2022 European Control Conference (ECC)
July 12-15, 2022. London, United Kingdom

Privacy Signaling Games with Binary Alphabets

Photios A. Stavrou[1], Serkan Sarıtaş[2] and Mikael Skoglund[3]

***Abstract*— In this paper, we consider a privacy signaling game problem for binary alphabets and single-bit transmission where a transmitter has a pair of messages, one of which is a casual message that needs to be conveyed, whereas the other message contains sensitive data and needs to be protected. The receiver wishes to estimate both messages to acquire as much information as possible. For this setup, we study the interactions between the transmitter and the receiver with non-aligned information-theoretic objectives (modeled by mutual information and hamming distance) due to the privacy concerns of the transmitter. We derive conditions under which Nash and/or Stackelberg equilibria exist and identify the optimal responses of the encoder and decoders strategies for each type of game. One particularly surprising result is that when both types of equilibria exist, they admit the same encoding and decoding strategies. We corroborate our analysis with simulation studies.**

I. Introduction

Decision-making is pivotal for a wide range of real-world applications such as social networks, networked control systems, smart grids, and recommendation systems. In these applications, usually, several customers (users) in a network may share extensive amounts of information with some service provider (i.e., utility company) because the latter wishes to know as much as possible about the service offered at the customer to improve the quality of service. However, this may come with a price as sometimes the users in the network may be prone to network-based attacks from malicious elements aiming to steal some sensitive information. Therefore, the users, in addition to the continuous improvement of the quality of service offered by a provider, wish to maintain a certain level of privacy. A type of privacy objective can be assumed when the information transmitted by some user to the service provider may be correlated with certain private information they want to protect. For example, in smart grids, the smart meter provides real-time information on energy supplies from the energy provider on the demands of the consumer (user), which can be utilized for unauthorized purposes, e.g., to infer the private information of the consumer, such as their habits and behaviors, see, e.g., [1], [2]. Identifying privacy-preserving mechanisms or approaches under various contexts related to information theory and control applications can be found in an anthology of papers, for instance, in [3]–[8].

A. Motivational Example

Consider the scenario illustrated in Fig.1. In that scenario, a smart house is illustrated in which a smart meter records the energy consumption and exchanges consumption data with energy suppliers, which can be used for monitoring and billing. Evidently, the existence of home residents and the electricity usage recorded by the smart meter at home are, in practice, correlated. Nevertheless, the presence of the house residents at home should be kept secret to possible outsiders (burglars, adversaries, etc.) whereas, at the same time, the energy consumption should be available to the electricity service providers. Therefore, the smart meter should be designed so that the service providers can access the electricity usage data, whereas the outsiders should not be able to deduce if the residents are home or not by checking the information from the smart meter.

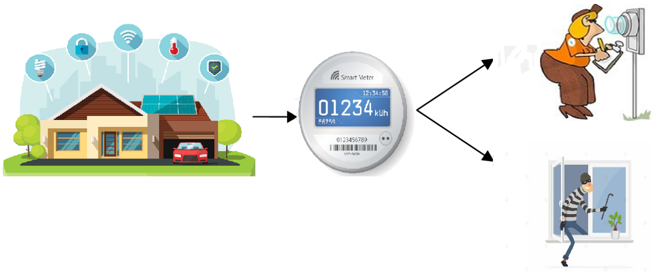


Fig. 1: Motivational example.

B. Literature review

The studies on cheap talk and signaling games were initiated by Crawford and Sobel in their seminal work [9], and found applications in various topics, e.g., in networked systems [10], [11], recommendation systems [12], [13], and economics [14], [15]. Starting with [16], there are many studies that consider the Stackelberg equilibrium of signaling games; an incomplete list includes [17]–[26] (see also the references therein). Many of these works assume that the non-alignment between the objective functions of the encoder and the decoder is a function of a Gaussian random variable (RV) correlated with the Gaussian source and hidden from the decoder (unlike the original case where it is fixed and commonly known by the encoder and the decoder [9], which is also studied in [17], [21], [24]). Nash and Stackelberg equilibria of signaling games are investigated in [25] when

[1] Photios A. Stavrou is with the Communication Systems Department at EURECOM, Campus SophiaTech, 06904, France. fotios.stavrou@eurecom.fr

[2] Serkan Sarıtaş is with the Department of Electrical and Electronics Engineering, Middle East Technical University, 06800, Ankara, Turkey. ssaritas@metu.edu.tr

[3] Mikael Skoglund is with the Division of Information Science and Engineering, KTH Royal Institute of Technology, SE-10044, Stockholm, Sweden. skoglund@kth.se

978-3-907144-07-7 ©2022 EUCA

there is a mismatch in priors of players. We refer to [15], [17], [21] for more discussion on the literature and some extensions (including Nash equilibrium analyses and multi-stage extensions) on cheap talk and signaling games.

In the context of strategic information transmission, several works consider the scenario where the sender takes the privacy of certain information into account by deploying a suitable privacy measure under either the Nash or Stackelberg criteria. For instance, in [27], a communication scenario between a sender and a receiver is investigated using the Stackelberg equilibrium. A family of nontrivial equilibria, in which the communicated messages carry information, is constructed, and its properties are studied. In [28], the authors study a Stackelberg game where the utility measure for the public parameter is quadratic and the privacy measure is entropy-based. Additional results therein include characterizations of the equilibrium under noisy and noiseless communication scenarios and analysis of the corresponding coding policies. In [29], the effect of privacy via a Nash game is studied between a sender and a receiver. As a measure of merit, the authors use mutual information between the private information and the communicated message to quantify the amount of the leaked information. For discrete RVs, they provide a numerical algorithm to find an equilibrium, whereas for Gaussian RVs, a bound on the estimation error is provided, and affine policies are shown to achieve this bound. In [30], a privacy-signaling game problem for the setup in Fig. 2 is considered in which a transmitter with privacy concerns observes a pair of correlated random vectors which are modeled as jointly Gaussian. Among other results, it is shown that a payoff dominant Nash equilibrium among all admissible policies is attained by a set of explicitly characterized linear policies and coincides with a Stackelberg equilibrium. Regarding the state of the art privacy metrics, we refer to [31] for a selection of over eighty privacy metrics and their categorization. Herein, we consider one of the discussed metrics therein, namely, Hamming distance.

In this paper, we consider the scenario that was first introduced in [30]. In this setup, a transmitter has a pair of messages, one of which is a casual message that needs to be conveyed, whereas the other message contains sensitive data and needs to be protected. On the other hand, the receiver wishes to estimate both messages with the aim of acquiring as much information as possible. For this setup, we study the interactions between the transmitter and the receiver whose objectives are not-aligned due to the privacy concerns of the transmitter in a game-theoretic framework. However, in contrast to [30] that deals with jointly Gaussian random vectors (of possibly different lengths) and linear policies, *here we deal with binary alphabets and consider different objectives for our single-bit transmitter and receiver.*

C. Contributions

The main contributions of this paper can be summarized as follows:

(i) We model a binary privacy signaling game, assuming a single-bit transmission, between an encoder and a decoder in which the encoder aims to hide one of two correlated binary RVs (i.e., private message) and to transmit the other (i.e., public message) while the decoder's goal is to learn about both of the RVs as much as possible. We use mutual information as a metric to measure the information exchange, a Hamming distortion to measure the level of privacy, and a weighting coefficient that determines the importance of privacy from the perspective of the encoder.

(ii) We characterize the objective functions of the encoder and the decoder in terms of priors and strategies (see Lemma 1), derive the best response of the encoder (resp. decoder) for a given decoder (resp. encoder) (see Lemmas 2 and 3). Then, using these best response maps, we characterize the Stackelberg and Nash equilibria (see Theorems 1 and 2, respectively).

(iii) We show that under certain conditions on the source prior probabilities, Nash and Stackelberg equilibria exist and coincide. Otherwise, there does not exist a Nash equilibrium, and the privacy coefficient only affects the Stackelberg equilibrium. In particular, for large privacy coefficients, the encoder may even prefer to hide information about the public message.

Due to space constraints, here we provide only the statements for lemmas and theorems; the proofs are available in [32].

II. Problem formulation and preliminaries

In this paper, we consider the scenario illustrated in Fig. 2 that was first introduced in [30]. We assume that the transmitter encodes a pair of correlated random variables $(\mathbf{x},\mathbf{y}) \in \mathcal{X} \times \mathcal{Y}$, $\mathcal{X} = \mathcal{Y} = \{0,1\}$ into $\mathbf{z} \in \mathcal{Z} = \{0,1\}$ using an encoding function denoted by $z = \gamma^e(x,y)$ and the receiver wants to decode both messages based on the observation $\mathbf{z} = z$. Note that the transmitter desires to transmit information about $\mathbf{y}$ and sees $\mathbf{x}$ as a private parameter that needs to be hidden from the receiver. In contrast, the receiver wants to accurately estimate both public and private messages given the observation $\mathbf{z} = z$. We denote the decoding functions for estimating $\mathbf{x}$ and $\mathbf{y}$ by $\hat{x} = \gamma^{d_\mathbf{x}}(z)$ and $\hat{y} = \gamma^{d_\mathbf{y}}(z)$, respectively.

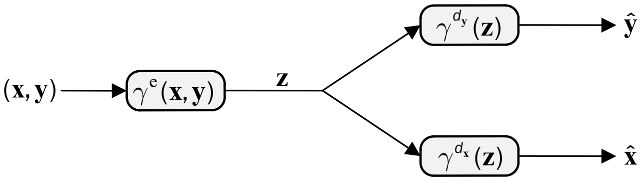

$(\mathbf{x},\mathbf{y})$
$\gamma^e(\mathbf{x},\mathbf{y})$
$\mathbf{z}$
$\gamma^{d_\mathbf{y}}(\mathbf{z})$
$\hat{\mathbf{y}}$
$\gamma^{d_\mathbf{x}}(\mathbf{z})$
$\hat{\mathbf{x}}$

Fig. 2: Our setup.

Since the transmitter needs to encode two messages generated by the joint distribution of $(\mathbf{x},\,\mathbf{y})$, i.e., $\mathbf{P}(x,y)$, it means that hiding $\mathbf{x}$ or transmitting $\mathbf{y}$ are somehow inter-dependent actions. Since our scenario is for binary alphabets, in the sequel, we will denote the joint distribution or probability mass function of $(\mathbf{x},\mathbf{y})$ to be given by the following column stochastic matrix:

$$\mathbf{P}(x,y)=\begin{bmatrix}\mathbf{P}(x=0,y=0)\\\mathbf{P}(x=0,y=1)\\\mathbf{P}(x=1,y=0)\\\mathbf{P}(x=1,y=1)\end{bmatrix}=\begin{bmatrix}a\\b\\c\\d\end{bmatrix}, \qquad (1)$$

where $d\triangleq 1-(a+b+c)$ with $(a,b,c,d)\in[0,1]\times[0,1]\times[0,1]\times[0,1]$ (also denoted for simplicity $[0,1]^4$). The objective of the transmitter is to maximize the public information $\mathbf{y}$ for the receiver and at the same time to hide as much as possible the sensitive information $\mathbf{x}$. These can be cast by the following objective function

$$J^e(\gamma^e,\gamma^{d_\mathbf{x}},\gamma^{d_\mathbf{y}})=I(\mathbf{y};\hat{\mathbf{y}})+\rho\mathbf{E}\{d_H(\mathbf{x},\hat{\mathbf{x}})\}, \qquad (2)$$

which is to be maximized by the encoder, where $I(\mathbf{y};\hat{\mathbf{y}})$ is the mutual information between $\mathbf{y}$ and $\hat{\mathbf{y}}$ [33], $\rho>0$ is a weighting coefficient that determines the level of desired privacy of $\mathbf{x}$, $\mathbf{E}\{d_H(\mathbf{x},\hat{\mathbf{x}})\}$ is some loss function which for this paper is assumed to be modeled by Hamming distortion, i.e.,

$$d_H(\mathbf{x},\hat{\mathbf{x}})=\begin{cases}1 & \mathbf{x}\neq\hat{\mathbf{x}}\\0 & \mathbf{x}=\hat{\mathbf{x}}\end{cases}, \qquad (3)$$

responsible to capture the privacy term $\mathbf{x}$. The objective of the receiver is to maximize the information of both public information $\mathbf{y}$ and sensitive information $\mathbf{x}$. This can be cast by the following objective function

$$J^d(\gamma^e,\gamma^{d_\mathbf{x}},\gamma^{d_\mathbf{y}})=I(\mathbf{y};\hat{\mathbf{y}})-\mathbf{E}\{d_H(\mathbf{x},\hat{\mathbf{x}})\}, \qquad (4)$$

which is to be maximized by the decoder. Since the costs of the encoder and the decoder are not aligned, the problem is studied under a game theoretic framework, and Stackelberg and Nash equilibria are investigated. In the Nash (simultaneous-move) game, the encoder and the decoder announce their strategies at the same time. More precisely, suppose that the set of possible strategies at the encoder is denoted by Γ^e and those at the decoders by $\Gamma^{d_\mathbf{y}}$ and $\Gamma^{d_\mathbf{x}}$, respectively, such that $\gamma^e\in\Gamma^e$, $\gamma^{d_\mathbf{y}}\in\Gamma^{d_\mathbf{y}}$, $\gamma^{d_\mathbf{x}}\in\Gamma^{d_\mathbf{x}}$. Then, a triplet of policies $(\gamma^{e,*},\gamma^{d_\mathbf{y},*},\gamma^{d_\mathbf{x},*})$ is said to be a *Nash equilibrium* [34] if

$$\begin{aligned}&J^e(\gamma^{e,*},\gamma^{d_\mathbf{y},*},\gamma^{d_\mathbf{x},*})\geq J^e(\gamma^e,\gamma^{d_\mathbf{y},*},\gamma^{d_\mathbf{x},*}),\quad\forall\gamma^e\in\Gamma^e,\\&J^d(\gamma^{e,*},\gamma^{d_\mathbf{y},*},\gamma^{d_\mathbf{x},*})\\&\qquad\geq J^d(\gamma^{e,*},\gamma^{d_\mathbf{y}},\gamma^{d_\mathbf{x}})\quad\forall\gamma^{d_\mathbf{y}}\in\Gamma^{d_\mathbf{y}},\gamma^{d_\mathbf{x}}\in\Gamma^{d_\mathbf{x}}.\end{aligned} \qquad (5)$$

As observed in (5), none of the players prefer to change their optimal strategies at the equilibrium, i.e., there is no profitable unilateral deviation from any of the players. In the Stackelberg game, the leader (encoder) commits to a particular policy and announces it to the follower (decoder). The decoder takes its optimal action upon observing the encoder's committed strategy. More precisely, a triplet of strategies $(\gamma^{e,*},\gamma^{d_\mathbf{y},*},\gamma^{d_\mathbf{x},*})$ is said to be a *Stackelberg equilibrium* [34] if

$$\begin{aligned}&J^e(\gamma^{e,*},\gamma^{d_\mathbf{y},*}(\gamma^{e,*}),\gamma^{d_\mathbf{x},*}(\gamma^{e,*}))\\&\qquad\geq J^e(\gamma^e,\gamma^{d_\mathbf{y},*}(\gamma^e),\gamma^{d_\mathbf{x},*}(\gamma^e)),\quad\forall\gamma^e\in\Gamma^e,\\&\text{where}(\gamma^{d_\mathbf{y},*}(\gamma^e),\gamma^{d_\mathbf{x},*}(\gamma^e))\text{ satisfy}\\&J^d(\gamma^e,\gamma^{d_\mathbf{y},*}(\gamma^e),\gamma^{d_\mathbf{x},*}(\gamma^e))\\&\quad\geq J^d(\gamma^e,\gamma^{d_\mathbf{y}}(\gamma^e),\gamma^{d_\mathbf{x}}(\gamma^e))\quad\forall\gamma^{d_\mathbf{y}}\in\Gamma^{d_\mathbf{y}},\gamma^{d_\mathbf{x}}\in\Gamma^{d_\mathbf{x}}.\end{aligned}$$

Note that the follower (decoder) takes its action after observing the strategy γ^e of the leader (encoder), thus the strategies $(\gamma^{d_\mathbf{y}}(\gamma^e),\gamma^{d_\mathbf{x}}(\gamma^e))$ of the decoder are a function of γ^e.

III. Main Results

Before we start with our main results, we first introduce the general structure of the "stochastic" encoder and decoder policies for our setup. In particular, the encoder is given by the transition matrix

$$\mathbf{P}^e(z|x,y)=\begin{bmatrix}\kappa_1 & \kappa_2 & \kappa_3 & \kappa_4\\1-\kappa_1 & 1-\kappa_2 & 1-\kappa_3 & 1-\kappa_4\end{bmatrix}, \qquad (6)$$

where $(\kappa_1,\kappa_2,\kappa_3,\kappa_4)\in[0,1]^4$, whereas the transition matrices at the decoder are given by the column stochastic matrices

$$\mathbf{P}^{d_\mathbf{y}}(\hat{y}|z)=\begin{bmatrix}\delta_1 & \delta_2\\1-\delta_1 & 1-\delta_2\end{bmatrix}, \qquad (7)$$

$$\mathbf{P}^{d_\mathbf{x}}(\hat{x}|z)=\begin{bmatrix}\epsilon_1 & \epsilon_2\\1-\epsilon_1 & 1-\epsilon_2\end{bmatrix}, \qquad (8)$$

where $(\delta_1,\delta_2,\epsilon_1,\epsilon_2)\in[0,1]^4$. To derive our main results, we make use of the following assumption.

Assumption 1. *(Structural assumption on (6)) Restrict the information structure in (6) to one that $\kappa_3=1-\kappa_2$ and $\kappa_4=1-\kappa_1$.*

Remark 1. *(Comments on Assumption 1) By putting such a restriction on κ_3 and κ_4 (i.e., a "symmetric" encoder assumption), we prevent infinitely many quadruples $(\kappa_1,\kappa_2,\kappa_3,\kappa_4)$ resulting in essentially equivalent encoders with respect to performance. Furthermore, after eliminating redundant quadruples by Assumption 1, it is possible to obtain the (joint) convexity of $I(\mathbf{y};\hat{\mathbf{y}})$ with respect to κ_1 and κ_2. Otherwise, i.e., without Assumption 1, there is no conclusion on the (joint) convexity/concavity of $I(\mathbf{y};\hat{\mathbf{y}})$ with respect to the quadruple $(\kappa_1,\kappa_2,\kappa_3,\kappa_4)$.*

Next, we prove a lemma that reformulates the objective functions of (2), (4). We note that this lemma holds even if Assumption 1 does not hold.

Lemma 1. *(Characterization) For the information structure of the stochastic encoder and decoder in (6)-(8), the objective functions in (2), (4) can be characterized as follows*

$$J^e(\gamma^e,\gamma^{d_\mathbf{x}},\gamma^{d_\mathbf{y}})=I(\mathbf{y};\hat{\mathbf{y}})+\rho\mathbf{E}\{d_H(\mathbf{x},\hat{\mathbf{x}})\}, \qquad (9)$$

$$J^d(\gamma^e,\gamma^{d_\mathbf{x}},\gamma^{d_\mathbf{y}})=I(\mathbf{y};\hat{\mathbf{y}})-\mathbf{E}\{d_H(\mathbf{x},\hat{\mathbf{x}})\}, \qquad (10)$$

where[1]

$$I(\mathbf{y};\hat{\mathbf{y}}) = \\ H_b(q_1) + H_b(P_1+P_2) + P_1\log(P_1) + P_2\log(P_2) + \\ (q_1-P_1)\log(q_1-P_1) + (1-q_1-P_2)\log(1-q_1-P_2)\,, \tag{11}$$

$$\mathbf{E}\{d_H(\mathbf{x},\hat{\mathbf{x}})\} = a(1-n_1) + b(1-n_2) + cn_3 + dn_4, \tag{12}$$

with $H_b(p)$ denoting the binary entropy function, i.e., $H_b(p) \triangleq -p\log p - (1-p)\log(1-p)$, $q_1 \triangleq a+c$, $P_1 \triangleq at_1 + ct_3$, $P_2 \triangleq bt_2 + dt_4$, $t_i \triangleq \delta_1\kappa_i + \delta_2(1-\kappa_i)$ and $n_i \triangleq \epsilon_1\kappa_i + \epsilon_2(1-\kappa_i)$ for $i = 1,2,3,4$.

After formulating the objectives of the encoder and the decoder, next, we characterize their optimal strategies, in particular, their best responses for any other given strategy.

Lemma 2. *(Best Response: Encoder) Suppose that Assumption 1 holds. Then, for given decoder strategies $\gamma^{d_{\mathbf{x}}}$ and $\gamma^{d_{\mathbf{y}}}$, the objective function of the encoder in (9) is a jointly convex function of the pair (κ_1,κ_2), and the maximum is achieved at one of the extreme points, i.e., $\kappa_1\kappa_2 = \{00, 01, 10, 11\}$.*

Lemma 3. *(Best Response: Decoder) Suppose that Assumption 1 holds. Then the following hold.*

(i) *For a given encoder strategy γ^e, $I(\mathbf{y};\hat{\mathbf{y}})$ in (11) is a jointly convex function of the pair (δ_1,δ_2), and the maximum is achieved either when $\delta_1\delta_2 = 01$ or $\delta_1\delta_2 = 10$.*

(ii) *For a given encoder strategy γ^e, the average distortion $\mathbf{E}\{d_H(\mathbf{x},\hat{\mathbf{x}})\}$ in (12) is minimized using the decoder strategy characterized in Table I, where $\theta \triangleq \kappa_1(a+d) + \kappa_2(b+c)$.*

TABLE I: Optimal decoder strategy to minimize the average distortion.

Condition	ϵ_1	ϵ_2	$\mathbf{E}\{d_H(\mathbf{x},\hat{\mathbf{x}})\}$
$a+b \le \theta \le c+d$	0	0	$a+b$
$\theta \le a+b,\ \theta \le c+d$	0	1	θ
$\theta \ge a+b,\ \theta \ge c+d$	1	0	$1-\theta$
$a+b \ge \theta \ge c+d$	1	1	$c+d$

Next, we proceed to derive conditions for which Nash and/or Stackelberg equilibria exist together with their corresponding optimal strategies.

Theorem 1. *(Stackelberg) Suppose that Assumption 1 holds and $\min\{a+b, c+d, a+d, b+c\}$ is $a+b$ or $c+d$. Then, the following strategies form a Stackelberg equilibrium:*

$$\begin{aligned} &\kappa_1\kappa_2 = 01 \text{ or } \kappa_1\kappa_2 = 10 \quad \textit{(Encoder)} \\ &\delta_1\delta_2 = 01 \text{ or } \delta_1\delta_2 = 10 \quad \textit{(Decoder-}\hat{\mathbf{y}}\textit{)} \\ &\epsilon_1\epsilon_2 = \begin{cases} 00 \text{ if } a+b \le c+d \\ 11 \text{ if } a+b \ge c+d \end{cases} . \quad \textit{(Decoder-}\hat{\mathbf{x}}\textit{)} \end{aligned} \tag{13}$$

Otherwise, if $\min\{a+b, c+d, a+d, b+c\}$ is $a+d$ or $b+c$, then, for sufficiently small ρ, the equilibrium strategies of (13) are still valid. In contrast, for sufficiently large ρ, the decoder strategies in (13) are still the same and the optimum encoder strategy lies at the boundary of the $\kappa_1\kappa_2$ region which satisfy either $a+b \le \theta \le c+d$ or $a+b \ge \theta \ge c+d$, where $\theta \triangleq \kappa_1(a+d) + \kappa_2(b+c)$ (defined as before).

Theorem 2. *(Nash) Suppose that Assumption 1 holds and $\min\{a+b, c+d, a+d, b+c\}$ is $a+b$ or $c+d$. Then, the same strategies as in (13) form a Nash equilibrium. Otherwise, there does not exist a Nash equilibrium.*[2]

Next we give some technical comments related to our results in Theorems 1, 2.

Remark 2. *(Technical comments)* **(TC1)** *When in Theorems 1, 2, $\min\{a+b, c+d, a+d, b+c\}$ is $a+b$ or $c+d$, the optimal encoder selects either $\kappa_1\kappa_2 = 01$ or $\kappa_1\kappa_2 = 10$, which correspond to sending information only about $\mathbf{y}$ (e.g., $\kappa_1\kappa_2 = 01$ implies $\mathbf{P}^e(z|x,y) = \mathbf{P}^e(z|y)$). In this case, since the received message $\mathbf{z}$ does not contain any direct information about $\mathbf{x}$, the decoder-$\hat{\mathbf{x}}$ uses only priors of $\mathbf{P}(x)$ and achieves the average distortion $\mathbf{E}\{d_H(\mathbf{x},\hat{\mathbf{x}})\} = \min\{a+b, c+d\}$.* **(TC2)** *If $\min\{a+b, c+d, a+d, b+c\}$ is $a+d$ or $b+c$ and the encoder still uses $\kappa_1\kappa_2 = 01$ or $\kappa_1\kappa_2 = 10$, then, the decoder makes use of the conditional probability $\mathbf{P}(x|y)$ (since $\mathbf{z}$ is directly related to $\mathbf{y}$), which further means that the average distortion $\mathbf{E}\{d_H(\mathbf{x},\hat{\mathbf{x}})\} = \min\{a+d, b+c\}$. Hence in order to increase the privacy level (i.e., increase the average distortion to $\min\{a+b, c+d\}$), the encoder uses different strategies that result in smaller value of mutual information (see Fig. 5). To make this point clear, we note that the strategies in (13) result in $I(\mathbf{y};\hat{\mathbf{y}}) = H_b(q_1)$ and $\mathbf{E}\{d_H(\mathbf{x},\hat{\mathbf{x}})\} = \min\{a+b, c+d\}$. However, for large privacy weighting coefficient ρ, as shown in Theorem 1, the average distortion does not change, and the mutual information, as a function of the pair (κ_1,κ_2), can be characterized as in (11) with $P_1 = a\kappa_1 + c(1-\kappa_2)$ and $P_2 = b\kappa_2 + d(1-\kappa_1)$. The resulting mutual information value will be less than $H_b(q_1)$, which means that the encoder ventures to send less information about the public message to be able to hide information about the private message.*

Remark 3. *(Connection to similar work) In [30], a similar setup is considered in which the random sources are jointly Gaussian, and the squared error is utilized as a privacy and information exchange metric. Similar to our result, it is shown that Stackelberg and payoff dominant Nash equilibria coincide. However, due to the difference between our source assumption (i.e., binary), information exchange metric (i.e., mutual information), and privacy metric (i.e., Hamming distortion), we have some cases under which Stackelberg equilibria exist, but there is no Nash equilibrium.*

IV. Numerical Results

In this section, we validate our theoretical results via simulations. For all simulations, we let $a = 0.3$, $b = 0.1$,

[1] The logarithms are taken with base two throughout the paper.

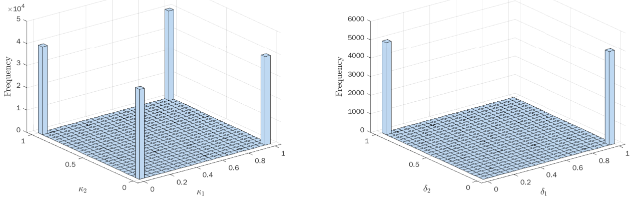
[2] We exclude the trivial case of equal priors $a = b = c = d = 0.25$ in which any strategy pair ends up an equilibrium.

$c = 0.2$, and $d = 1 - a - b - c = 0.4$. We start by validating the best responses of the players as follows:

(i) First we let $\rho = 1$. Then, for given decoder strategies $(\delta_1, \delta_2) \in [0,1]^2$ and $(\epsilon_1, \epsilon_2) \in [0,1]^2$, we calculate corresponding encoder costs by (9) for every possible symmetric encoder actions $\gamma^e = (\kappa_1, \kappa_2)$, to find the optimal one, i.e., the maximizer. We repeat this process for every δ_1, δ_2, ϵ_1, and ϵ_2, which can take one of the 20 evenly spaced values between 0 and 1. As illustrated in Fig. 3a, for 16×10^4 different combinations, $\kappa_1\kappa_2$ takes only four different values, which are $\delta_1\delta_2 = 00$, $\delta_1\delta_2 = 01$, $\delta_1\delta_2 = 10$, or $\delta_1\delta_2 = 11$. Thus, the best response of the encoder stated in Lemma 2 is proved numerically, too. Note that, in our simulations, when there are multiple optima, the encoder selects any of them randomly.

(ii) For a given encoder strategy $(\kappa_1, \kappa_2) \in [0,1]^2$, we calculate corresponding decoder costs by (10) (indeed, only the mutual information $I(\mathbf{y}; \hat{\mathbf{y}})$ part) for every possible decoder actions $\gamma^{d_{\mathbf{y}}} = (\delta_1, \delta_2)$, to find the optimal one, i.e., the maximizer. We repeat this process for every κ_1 and κ_2, which can take one of the 100 evenly spaced values between 0 and 1. As illustrated in Fig. 3b, for 10^4 different $\kappa_1\kappa_2$ values, $\delta_1\delta_2$ takes only two different values, which are $\delta_1\delta_2 = 01$ or $\delta_1\delta_2 = 10$. Thus, the best response of the decoder stated in Lemma 3.(i) is proved numerically, too. Note that, in our simulations, when there are multiple optima, the decoder selects any of them randomly. Thus, always optimal actions $\delta_1\delta_2 = 01$ or $\delta_1\delta_2 = 10$ are selected approximately equal number of times.

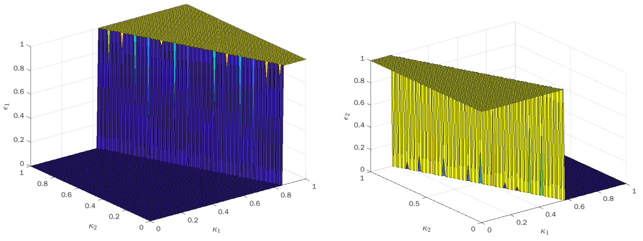
(iii) Similar to the previous analysis, for a given encoder strategy $(\kappa_1, \kappa_2) \in [0,1]^2$, now we calculate corresponding decoder costs by (10) (indeed, only the average distortion $\mathbf{E}\{d_H(\mathbf{x}, \hat{\mathbf{x}})\}$ part) for every possible decoder actions $\gamma^{d_{\mathbf{x}}} = (\epsilon_1, \epsilon_2)$, to find the optimal one, i.e., the minimizer. We repeat this process for every κ_1 and κ_2, and obtain Fig. 4. The optimal actions are $\epsilon_1\epsilon_2 = 00$, $\epsilon_1\epsilon_2 = 01$ or $\epsilon_1\epsilon_2 = 10$. Thus, the best response of the decoder stated in Lemma 3.(ii) is proved numerically, too. Note that, by Table I, $\epsilon_1\epsilon_2$ cannot be 11 since $a + b \geq c + d$ is not satisfied for our selection.



(a) The distribution of the best response of the encoder $\gamma^e = (\kappa_1, \kappa_2)$ for given decoder actions. As it can be seen, $\kappa_1\kappa_2$ can only be any of 00, 01, 10, and 11.

(b) The distribution of the best response of the decoder $\gamma^{d_{\mathbf{y}}} = (\delta_1, \delta_2)$ for given encoder actions. As it can be seen, $\delta_1\delta_2$ is either 01 or 10.

Fig. 3: Simulation results on the best responses of the players.



(a)

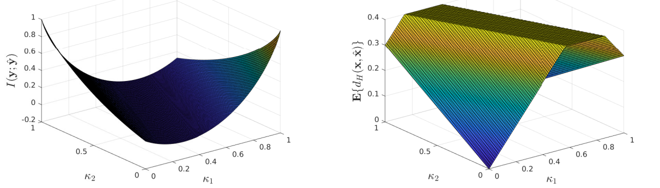
(b)

Fig. 4: The best response of the decoder $\gamma^{d_{\mathbf{x}}} = (\epsilon_1, \epsilon_2)$ for given encoder actions. As it can be seen, $\epsilon_1\epsilon_2$ can only be any of 00, 01, and 10.

After getting the best response maps of the players, we can utilize these results to obtain the Stackelberg equilibrium. In Fig. 5, we plot $I(\mathbf{y}; \hat{\mathbf{y}})$ and $\mathbf{E}\{d_H(\mathbf{x}, \hat{\mathbf{x}})\}$ as a function of the encoder strategy (κ_1, κ_2) for given decoder strategies. In particular, Fig. 5a illustrates the best response of the decoder-$\hat{\mathbf{y}}$ due to a Stackelberg assumption, i.e., $\delta_1\delta_2 = 01$ or $\delta_1\delta_2 = 10$ (via Lemma 3), and the maximum $I(\mathbf{y}; \hat{\mathbf{y}})$ is achieved when $\kappa_1\kappa_2 = 01$ or $\kappa_1\kappa_2 = 10$ (see Theorem 1). In Fig. 5b, via Lemma 3, the best response of the decoder-$\hat{\mathbf{x}}$ is considered due to a Stackelberg assumption. Since $\min\{a + b, c + d, a + d, b + c\} = b + c$, $\kappa_1\kappa_2 = 01$ or $\kappa_1\kappa_2 = 10$ is not in the optimal region to maximize $\mathbf{E}\{d_H(\mathbf{x}, \hat{\mathbf{x}})\}$ (see Theorem 1 and Remark 2). The effect on this confusion can be observed for large value of ρ. Indeed, for small enough privacy weighting coefficient ρ, as it can be seen in Fig. 6a, the maximizers of $I(\mathbf{y}; \hat{\mathbf{y}})$ are still the maximizers of $J^e(\gamma^e, \gamma^{d_{\mathbf{x}}}, \gamma^{d_{\mathbf{y}}})$ in (9). On the other hand, for large enough privacy weighting coefficient ρ, $\mathbf{E}\{d_H(\mathbf{x}, \hat{\mathbf{x}})\}$ gets more dominant in $J^e(\gamma^e, \gamma^{d_{\mathbf{x}}}, \gamma^{d_{\mathbf{y}}})$ in (9). This case is illustrated in Fig. 6b.

V. Conclusion and Future Research

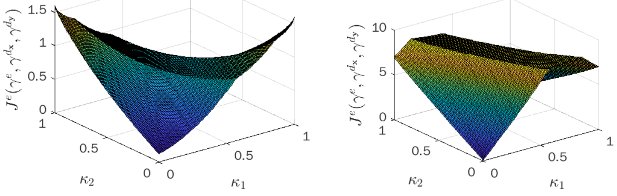
In this paper, we studied Nash and Stackelberg equilibria of privacy signaling games with binary alphabets with single-bit transmission between an encoder and a decoder with misaligned objectives. We derived the conditions under



(a)

(b)

Fig. 5: $I(\mathbf{y}; \hat{\mathbf{y}})$ and $\mathbf{E}\{d_H(\mathbf{x}, \hat{\mathbf{x}})\}$ as a function of the encoder strategy (κ_1, κ_2) to analyze the Stackelberg equilibrium.



(a) $\rho = 2$.

(b) $\rho = 20$.

Fig. 6: $J^e(\gamma^e, \gamma^{d_\mathsf{x}}, \gamma^{d_\mathsf{y}})$ as a function of the encoder strategy (κ_1, κ_2) to analyze the Stackelberg equilibrium. As it can be seen, for small ρ, $\kappa_1\kappa_2 = 01$ and $\kappa_1\kappa_2 = 10$ are still optimal, whereas, the encoder selects intermediate κ_1 and κ_2 values as ρ gets larger.

which Nash and/or Stackelberg equilibria exist.

Our model has several possible interesting extensions. The most important question that needs to be answered is the extension of the framework beyond the single-bit transmission, that is to say, the transmitted messages are random vectors. Another interesting extension would be to consider scenarios with alternative objective functions and privacy criteria (e.g., log-loss function).

Acknowledgement

This work was funded in part by the Swedish Research Council under contract 2019-03606.

References

[1] P. McDaniel and S. McLaughlin, "Security and privacy challenges in the smart grid," *IEEE Security Privacy*, vol. 7, no. 3, pp. 75–77, 2009.

[2] S. Finster and I. Baumgart, "Privacy-aware smart metering: A survey," *IEEE Communications Surveys Tutorials*, vol. 17, no. 2, pp. 1088–1101, 2015.

[3] J. Le Ny and G. J. Pappas, "Differentially private filtering," *IEEE Transactions on Automatic Control*, vol. 59, no. 2, pp. 341–354, 2014.

[4] J. Gómez-Vilardebó and D. Gündüz, "Smart meter privacy for multiple users in the presence of an alternative energy source," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 1, pp. 132–141, 2015.

[5] Z. Li, T. J. Oechtering, and D. Gündüz, "Privacy against a hypothesis testing adversary," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 6, pp. 1567–1581, 2019.

[6] E. Nekouei, T. Tanaka, M. Skoglund, and K. H. Johansson, "Information-theoretic approaches to privacy in estimation and control," *Annual Reviews in Control*, vol. 47, pp. 412–422, 2019.

[7] Y. Lu and M. Zhu, "On privacy preserving data release of linear dynamic networks," *Automatica*, vol. 115, p. 108839, 2020.

[8] B. Cavarec, P. A. Stavrou, M. Bengtsson, and M. Skoglund, "Designing privacy filters for hidden Markov processes," in *European Control Conference (ECC)*, 2021.

[9] V. P. Crawford and J. Sobel, "Strategic information transmission," *Econometrica*, vol. 50, pp. 1431–1451, 1982.

[10] I. Shames, A. M. H. Teixeira, H. Sandberg, and K. H. Johansson, "Agents misbehaving in a network: a vice or a virtue?" *IEEE Network*, vol. 26, no. 3, pp. 35–40, May 2012.

[11] B. Larrousse, O. Beaude, and S. Lasaulce, "Crawford-Sobel meet Lloyd-Max on the grid," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 6127–6131.

[12] J. Miklós-Thal and H. Schumacher, "The value of recommendations," *Games and Economic Behavior*, vol. 79, pp. 132–147, 2013.

[13] O. Ben-Porat and M. Tennenholtz, "A game-theoretic approach to recommendation systems with strategic content providers," in *International Conference on Neural Information Processing Systems (NeurIPS)*, 2018, p. 1118–1128.

[14] J. G. Riley, "Silver signals: Twenty-five years of screening and signaling," *Journal of Economic Literature*, vol. 39, no. 2, pp. 432–478, June 2001.

[15] J. Sobel, "Signaling games," in *Encyclopedia of Complexity and Systems Science*, R. A. Meyers, Ed. Springer New York, 2009, pp. 8125–8139.

[16] E. Kamenica and M. Gentzkow, "Bayesian persuasion," *American Economic Review*, vol. 101, no. 6, pp. 2590–2615, Oct. 2011.

[17] S. Sarıtaş, S. Yüksel, and S. Gezici, "Quadratic multi-dimensional signaling games and affine equilibria," *IEEE Transactions on Automatic Control*, vol. 62, no. 2, pp. 605–619, Feb. 2017.

[18] F. Farokhi, A. M. H. Teixeira, and C. Langbort, "Estimation with strategic sensors," *IEEE Transactions on Automatic Control*, vol. 62, no. 2, pp. 724–739, Feb. 2017.

[19] E. Akyol, C. Langbort, and T. Başar, "Information-theoretic approach to strategic communication as a hierarchical game," *Proceedings of the IEEE*, vol. 105, no. 2, pp. 205–218, Feb. 2017.

[20] M. O. Sayin, E. Akyol, and T. Başar, "Hierarchical multistage Gaussian signaling games in noncooperative communication and control systems," *Automatica*, vol. 107, pp. 9–20, 2019.

[21] S. Sarıtaş, S. Yüksel, and S. Gezici, "Dynamic signaling games with quadratic criteria under Nash and Stackelberg equilibria," *Automatica*, vol. 115, p. 108883, May 2020.

[22] M. le Treust and T. Tomala, "Strategic communication with decoder side information," in *IEEE International Symposium on Information Theory (ISIT)*, 2021, pp. 2696–2701.

[23] M. L. Treust and T. Tomala, "Persuasion with limited communication capacity," *Journal of Economic Theory*, vol. 184, p. 104940, 2019.

[24] S. Sarıtaş, G. Dán, and H. Sandberg, "Passive fault-tolerant estimation under strategic adversarial bias," in *American Control Conference (ACC)*, 2020, pp. 4644–4651.

[25] E. Kazıklı, S. Sarıtaş, S. Gezici, and S. Yüksel, "Quadratic signaling with prior mismatch at an encoder and decoder: Equilibria, continuity and robustness properties," *IEEE Transactions on Automatic Control*, pp. 1–1, 2022.

[26] S. Sarıtaş, P. A. Stavrou, R. Thobaben, and M. Skoglund, "Quadratic signaling games with channel combining ratio," in *IEEE International Symposium on Information Theory (ISIT)*, 2021, pp. 2690–2695.

[27] F. Farokhi, H. Sandberg, I. Shames, and M. Cantoni, "Quadratic Gaussian privacy games," in *54th IEEE Conference on Decision and Control (CDC)*, 2015, pp. 4505–4510.

[28] E. Akyol, C. Langbort, and T. Başar, "Privacy constrained information processing," in *54th IEEE Conference on Decision and Control (CDC)*, 2015, pp. 4511–4516.

[29] F. Farokhi and G. Nair, "Privacy-constrained communication," *IFAC-PapersOnLine*, vol. 49, no. 22, pp. 43–48, 2016.

[30] E. Kazikli, S. Gezici, and S. Yüksel, "Quadratic privacy-signaling games and the MMSE information bottleneck problem for Gaussian sources," *arxiv.org*, vol. abs/2005.05743v3, 2022. [Online]. Available: https://arxiv.org/abs/2005.05743v3

[31] I. Wagner and D. Eckhoff, "Technical privacy metrics: A systematic survey," *ACM Comput. Surv.*, vol. 51, no. 3, June 2018.

[32] P. A. Stavrou, S. Sarıtaş, and M. Skoglund, "Privacy signaling games with binary alphabets," *arxiv.org*, vol. abs/2111.05947, 2021. [Online]. Available: https://arxiv.org/abs/2111.05947v2

[33] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, Inc., Hoboken, New Jersey, 2006.

[34] T. Başar and G. J. Olsder, *Dynamic Noncooperative Game Theory*. Philadelphia, PA: SIAM Classics in Applied Mathematics, 1999.