Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC-PapersOnLine 49-22 (2016) 043–048

Privacy-Constrained Communication ⋆

Farhad Farokhi * **Girish Nair** *

* *Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, VIC 3010, Australia (e-mail: {ffarokhi,gnair}@unimelb.edu.au)*

Abstract: A game is introduced to study the effect of privacy in strategic communication between a well-informed sender and a receiver. The receiver wants to accurately estimate a random variable. The sender, however, wants to communicate a message that balances a trade-off between providing an accurate measurement and minimizing the amount of leaked private information, which is assumed to be correlated with the to-be-estimated variable. The mutual information between the transmitted message and the private information is used as a measure of the amount of leaked information. An equilibrium is constructed and its properties are investigated.

Keywords: Data privacy; Information Theory; Communication Systems; Statistical inference

1. INTRODUCTION

Participatory and crowd-sensing technologies rely on honest data from recruited users to generate estimates of variables, such as traffic condition. Providing accurate information by the users can undermine their privacy. For instance, a road user that provides her start and finish points as well as travel time to a participatory-sensing app can significantly improve the quality of the traffic estimation at the expense of exposing her private life. Therefore, she can benefit from providing "false" information, not to deceive the system but to protect her privacy. The amount of the deviation from the truth is determined by the value of privacy that can vary across the population. To better understand this effect, we use a game-theoretic framework to model the conflict of interest and to study the effect of privacy in strategic communication.

We develop a model in which the receiver is interested in estimating a random variable. To this aim, it asks a better-informed sender to provide a measurement. The sender wants to find a trade-off between her desire to provide an accurate measurement of the variable while minimizing the amount of leaked private information, which is potentially correlated with that variable or its measurement. We assume that the sender has access to a possibly noisy measurement of the variable and a perfect measurement of her private information. We use *mutual information* between the communicated message and the private information to capture the amount of the leaked information. We present a numerical algorithm for finding an equilibrium (i.e., policies from which no one has an incentive to unilaterally deviate) of the presented game when the random variables are discrete. In the continuous case, a fundamental bound on the estimation error is provided for Gaussian random variables and the equilibrium over affine policies are shown to achieve the bound when the emphasis on the privacy grows.

The problem considered in this paper is close, in essence, to the idea of differential privacy and its application in estimation and signal processing, e.g. (Dwork, 2008; Friedman and Schuster, 2010; Huang et al., 2014; Le Ny and Pappas, 2014). Those studies rely on adding noise, typically Laplace noises, to guarantee the privacy of the users by making the outcome less sensitive to local parameter variations. Various studies were devoted to finding "optimal" noise distribution in differential privacy (Soria-Comas and Domingo-Ferrer, 2013; Geng and Viswanath, 2014) or other variants such as Lipschitz privacy (Koufogiannis et al., 2015). Contrary to these studies, we study privacy-constrained communication using game theory by explicitly modelling the conflict of interest between the senders and the receiver stemming from the privacy constraint. Further, we show that in the case of continuous random variables even when the emphasis on the privacy grows, the sender either does not add noise to the transmitted messages or the intensity of the noise does not grow with the value of privacy, which are not the case in the differential privacy literature. The interesting fact that the sender does not add noise was proved in (Akyol et al., 2015) for scalar random variables. However, the above-mentioned observations regarding the intensity of the noise are valid irrespective of the dimensions of the random variables.

In the information theory literature, wiretap channels have been studied heavily dating from the pioneering work in (Wyner, 1975). In these problems, the sender wishes to devise encoding schemes to create a secure channel for communicating with the receiver while hiding her data from an eavesdropper. This is a secrecy problem. In contrast, other studies have considered the privacy problem in which the masking or equivocation of information corresponds to either the intended primary receiver rather than an eavesdropper (Sankar et al., 2013; Courtade et al., 2012) or a secondary receiver with as much information as the primary one (Yamamoto, 1983, 1988). Information-theoretic guarantees on the amount of leaked private information when utilizing the differential privacy framework was given in (Alvim et al., 2011; du Pin Calmon and Fawaz, 2012). Privacy-aware machine learning was discussed in (Wainwright et al., 2012), where the mutual information between the transmitted message and the original data points is used to measure and constrain the loss of privacy. The contributions of the paper, which sets it apart from the above-mentioned studies, is that we use

⋆ The work of F. Farokhi was supported by a McKenzie Fellowship, ARC grant LP130100605, a grant from Melbourne School of Engineering. The work of G. Nair was supported by ARC grants DP140100819 and FT140100527.

10.1016/j.ifacol.2016.10.370

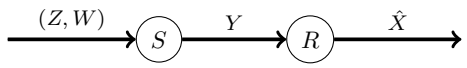

Fig. 1. Communication structure between the sender S and the receiver R.

a strategic game to model the interactions between the sender and the receiver. As we see shortly, there exists an equilibrium in which the players cooperate and thus the problem transforms into a team play; however, there are other equilibria in which the players would not necessarily cooperate.

The idea of strategic communication has been studied in the economics literature in the context of cheap-talk games (Crawford and Sobel, 1982) in which well-informed senders communicates with a receiver that makes a decision regarding the society's welfare. In those games, the sender(s) and the receiver have a clear conflict of interest, which results in potentially dishonest messages. Contrary to those studies, here, the conflict of interest is motivated by the sense of privacy of the sender, which changes the form of the cost functions. Cheap-talk games were recently adapted to investigate privacy in communication and estimation (Farokhi et al., 2015). That study, however, focuses quadratic cost functions and Gaussian random variables. Privacy constrained information processing with an entropy based privacy measure was studied by (Akyol et al., 2015). Contrary to these studies, we consider both discrete and continuous random variables with arbitrary distributions. For the continuous Gaussian random variables, we also present a fundamental bound on the variance of the estimation error at the equilibrium, which is missing from the literature.

The rest of the paper is organized as follows. In Section 2, the problem formulation for discrete random variables is introduced and the equilibria of the game are constructed. The results are extended to continuous random variables in Section 3. The paper is concluded in Section 4.

2. DISCRETE RANDOM VARIABLES

We consider strategic communication between a sender and a receiver as depicted in Fig. 1. The receiver wants to have an accurate measurement of a discrete random variable $X \in \mathcal{X}$, where $\mathcal{X}$ denotes the set of all the possibilities. To this aim, the receiver deploys a sensor (which is a part of the sender) to provide a measurement of the variable. The measurement is denoted by $Z \in \mathcal{X}$. The sender also has another discrete random variable denoted by $W \in \mathcal{W}$, which is correlated with X and/or Z. This random variable is the sender's private information, i.e., it is not known by the receiver. The sender wants to transmit a message $Y \in \mathcal{Y}$ that contains useful information about the measured variable while minimizing the amount of the leaked private information (note that, because of the correlation between W and X and/or Z, an honest report of Z may shine some light on the realization of W). Throughout this paper, for notational consistency, we use capital letters to denote the random variables, e.g., X, and small letters to denote a value, e.g., x.

Assumption 1. The discrete random variables X, Z, W are distributed according to a joint probability distribution $p : \mathcal{X} \times \mathcal{X} \times \mathcal{W} \to [0, 1]$, i.e., $\mathbb{P}\{X = x, Z = z, W = w\} = p(x, z, w)$ for all $(x, z, w) \in \mathcal{X} \times \mathcal{X} \times \mathcal{W}$.

The conflict of interest between the sender and the receiver can be modelled and analysed as a game. This conflict of interest can manifest itself in the following ways:

1) In participatory-sensing schemes, the sender's measurement of the state potentially depends on the way that the sender experiences the underlying process or services. For instance, in traffic estimation, the sender's measurement is fairly accurate on the route that she has travelled and, thus, an honest revelation of Z provides a window into the life of the commuter. However, the underlying state X is not related to the private information of sender W since she is only an infinitesimal part of the traffic flow. In such a case, we have
$$\begin{aligned}&\mathbb{P}\{X = x, Z = z, W = w\}\\ &\quad = \mathbb{P}\{Z = z|X = x, W = w\}\mathbb{P}\{X = x, W = w\}\\ &\quad = \mathbb{P}\{Z = z|X = x, W = w\}\mathbb{P}\{X = x\}\mathbb{P}\{W = w\},\end{aligned}$$
where the second equality follows from independence of random variables W and X.
2) In many services, such as buying insurance coverage or participating in polling surveys, an individual should provide an accurate history of her life or beliefs. In these cases, the variable X highly depends on the private information of the sender W (if not equal to it). In such cases,the measurement Z may not contain any error as well.

The results of this paper rely on the knowledge of that the cross correlation of the random variables X and W. In practice, these information might not be available due to the complexity of the privacy problem. However, the presented results provide valuable insights regarding the nature of the problem and the arising complexities. In what follows, the privacy game is properly defined.

2.1 Receiver

The receiver constructs its best estimate $\hat{X} \in \mathcal{X}$ using the conditional distribution $\mathbb{P}\{\hat{X} = \hat{x}|Y = y\} = \beta_{\hat{x}y}$ for all $(\hat{x}, y) \in \mathcal{X} \times \mathcal{Y}$. The matrix $\beta = (\beta_{\hat{x}y})_{(\hat{x},y)\in\mathcal{X}\times\mathcal{Y}} \in B$ is the policy of the receiver with the set of feasible policies defined as
$$B=\left\{\beta:\beta_{\hat{x}y} \in [0, 1],\forall(\hat{x}, y) \in \mathcal{X} \times \mathcal{Y} \wedge \sum_{\hat{x}\in\mathcal{X}}\beta_{\hat{x}y}=1,\forall y \in \mathcal{Y}\right\}.$$
The receiver prefers an accurate measurement of the variable X. Therefore, the receiver wants to minimize the cost function $\mathbb{E}\{d(X, \hat{X})\}$ with the mapping $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_{\geq 0}$ being a measure of distance between the entries of set $\mathcal{X}$. An example of such a distance is
$$d(x, \hat{x}) = \begin{cases} 0, & x = \hat{x},\\ 1, & x \neq \hat{x}.\end{cases} \tag{1}$$
When using the distance mapping in (1), the term $\mathbb{E}\{d(X, \hat{X})\}$ becomes the probability of error at the receiver. The results of this paper are valid irrespective of the choice of this mapping.

2.2 Sender

The sender constructs its message $y \in \mathcal{Y}$ according to the conditional probability distribution $\mathbb{P}\{Y = y|Z = z, W = w\} = \alpha_{yzw}$ for all $(y, z, w) \in \mathcal{Y} \times \mathcal{X} \times \mathcal{W}$. Therefore, the tensor $\alpha = (\alpha_{yzw})_{(y,z,w)\in\mathcal{Y}\times\mathcal{X}\times\mathcal{W}} \in A$ denotes the policy of the sender. The set of feasible policies is given by
$$A = \left\{\alpha : \alpha_{yzw} \in [0, 1], \forall(y, z, w) \in \mathcal{Y} \times \mathcal{X} \times \mathcal{W} \wedge \sum_{y\in\mathcal{Y}} \alpha_{yzw} = 1, \forall(z, w) \in \mathcal{X} \times \mathcal{W}\right\}.$$