}} \in \Theta^{\mathrm{s}}} \sigma^{\mathrm{s}*}(m|\phi^{\mathrm{s}})\pi^{\mathrm{r}}(\phi^{\mathrm{s}}|\theta^{\mathrm{r}})}$$

if $\sum_{\phi^{\mathrm{s}} \in \Theta^{\mathrm{s}}} \sigma^{\mathrm{s}*}(m|\phi^{\mathrm{s}})\pi^{\mathrm{r}}(\phi^{\mathrm{s}}|\theta^{\mathrm{r}}) \neq 0$.

There are three categories of PBE: *separating, pooling,* and *partially-separating* PBE in traditional signaling games [12], [13]. At separating PBE, the senders having different types transmit opposite messages, i.e., $\sigma^{\mathrm{s}}(m|\theta^{\mathrm{s}}_0) = 1 - \sigma^{\mathrm{s}}(m|\theta^{\mathrm{s}}_1) \in \{0, 1\}$ for any $m \in \mathcal{M}$. At pooling PBE, the sender's strategies are independent of the type, i.e., $\sigma^{\mathrm{s}}(m|\theta^{\mathrm{s}}_0) = \sigma^{\mathrm{s}}(m|\theta^{\mathrm{s}}_1)$ for any $m \in \mathcal{M}$. Otherwise, the PBE are referred to as partially-separating PBE. The separating and pooling PBE describe two extreme cases. In the honeypot example, the separating PBE mean that the legitimate system always provides full service while the honeypot always responds nothing. In contrast, the pooling PBE mean that the system provides the same service regardless of its attribute, although the honeypot may record the malicious intrusion.

C. Equilibrium Analysis of Epistemic Signaling Games

For the equilibrium analysis, we introduce some notation. Define

$$\begin{cases} \Delta u^{\mathrm{s}}_{10}(\theta^{\mathrm{s}}_0) := u^{\mathrm{s}}(\theta^{\mathrm{s}}_0, m_0, a_1) - u^{\mathrm{s}}(\theta^{\mathrm{s}}_0, m_0, a_0), \\ \Delta u^{\mathrm{s}}_{10}(\theta^{\mathrm{s}}_1) := u^{\mathrm{s}}(\theta^{\mathrm{s}}_1, m_1, a_1) - u^{\mathrm{s}}(\theta^{\mathrm{s}}_1, m_1, a_0), \end{cases}$$

which satisfy $\Delta u^{\mathrm{s}}_{10}(\theta^{\mathrm{s}}_1) < 0 < \Delta u^{\mathrm{s}}_{10}(\theta^{\mathrm{s}}_0)$. Note that $\Delta u^{\mathrm{s}}_{10}(\theta^{\mathrm{s}}_0) + \Delta u^{\mathrm{s}}_{10}(\theta^{\mathrm{s}}_1) = 0$ holds from the assumption on symmetry of the sender's utility. We define the constant $C' := C/\Delta u^{\mathrm{s}}_{10}(\theta^{\mathrm{s}}_0)$, which is the cost normalized

by $\Delta u_{10}^{\mathrm{s}}(\theta_0^{\mathrm{s}})$. Moreover, define $\Delta u_{10}^{\mathrm{r}}(\theta_0^{\mathrm{s}}) := u^{\mathrm{r}}(\theta_0^{\mathrm{s}}, a_1) - u^{\mathrm{r}}(\theta_0^{\mathrm{s}}, a_0)$, and $\Delta u_{10}^{\mathrm{r}}(\theta_1^{\mathrm{s}}) := u^{\mathrm{r}}(\theta_1^{\mathrm{s}}, a_1) - u^{\mathrm{r}}(\theta_1^{\mathrm{s}}, a_0)$, which satisfy $\Delta u_{10}^{\mathrm{r}}(\theta_0^{\mathrm{s}}) < 0 < \Delta u_{10}^{\mathrm{r}}(\theta_1^{\mathrm{s}})$.

We first characterize best responses to a given opponent's strategy.

Lemma 1: Define $\gamma^{\mathrm{s}} : \mathcal{S}^{\mathrm{r}} \to \mathbb{R}$ and $\gamma^{\mathrm{r}} : \mathcal{S}^{\mathrm{s}} \times \mathcal{M} \times \Theta^{\mathrm{r}} \to \mathbb{R}$ by

$$\gamma^{\mathrm{s}}(\sigma^{\mathrm{r}}) := \sum_{\theta^{\mathrm{r}} \in \Theta^{\mathrm{r}}} \Delta\sigma_{10}^{\mathrm{r}}(a_1 | \theta^{\mathrm{r}}) \pi^{\mathrm{s}}(\theta^{\mathrm{r}}),$$
$$\gamma^{\mathrm{r}}(\sigma^{\mathrm{s}}, m, \theta^{\mathrm{r}}) := \sum_{\theta^{\mathrm{s}} \in \Theta^{\mathrm{s}}} \sigma^{\mathrm{s}}(m | \theta^{\mathrm{s}}) \pi^{\mathrm{r}}(\theta^{\mathrm{s}} | \theta^{\mathrm{r}}) \Delta u_{10}^{\mathrm{r}}(\theta^{\mathrm{s}}),$$
$$\tag{2}$$

where $\Delta\sigma_{10}^{\mathrm{r}}(a | \theta^{\mathrm{r}}) := \sigma^{\mathrm{r}}(a | m_1, \theta^{\mathrm{r}}) - \sigma^{\mathrm{r}}(a | m_0, \theta^{\mathrm{r}})$. For a given receiver's strategy σ^{r}, the sender's best response $\sigma^{\mathrm{s}*}(m_0 | \theta_0^{\mathrm{s}})$ is given by

$$\sigma^{\mathrm{s}*}(m_0 | \theta_0^{\mathrm{s}}) = \begin{cases} 1 & \text{if } \gamma^{\mathrm{s}}(\sigma^{\mathrm{r}}) < C', \\ \alpha & \text{if } \gamma^{\mathrm{s}}(\sigma^{\mathrm{r}}) = C', \\ 0 & \text{otherwise}, \end{cases}$$

for any $\alpha \in [0, 1]$. For θ_1^{s}, the best response is given as the message opposite to the one by θ_0^{s} when $\gamma^{\mathrm{s}}(\sigma^{\mathrm{r}}) \neq C'$ and otherwise given as an arbitrary number in $[0, 1]$. For a given sender's strategy σ^{s}, the receiver's best response $\sigma^{\mathrm{r}*}(a_0 | m, \theta^{\mathrm{r}})$ is given by

$$\sigma^{\mathrm{r}*}(a_0 | m, \theta^{\mathrm{r}}) = \begin{cases} 1 & \text{if } \gamma^{\mathrm{r}}(\sigma^{\mathrm{s}}, m, \theta^{\mathrm{r}}) < 0, \\ \alpha & \text{if } \gamma^{\mathrm{r}}(\sigma^{\mathrm{s}}, m, \theta^{\mathrm{r}}) = 0, \\ 0 & \text{otherwise}, \end{cases}$$

when the message is on-path.

The separating PBE and pooling PBE are given as follows.

Theorem 1: If $C \geq \Delta u_{10}^{\mathrm{s}}(\theta_0^{\mathrm{s}})$, all PBE are separating PBE given by

$$\begin{cases} \sigma^{\mathrm{s}*}(m_0 | \theta_0^{\mathrm{s}}) = \sigma^{\mathrm{s}*}(m_1 | \theta_0^{\mathrm{s}}) = 1, \\ \sigma^{\mathrm{r}*}(a_0 | m_0, \theta^{\mathrm{r}}) = \sigma^{\mathrm{r}*}(a_1 | m_1, \theta^{\mathrm{r}}) = 1, \quad \forall \theta^{\mathrm{r}} \in \Theta^{\mathrm{r}} \end{cases} \tag{3}$$

with the posterior belief $\pi^{\mathrm{r}}(\theta_0^{\mathrm{s}} | m_0, \theta^{\mathrm{r}}) = \pi^{\mathrm{r}}(\theta_1^{\mathrm{s}} | m_1, \theta^{\mathrm{r}}) = 1$ for any $\theta^{\mathrm{r}} \in \Theta^{\mathrm{r}}$. If $C < \Delta u_{10}^{\mathrm{s}}(\theta_0^{\mathrm{s}})$ and $\pi^{\mathrm{s}}(\theta_0^{\mathrm{s}}) \geq C'$, there exist pooling PBE characterized by

$$\begin{cases} \sigma^{\mathrm{s}*}(m_0 | \theta^{\mathrm{s}}) = 1, \quad \forall \theta^{\mathrm{s}} \in \Theta^{\mathrm{s}}, \\ \begin{cases} \sigma^{\mathrm{r}*}(a_0 | m_0, \theta_0^{\mathrm{r}}) = 1, \quad \sigma^{\mathrm{r}*}(a_1 | m_0, \theta_1^{\mathrm{r}}) = 1, \\ \sigma^{\mathrm{r}*}(a_1 | m_1, \theta_0^{\mathrm{r}}) \pi^{\mathrm{s}}(\theta_0^{\mathrm{s}}) - \sigma^{\mathrm{r}*}(a_0 | m_1, \theta_1^{\mathrm{r}}) \pi^{\mathrm{s}}(\theta_1^{\mathrm{s}}) = C' \end{cases} \end{cases} \tag{4}$$

with a suitable off-path posterior belief. If $C < \Delta u_{10}^{\mathrm{s}}(\theta_0^{\mathrm{s}})$ and $\pi^{\mathrm{s}}(\theta_1^{\mathrm{r}}) \geq C'$, there exists PBE pooling at m_1, which can be characterized in a similar manner. Furthermore, if $\pi^{\mathrm{s}}(\theta_0^{\mathrm{s}}) < C'$ and $\pi^{\mathrm{s}}(\theta_1^{\mathrm{s}}) < C'$, then the game admits no pooling PBE.

In Theorem 1, the claim on separating PBE implies that the sender is always honest when the cost C is too high. The other claim on pooling PBE implies that giving no information can become a reasonable sender's strategy when the cost C is not too high.

Subsequently, we characterize partially-separating PBE. As a preparation, we state the following lemma.

Lemma 2: The function γ^{r} defined in (2) satisfies

$$\gamma^{\mathrm{r}}(\sigma^{\mathrm{s}}, m, \theta_0^{\mathrm{r}}) < \gamma^{\mathrm{r}}(\sigma^{\mathrm{s}}, m, \theta_1^{\mathrm{r}}), \quad \forall \sigma^{\mathrm{s}} \in \mathcal{S}^{\mathrm{s}}, \forall m \in \mathcal{M} \tag{5}$$

TABLE II
PARTIALLY-SEPARATING PBE OF EPISTEMIC SIGNALING GAMES

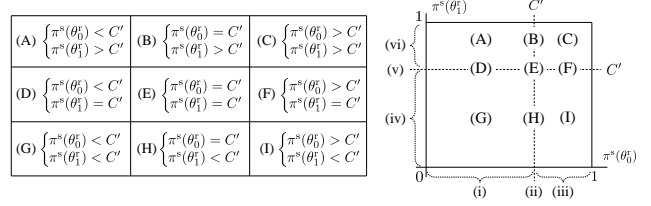| case (A): (i,vi)	case (B): (i,ii,iii,vi)	case (C): (iii,vi)
case (D): (i,iv,v,vi)	case (E): (i,ii,iii,iv,v,vi)	case (F): (iii,iv,v,vi)
case (G): (i,iv)	case (H): (i,ii,iii,iv)	case (I): (iii,iv)

Fig. 3. Regions of the sender's belief. The partially-separating PBE are determined from the region which the sender's belief belongs to.

and

$$\begin{cases} \gamma^{\mathrm{r}}(\sigma^{\mathrm{s}}, m_0, \theta_0^{\mathrm{r}}) + \gamma^{\mathrm{r}}(\sigma^{\mathrm{s}}, m_1, \theta_0^{\mathrm{r}}) < 0, \\ \gamma^{\mathrm{r}}(\sigma^{\mathrm{s}}, m_0, \theta_1^{\mathrm{r}}) + \gamma^{\mathrm{r}}(\sigma^{\mathrm{s}}, m_1, \theta_1^{\mathrm{r}}) > 0, \end{cases} \quad \forall \sigma^{\mathrm{s}} \in \mathcal{S}^{\mathrm{s}}. \tag{6}$$

Owing to Lemma 2, we can reduce the number of possible cases.

Lemma 3: The sender's optimal strategy $\sigma^{\mathrm{s}*} \in \mathcal{S}^{\mathrm{s}}$ at any partially-separating PBE satisfies $\gamma^{\mathrm{r}}(\sigma^{\mathrm{s}*}, m_0, \theta_0^{\mathrm{r}}) < 0 < \gamma^{\mathrm{r}}(\sigma^{\mathrm{s}*}, m_1, \theta_1^{\mathrm{r}})$.

Using Lemma 3, we can characterize the partially-separating PBE.

Theorem 2: Assume $C < \Delta u_{10}^{\mathrm{s}}(\theta_0^{\mathrm{s}})$. In epistemic signaling games, there exist partially-separating PBE independent of the sender's belief characterized by

$$\begin{cases} \sigma^{\mathrm{s}*} : \begin{cases} \gamma^{\mathrm{r}}(\sigma^{\mathrm{s}*}, m_0, \theta_0^{\mathrm{r}}) < 0, & \gamma^{\mathrm{r}}(\sigma^{\mathrm{s}*}, m_1, \theta_0^{\mathrm{r}}) = 0, \\ \gamma^{\mathrm{r}}(\sigma^{\mathrm{s}*}, m_0, \theta_1^{\mathrm{r}}) = 0, & \gamma^{\mathrm{r}}(\sigma^{\mathrm{s}*}, m_1, \theta_1^{\mathrm{r}}) > 0, \end{cases} \\ \sigma^{\mathrm{r}*} : \gamma^{\mathrm{s}}(\sigma^{\mathrm{r}*}) = C', \end{cases} \tag{7}$$

and the other partially-separating PBE dependent on the sender's belief are characterized by Table II, where the cases are specifically given in Fig. 3 and the equilibrium candidates are given in Table III.

An interpretation of the derived PBE can be given as follows. First, the receiver's strategy is determined from the condition $\gamma^{\mathrm{s}}(\sigma^{\mathrm{r}*}) = C'$ at any PBE, which means that the receiver always balances the sender's expected utilities corresponding to the messages m_0 and m_1 independently of her type. On the other hand, the sender's equilibrium strategy regions are illustrated in Fig. 4 where the solid line segments depict the conditions $\gamma^{\mathrm{s}}(\sigma^{\mathrm{s}*}, m_1, \theta_0^{\mathrm{s}}) = 0$ and $\gamma^{\mathrm{s}}(\sigma^{\mathrm{s}*}, m_0, \theta_1^{\mathrm{s}}) = 0$. The endpoints $(1, 0)$ and $(0, 1)$ correspond to the pooling PBE, and the intersection corresponds to the PBE independent of the sender's belief given by (7). It can be observed that the region at the PBE (i) and (vi) are connected to that at the PBE pooling at m_1. This PBE is taken when the sender believes that the receiver's type is θ_1^{r} with a high probability. In this sense, the PBE (i) and (vi) are reasonable consequences when the sender believes θ_1^{r}. Similarly, the PBE (iii) and (iv) can be reasonable when the sender believes θ_0^{r}.

TABLE III
EQUILIBRIUM CANDIDATES

(i)	$\sigma^{\mathrm{s}*}:$	$\begin{cases}\gamma^{\mathrm{r}}(\sigma^{\mathrm{s}*},m_0,\theta_0^{\mathrm{r}})<0, \gamma^{\mathrm{r}}(\sigma^{\mathrm{s}*},m_1,\theta_0^{\mathrm{r}})>0,\\ \gamma^{\mathrm{r}}(\sigma^{\mathrm{s}*},m_0,\theta_1^{\mathrm{r}})=0, \gamma^{\mathrm{r}}(\sigma^{\mathrm{s}*},m_1,\theta_1^{\mathrm{r}})>0,\end{cases}$	(ii)	$\sigma^{\mathrm{s}*}:$	$\begin{cases}\gamma^{\mathrm{r}}(\sigma^{\mathrm{s}*},m_0,\theta_0^{\mathrm{r}})<0, \gamma^{\mathrm{r}}(\sigma^{\mathrm{s}*},m_1,\theta_0^{\mathrm{r}})>0,\\ \gamma^{\mathrm{r}}(\sigma^{\mathrm{s}*},m,\theta_1^{\mathrm{r}})>0, \forall m\in\mathcal{M},\end{cases}$
	$\sigma^{\mathrm{r}*}:$	$\begin{cases}\sigma^{\mathrm{r}*}(a_0|m_0,\theta_0^{\mathrm{r}})=1, \sigma^{\mathrm{r}*}(a_1|m_1,\theta_0^{\mathrm{r}})=1,\\ \sigma^{\mathrm{r}*}(a_1|m_0,\theta_1^{\mathrm{r}})=C_1', \sigma^{\mathrm{r}*}(a_1|m_1,\theta_1^{\mathrm{r}})=1\end{cases}$		$\sigma^{\mathrm{r}*}:$	$\begin{cases}\sigma^{\mathrm{r}*}(a_0|m_0,\theta_0^{\mathrm{r}})=1, \sigma^{\mathrm{r}*}(a_1|m_1,\theta_0^{\mathrm{r}})=1,\\ \sigma^{\mathrm{r}*}(a_1|m,\theta_1^{\mathrm{r}})=1, \forall m\in\mathcal{M}\end{cases}$
(iii)	$\sigma^{\mathrm{s}*}:$	$\begin{cases}\gamma^{\mathrm{r}}(\sigma^{\mathrm{s}*},m_0,\theta_0^{\mathrm{r}})<0, \gamma^{\mathrm{r}}(\sigma^{\mathrm{s}*},m_1,\theta_0^{\mathrm{r}})=0,\\ \gamma^{\mathrm{r}}(\sigma^{\mathrm{s}*},m,\theta_1^{\mathrm{r}})>0, \forall m\in\mathcal{M},\end{cases}$	(iv)	$\sigma^{\mathrm{s}*}:$	$\begin{cases}\gamma^{\mathrm{r}}(\sigma^{\mathrm{s}*},m_0,\theta_0^{\mathrm{r}})<0, \gamma^{\mathrm{r}}(\sigma^{\mathrm{s}*},m_1,\theta_0^{\mathrm{r}})=0,\\ \gamma^{\mathrm{r}}(\sigma^{\mathrm{s}*},m_0,\theta_1^{\mathrm{r}})<0, \gamma^{\mathrm{r}}(\sigma^{\mathrm{s}*},m_1,\theta_1^{\mathrm{r}})>0,\end{cases}$
	$\sigma^{\mathrm{r}*}:$	$\begin{cases}\sigma^{\mathrm{r}*}(a_0|m_0,\theta_0^{\mathrm{r}})=1, \sigma^{\mathrm{r}*}(a_0|m_1,\theta_0^{\mathrm{r}})=C_3',\\ \sigma^{\mathrm{r}*}(a_1|m,\theta_1^{\mathrm{r}})=1, \forall m\in\mathcal{M}\end{cases}$		$\sigma^{\mathrm{r}*}:$	$\begin{cases}\sigma^{\mathrm{r}*}(a_0|m_0,\theta_0^{\mathrm{r}})=1, \sigma^{\mathrm{r}*}(a_0|m_1,\theta_0^{\mathrm{r}})=C_4',\\ \sigma^{\mathrm{r}*}(a_0|m_0,\theta_1^{\mathrm{r}})=1, \sigma^{\mathrm{r}*}(a_1|m_1,\theta_1^{\mathrm{r}})=1\end{cases}$
(v)	$\sigma^{\mathrm{s}*}:$	$\begin{cases}\gamma^{\mathrm{r}}(\sigma^{\mathrm{s}*},m,\theta_0^{\mathrm{r}})<0, \forall m\in\mathcal{M},\\ \gamma^{\mathrm{r}}(\sigma^{\mathrm{s}*},m_0,\theta_1^{\mathrm{r}})<0, \gamma^{\mathrm{r}}(\sigma^{\mathrm{s}*},m_1,\theta_1^{\mathrm{r}})>0,\end{cases}$	(vi)	$\sigma^{\mathrm{s}*}:$	$\begin{cases}\gamma^{\mathrm{r}}(\sigma^{\mathrm{s}*},m,\theta_0^{\mathrm{r}})<0, \forall m\in\mathcal{M},\\ \gamma^{\mathrm{r}}(\sigma^{\mathrm{s}*},m_0,\theta_1^{\mathrm{r}})=0, \gamma^{\mathrm{r}}(\sigma^{\mathrm{s}*},m_1,\theta_1^{\mathrm{r}})>0,\end{cases}$
	$\sigma^{\mathrm{r}*}:$	$\begin{cases}\sigma^{\mathrm{r}*}(a_0|m,\theta_0^{\mathrm{r}})=1, \forall m\in\mathcal{M},\\ \sigma^{\mathrm{r}*}(a_0|m_0,\theta_1^{\mathrm{r}})=1, \sigma^{\mathrm{r}*}(a_1|m_1,\theta_1^{\mathrm{r}})=1\end{cases}$		$\sigma^{\mathrm{r}*}:$	$\begin{cases}\sigma^{\mathrm{r}*}(a_0|m,\theta_0^{\mathrm{r}})=1, \forall m\in\mathcal{M},\\ \sigma^{\mathrm{r}*}(a_1|m_0,\theta_1^{\mathrm{r}})=C_6', \sigma^{\mathrm{r}*}(a_1|m_1,\theta_1^{\mathrm{r}})=1\end{cases}$
Constants: $C_1':=(1-C')/\pi^{\mathrm{s}}(\theta_1^{\mathrm{r}}), C_3':=1-C'/\pi^{\mathrm{s}}(\theta_0^{\mathrm{r}}), C_4':=(1-C')/\pi^{\mathrm{s}}(\theta_0^{\mathrm{r}}), C_6':=1-C'/\pi^{\mathrm{s}}(\theta_1^{\mathrm{r}})$					

Fig. 5 illustrates the transition of the sender's equilibrium strategy regions for increasing $\pi^{\mathrm{s}}(\theta_0^{\mathrm{r}})$. Starting with $\pi^{\mathrm{s}}(\theta_0^{\mathrm{r}})$ close to zero, the resulting PBE are (i) and (vi). When $\pi^{\mathrm{s}}(\theta_0^{\mathrm{r}}) = C'$ and $C' < 1/2$, the PBE (ii) and (iii) are additionally admitted. When $\pi^{\mathrm{s}}(\theta_0^{\mathrm{r}})$ slightly increases, the possible strategy region switches from the PBE (i) to (iii) through (ii). It can be observed that, for such a moderate belief, the sender takes both θ_0^{r} and θ_1^{r} into account. When $\pi^{\mathrm{s}}(\theta_0^{\mathrm{r}})$ increases more, the region switches again and reaches the PBE (iii) and (iv). A similar transition can be observed when $C' > 1/2$. Note that the separating PBE at the top right of Fig. 4 is the most honest strategy, and thus, the closer to the bottom left the strategy is, the more deceptive it is. In this sense, the strategies at the PBE (iii) can be regarded as more deceptive than those at (iv), although both are reasonable when the sender believes θ_0^{r}. Indeed, the PBE (iii) is taken when the deception cost is low. A similar interpretation is obtained for the PBE (i) and (vi). Finally, the PBE (ii) and (v) "bridge" the other PBE.

IV. CONCLUSION

This study proposes epistemic signaling games, a novel model of cyber deception with asymmetric recognition, based on the Mertens-Zamir model. The equilibria are analytically characterized.

APPENDIX

Proof of Lemma 1: With a slight abuse of notation, the difference between the sender's expected utilities using m_0 and m_1 is

$$\bar{u}^{\mathrm{s}}(m_0,\sigma^{\mathrm{r}}|\theta_0^{\mathrm{s}}) - \bar{u}^{\mathrm{s}}(m_1,\sigma^{\mathrm{r}}|\theta_0^{\mathrm{s}})$$
$$= -\sum_{a\in\mathcal{A},\theta^{\mathrm{r}}\in\Theta^{\mathrm{r}}}\pi^{\mathrm{s}}(\theta^{\mathrm{r}})u^{\mathrm{s}}(\theta_0^{\mathrm{s}},m_0,a)\Delta\sigma_{10}^{\mathrm{r}}(a|\theta^{\mathrm{r}}) + C$$
$$= -\gamma^{\mathrm{s}}(\sigma^{\mathrm{r}})\Delta u_{10}^{\mathrm{s}}(\theta_0^{\mathrm{s}}) + C,$$

which leads to the given best response for θ_0^{s}. Similarly, the same criterion can be obtained for θ_1^{s}.

For the receiver, with an on-path message m, we have

$$\bar{u}^{\mathrm{r}}(a_0|m,\theta^{\mathrm{r}}) - \bar{u}^{\mathrm{r}}(a_1|m,\theta^{\mathrm{r}})$$
$$= \sum_{\theta^{\mathrm{s}}\in\Theta^{\mathrm{s}}}\pi^{\mathrm{r}}(\theta^{\mathrm{s}}|m,\theta^{\mathrm{r}})\Delta u_{10}^{\mathrm{r}}(\theta^{\mathrm{s}})$$
$$= \gamma^{\mathrm{r}}(\sigma^{\mathrm{s}},m,\theta^{\mathrm{r}})/\sum_{\phi^{\mathrm{s}}\in\Theta^{\mathrm{s}}}\{\sigma^{\mathrm{s}}(m|\phi^{\mathrm{s}})\pi^{\mathrm{s}}(\phi^{\mathrm{s}}|\theta^{\mathrm{r}})\},$$
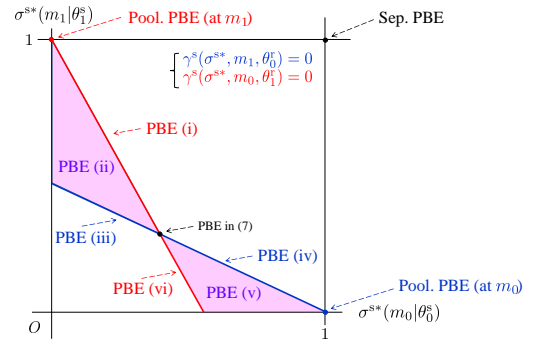

Fig. 4. Possible equilibrium strategies of the sender, as given by Theorem 2. The horizontal and vertical axes are the sender's mixed strategies of honest messaging for θ_0^{s} and θ_1^{s}, respectively. The points $(1,1)$, $(1,0)$, and $(0,1)$ are the separating PBE, and the PBE pooling at m_0 and m_1, respectively. The red and blue lines depict the regions of sender's strategies that satisfy $\gamma^{\mathrm{s}}(\sigma^{\mathrm{s}*},m_0,\theta_1^{\mathrm{r}})=0$ and $\gamma^{\mathrm{s}}(\sigma^{\mathrm{s}*},m_1,\theta_0^{\mathrm{r}})=0$, respectively. The intersection is associated with the PBE in (7). The upper left segment of the red line is associated with the PBE (i), and the lower right is associated with the PBE (vi). The upper left segment of the blue line is associated with the PBE (iii), and the lower right is associated with the PBE (iv).

which leads to the given best response. $\square$

Proof of Theorem 1: Assume $C \geq \Delta u_{10}^{\mathrm{s}}(\theta_0^{\mathrm{s}})$, i.e., $C' \geq 1$. Since $\gamma^{\mathrm{s}}(\sigma^{\mathrm{r}}) \leq 1$ for any $\sigma^{\mathrm{r}} \in \mathcal{S}^{\mathrm{r}}$ and $\theta^{\mathrm{s}} \in \Theta^{\mathrm{s}}$, sending an honest message is always optimal. Hence, the separating PBE is given by (3).

Next, assume $C < \Delta u_{10}^{\mathrm{s}}(\theta_0^{\mathrm{s}})$ and $\pi^{\mathrm{s}}(\theta_0^{\mathrm{r}}) \geq C'$. Take the sender's strategy $\sigma^{\mathrm{s}}(m_0|\theta^{\mathrm{s}}) = 1$ for any $\theta^{\mathrm{s}} \in \Theta^{\mathrm{s}}$. Then $\gamma^{\mathrm{r}}(\sigma^{\mathrm{s}},m_0,\theta_0^{\mathrm{r}}) < 0$ and $\gamma^{\mathrm{r}}(\sigma^{\mathrm{s}},m_0,\theta_1^{\mathrm{r}}) > 0$. Hence the receiver's best response to σ^{s} for m_0 is given by $\sigma^{\mathrm{r}*}$ in (4). Now σ^{s} becomes the best response to $\sigma^{\mathrm{r}*}$ if and only if $\gamma^{\mathrm{s}}(\sigma^{\mathrm{r}*}) = C'$, i.e., the off-path strategy satisfies the equation in (4). There exists an off-path strategy that satisfies the condition if and only if $\pi^{\mathrm{s}}(\theta_0^{\mathrm{r}}) \geq C'$. Therefore, the claim holds. For the case where $C < \Delta u_{10}^{\mathrm{s}}(\theta_0^{\mathrm{s}})$ and $\pi^{\mathrm{s}}(\theta_1^{\mathrm{r}}) \geq C'$, the pooling PBE can be derived in a similar manner. Finally, if both of $\pi^{\mathrm{s}}(\theta_0^{\mathrm{r}})$ and $\pi^{\mathrm{s}}(\theta_1^{\mathrm{r}})$ are less than C', then there exist no off-path strategies that satisfy the necessary equation in (4). Hence there exist no pooling PBE. $\square$

Proof of Lemma 2: We have $\Delta\gamma_{10}^{\mathrm{r}}(\sigma^{\mathrm{s}},m) =$
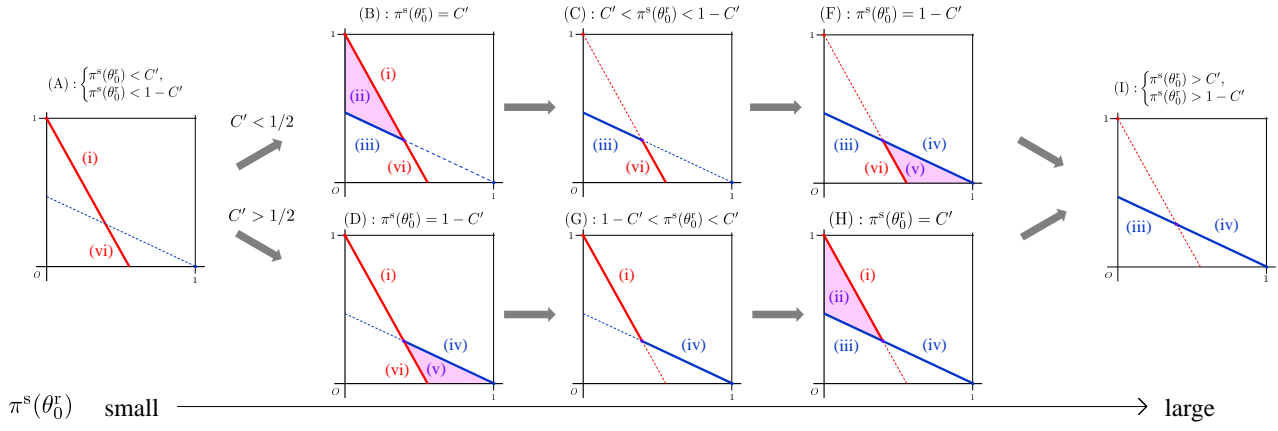
Fig. 5. Transition of the sender's equilibrium strategies for increasing $\pi^s(\theta_0^r)$. The sender's strategies are given by the region depicted by the red solid line when $\pi^s(\theta_0^r)$ is small. As $\pi^s(\theta_0^r)$ increases, the region shifts to the one depicted by the blue solid line.

$\sum_{\theta^s \in \Theta^s} \sigma^s(m|\theta^s) \Delta u_{10}^r(\theta^s) \Delta \pi_{10}^r(\theta^s)$ with $\Delta \gamma_{10}^r(\sigma^s, m) := \gamma^r(\sigma^s, m, \theta_1^r) - \gamma^r(\sigma^s, m, \theta_0^r)$ and $\Delta \pi_{10}^r(\theta^s) := \pi^r(\theta^s|\theta_1^r) - \pi^r(\theta^s|\theta_0^r)$. Since $\Delta \pi_{10}^r(\theta_1^s) = -\Delta \pi_{10}^r(\theta_0^s)$, we have

$$\Delta \gamma_{10}^r(\sigma^s, m)$$
$$= \{\underbrace{\sigma^s(m|\theta_1^s) \Delta u_{10}^r(\theta_1^s)}_{>0} - \underbrace{\sigma^s(m|\theta_0^s) \Delta u_{10}^r(\theta_0^s)}_{>0}\} \underbrace{\Delta \pi_{10}^r(\theta_1^s)}_{>0},$$

which leads to (5). Further, $\gamma^r(\sigma^s, m_0, \theta^r) + \gamma^r(\sigma^s, m_1, \theta^r) = \bar{u}^r(a_1|\theta^r) - \bar{u}^r(a_0|\theta^r)$, which leads to (6). $\square$

Proof of Lemma 3: Since there exists $m \in \mathcal{M}$ and $\theta^s \in \Theta^s$ such that $\sigma^{s*}(m|\theta^s) \in (0,1)$, we have $\gamma^s(\sigma^{r*}) = C'$. If $\gamma^r(\sigma^{s*}, m_0, \theta_0^r) \geq 0$, then $\gamma^r(\sigma^{s*}, m_1, \theta_0^r) < 0$ from (6). Thus $\sigma^{r*}(a_1|m_1, \theta_0^r) = 0$ and $\Delta\sigma_{10}^{r*}(a_1|\theta_0^r) \leq 0$. In addition, if $\gamma^r(\sigma^{s*}, m_0, \theta_0^r) \geq 0$, then $\gamma^r(\sigma^{s*}, m_0, \theta_1^r) > 0$ from (5). Thus $\sigma^{r*}(a_1|m_0, \theta_1^r) = 1$ and $\Delta\sigma_{10}^{r*}(a_1|\theta_1^r) \leq 0$. Therefore $\gamma^s(\sigma^{r*}) \leq 0 \neq C'$, which leads to a contradiction. The other claim can be proven in a similar manner. $\square$

Proof of Theorem 2: From Lemma 3, the possible combinations of the criterion for the best response are given by the nine cases: $\gamma^r(\sigma^{s*}, m_1, \theta_0^r) \gtreqless 0, \gamma^r(\sigma^{s*}, m_0, \theta_1^r) \gtreqless 0$ with $\gamma^s(\sigma^{r*}) = C'$, $\gamma^r(\sigma^{s*}, m_0, \theta_0^r) < 0$, $\gamma^r(\sigma^{s*}, m_1, \theta_1^r) > 0$.

We first show that the two cases

$$\begin{cases} \gamma^r(\sigma^{s*}, m_1, \theta_0^r) < 0 \\ \gamma^r(\sigma^{s*}, m_1, \theta_1^r) > 0 \end{cases}, \quad \begin{cases} \gamma^r(\sigma^{s*}, m_1, \theta_0^r) > 0 \\ \gamma^r(\sigma^{s*}, m_1, \theta_1^r) < 0 \end{cases}$$

do not happen. Assume that the former one holds. Then $\Delta\sigma_{10}^{r*}(a_1|\theta_0^r) = \Delta\sigma_{10}^{r*}(a_1|\theta_1^r) = 0$, and hence $\gamma^s(\sigma^{r*}) = 0 < C'$. Similarly, the latter one implies that $\gamma^s(\sigma^{r*}) = 1 > C'$. Those conditions lead to contradictions.

For the other cases, the PBE and their existence conditions are derived by a routine calculations. For example, for the PBE (7), it suffices to find σ^{s*} that satisfy the equations. This can be done using a standard linear algebra. $\square$

REFERENCES

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

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

[3] T. H. Project, *Know Your Enemy: Learning about Security Threats*. Addison-Wesley Professional, 2004.

[4] R. Zhuang, S. A. DeLoach, and X. Ou, "Towards a theory of moving target defense," in *Proc. ACM Workshop on Moving Target Defense*, 2014, p. 31–40.

[5] Q. Zhu, A. Clark, R. Poovendran, and T. Başar, "Deceptive routing games," in *Proc. 51st Conference on Decision and Control*, 2012, pp. 2704–2711.

[6] F. Farokhi, A. Teixeira, and C. Langbort, "Estimation with strategic sensors," *IEEE Trans. Autom. Control*, vol. 62, no. 2, pp. 724–739, 2017.

[7] S. Sarıtaş, S. Yüksel, and S. Gezici, "Quadratic multi-dimensional signaling games and affine equilibria," *IEEE Trans. Autom. Control*, vol. 62, no. 2, pp. 605–619, 2017.

[8] E. Miehling, R. Dong, C. Langbort, and T. Başar, "Strategic inference with a single private sample," in *Proc. 58th Conference on Decision and Control*, 2019, pp. 2188–2193.

[9] N. Heydaribeni and A. Anastasopoulos, "Linear equilibria for dynamic LQG games with asymmetric information and dependent types," in *Proc. 58th Conference on Decision and Control*, 2019, pp. 5971–5976.

[10] Y. Nugraha, A. Cetinkaya, T. Hayakawa, H. Ishii, and Q. Zhu, "Dynamic resilient network games considering connectivity," in *Proc. 59th Conference on Decision and Control*, 2020, pp. 3779–3784.

[11] M. Pirani, J. A. Taylor, and B. Sinopoli, "Strategic sensor placement on graphs," *Systems & Control Letters*, vol. 148, 2021.

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

[13] H. Çeker, J. Zhuang, S. Upadhyaya, Q. D. La, and B.-H. Soong, "Deception-based game theoretical approach to mitigate DoS attacks," in *Proc. Decision and Game Theory for Security*, 2016, pp. 18–38.

[14] J. Pawlick, E. Colbert, and Q. Zhu, "Modeling and analysis of leaky deception using signaling games with evidence," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 7, pp. 1871–1886, 2019.

[15] C. F. Camerer, T.-H. Ho, and J.-K. Chong, "A Cognitive Hierarchy Model of Games*," *The Quarterly Journal of Economics*, vol. 119, no. 3, pp. 861–898, 2004.

[16] J. Mertens and S. Zamir, "Formulation of Bayesian analysis for games with incomplete information," *International Journal of Game Theory*, vol. 14, pp. 1–29, 1985.

[17] E. Dekel and M. Siniscalchi, "Epistemic game theory," in *Handbook of Game Theory*. Elsevier, 2015, ch. 12, pp. 619–702.

[18] S. Zamir, "Bayesian games: Games with incomplete information," in *Encyclopedia of Complexity and Systems Science*. Springer, 2009, pp. 426–441.

[19] X. Fu, W. Yu, D. Cheng, X. Tan, K. Streff, and S. Graham, "On recognizing virtual honeypots and countermeasures," in *IEEE International Symposium on Dependable, Autonomic and Secure Computing*, 2006, pp. 211–218.

[20] N. Krawetz, "Anti-honeypot technology," *IEEE Security & Privacy*, vol. 2, no. 1, pp. 76–79, 2004.