

Competitive Privacy in the Smart Grid: An Information-theoretic Approach

Lalitha Sankar, Soumya Kar, Ravi Tandon, H. Vincent Poor
Dept. of Electrical Engineering,
Princeton University, Princeton, NJ 08544.
{lalitha,skar,rtandon,poor}@princeton.edu

Abstract—Advances in sensing and communication capabilities as well as power industry deregulation are driving the need for distributed state estimation at the regional transmission organizations (RTOs). This leads to a new *competitive privacy* problem amongst the RTOs since there is a tension between sharing data to ensure network reliability (utility/benefit to all RTOs) and withholding data for profitability and privacy reasons. The resulting tradeoff between utility, quantified via fidelity of its state estimate at each RTO, and privacy, quantified via the leakage of the state of one RTO at other RTOs, is captured precisely using a lossy source coding problem formulation for a two RTO network. For a two-RTO model, it is shown that the set of all feasible utility-privacy pairs can be achieved via a single round of communication when each RTO communicates taking into account the correlation between the measured data at both RTOs. The lossy source coding problem and solution developed here is also of independent interest.

I. INTRODUCTION

The electric power industry is undergoing profound changes as greater emphasis is placed on the importance of a smarter grid that supports sustainable energy utilization. Technically, enabled by advances in sensing, communication, and actuation, power system estimation and control are likely to involve many more fast information gathering and processing devices (e.g. Phasor Measurement Units) [1]. Economically, the deregulation of the electricity industry has led to the creation of many regional transmission organizations (RTOs) within a large interconnected power system [2] (see Fig. 1). Both technical and economic drivers suggest the need for more distributed estimation and control in power system operations.

Distributed Cooperation: While the distributed state estimation problem has been investigated for well over two decades, the focus to date has been on two-tier hierarchical models [3] in which each local control center (e.g., RTO) estimates independently without sharing any information amongst its neighbors, and at a higher level, a central coordinator receives the estimation results from the individual areas and coordinates them to obtain a system-wide solution. However, such an hierarchical approach does not scale with increasing measurement rates due to communication and reliability challenges

The research was supported in part by the Air Force Office of Scientific Research under MURI Grant FA9550-09-1-0643, in part by the Army Research Office under MURI Grant W911NF-11-1-0036, and in part by the National Science Foundation under Grants CNS-09-05398 and CCF-10-16671.

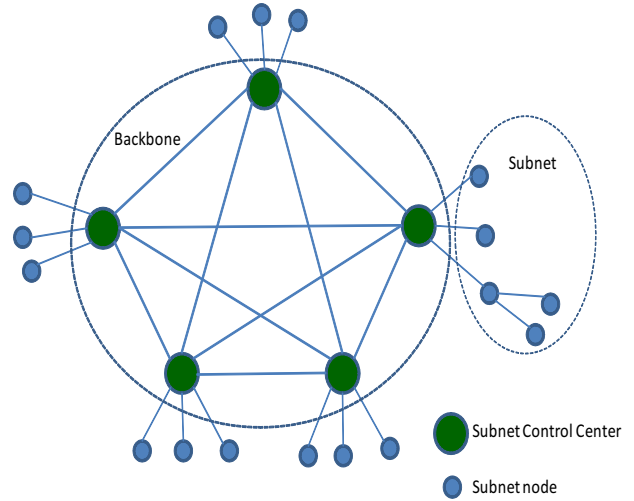


Fig. 1. Multiple RTOs (subnets) connected via the transmission network.

inherent in systems with one coordination center. Furthermore, the interconnectedness of the RTOs makes the problem of wide area monitoring and control (e.g. for the entire Eastern United States interconnection) important and immediate. This requirement of sharing the entire system state amongst all the RTOs is driving the need for a fully distributed approach to state estimation wherein the local control centers interactively estimate the system state as a whole.

There are several challenges that arise in the context of a fully distributed state estimation approach. We address a specific problem of collaboration and competition by introducing a framework that precisely quantifies the tradeoff between the utility (benefit) of collaboration and the resulting privacy leakage.

Competitive Privacy: The problem of end-user privacy has begun receiving attention with the deployment of smart meters to monitor and finely manage user power consumption [4]–[9]. In contrast, a new *competitive privacy* problem arises at the level of the RTOs due to the conflict between sharing data for distributed estimation and withholding data for economic (competitive) and end-user privacy reasons, i.e., there is a tradeoff between sharing data to ensure network reliability (utility/benefit to all RTOs) and withholding data to ensure

profitability and privacy.

Competitive privacy has been studied by the database community in the context of data integration from multiple autonomous sources that do not wish to reveal private sensitive business information to one another and yet benefit from the combined analysis [10]. However, the solutions proposed are unique for databases and do not directly apply to the distributed state estimation problem discussed above. An analytical framework that abstracts the problem of competitive privacy via appropriate source models and utility and privacy metrics is required and it is this problem that we address in this paper.

Utility vs. Privacy: Utility and privacy are competing goals: utility is maximized when the RTOs completely disclose their measurements to each other; however, privacy is minimal for this disclosure. On the other hand, not sharing any data guarantees maximal privacy but achieves zero utility. Thus, ensuring privacy requires distorting or perturbing the data which also provide utility guarantees. The theory of rate distortion allows us to study this tradeoff between privacy and utility using appropriate metrics for both constraints; we present such a formal framework in this paper for the competitive privacy problem.

Contributions: In this paper, we present a linear measurement model for the grid at the level of the RTOs that takes into account the interconnections amongst them. Viewing the power system state at each RTO as an information source, we model the measurements at each RTO as a linear combination of all the sources. The competitive privacy problem formulated thus leads to a new *distributed lossy source coding problem* wherein each RTO shares (encodes) a perturbed function of its measurements, and over a finite number of such communication rounds, estimates its state subject to two constraints: a) each RTO must be able to decode its own measurements to a desired fidelity, and b) at each RTO, the privacy leakage of the measurements shared by the other RTOs must be bounded.

Distributed source coding has received much attention recently with a growing interest in sensor networks. However, with a few exceptions, most distributed source coding problems remain open. One such exception is the quadratic Gaussian CEO problem in which a source is observed noisily by multiple sensors, each of which uses a finite-rate link to communicate it to a single receiver which in turn combines these messages to reconstruct the source to a desired level of fidelity [11]. The problem presented here further generalizes such a setting in the following ways: a) there are multiple sources, b) each sensor (here RTO) is now both an encoder and a decoder, i.e., there are multiple encoders and decoders, and c) there are fidelity requirements at each RTO for its own measurements and privacy requirements vis-à-vis other measurements.

Having modeled the general problem, we focus on a two-source setting that captures the distributed estimation problem for two large RTOs. Using distortion as a metric for utility and mutual information as a metric for privacy leakage, we demonstrate that a single-shot rate-distortion code at each RTO

that is cognizant of the side information (measurement vector) at the other RTO achieves the set of all feasible utility-privacy operating points.

The paper is organized as follows: in Section II, we develop the model and present abstract metrics for utility and privacy. In Section III, we present our main results. We conclude in Section IV.

II. MODEL AND METRICS

A. Model

Let M denote the total number of RTOs and let X_m denote the power system state variable at the m^{th} RTO, $m = 1, 2, \dots, M$. Assuming linearized system dynamics, the general measurement model at the k^{th} terminal at a sampling time instant is given by

$$Y_m = \sum_{j=1}^M H_{m,j} X_j + Z_m, \quad m = 1, 2, \dots, M, \quad (1)$$

where $X_m \sim \mathcal{N}(0, 1)$, for all m , are assumed to be mutually independent, and where $H_{m,j}$ denotes the Jacobian modeling the linearized dynamics between the j^{th} state and the measurement at the m^{th} RTO. The observation noise at the m^{th} RTO $Z_m \sim \mathcal{N}(0, \sigma_m^2)$ is assumed to be independent of X_m , for all m . We assume that the $H_{m,j}$ are fixed and known at all RTOs for the duration of communications; furthermore, we also assume that the statistics of the noise, the measurements, and the states are known at all the RTOs. In this paper, we focus on case of $M = 2$, i.e., we are interested in the distributed estimation problem between two (typically large) adjacent RTOs. Finally, we set $H_{1,1} = H_{2,2} = 1$, $H_{1,2} = \alpha \in (0, \infty)$, and $H_{2,1} = \beta \in (0, \infty)$.

We assume that the m^{th} RTO observes a sequence of n measurements $Y_m^n = [Y_{m,1} \ Y_{m,2} \ \dots \ Y_{m,n}]$, for all m , over a window of time prior to communications. The communication protocol is shown in Fig. 2 for $M = 2$. In each communication round, the two RTOs take turns such that each RTO encodes (quantizes) its measurements taking into account the fact the other RTO has correlated side information (the correlations come from the measurement vectors at each RTO being functions of the state variables at both RTOs). We assume a total of K such communication rounds before RTO m estimates \hat{X}_m^n , $m = 1, 2$, using the entire sequence of public communications and its own measurements.

B. Metrics

Utility: The motivation for cooperation at each RTO is to obtain a good estimate of its own system state. Thus, for our continuous Gaussian distributed state model, a reasonable metric for utility at the m^{th} RTO is the mean square error D_m of the original and the estimated state sequences X_m^n and \hat{X}_m^n , respectively.

Privacy: The measurements at each RTO in conjunction with the quantized data from the other RTO while enabling estimation of its own state can also potentially leak information about the state at the other RTO. We capture this information leakage via the mutual information.

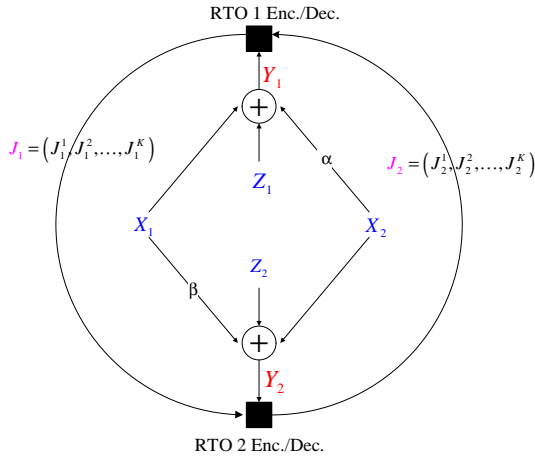


Fig. 2. Communication protocol for two RTOs.

C. Communication Protocol

A coding scheme for the network involves an encoder-decoder pair at each RTO over K communication rounds such that some desired utility and privacy values are achieved. Formally, an $(n, K, M_1, M_2, D_1, D_2, L_1, L_2)$ code for this network results from a K -round protocol in which RTO m , $m = 1, 2$, uses the encoding functions $\{f_{m,k}\}$ for the k^{th} round and a decoding function F_m at the end of the K rounds. In round k , the encoder at RTO m maps its measurements and the messages received until then from the other RTO to an index set \mathcal{J}_m^k where

$$\mathcal{J}_m^k \equiv \{1, 2, \dots, J_m^k\}, \quad k = 1, 2, \dots, K, \quad m = 1, 2, \quad (2)$$

is the index set at the m^{th} RTO for mapping the quantized sequence in the k^{th} protocol round via the encoder f_m^k defined as

$$f_m^k : \mathcal{Y}_m^n \times \mathcal{J}_m^1 \times \dots \times \mathcal{J}_m^{k-1} \rightarrow \mathcal{J}_m^k, \quad k = 1, 2, \dots, K, \quad m = 1, 2, \quad (3)$$

such that at the end of the K rounds, the decoding function F_m at the m^{th} RTO is a mapping from the space of measurements and received messages to that of the reconstructed sequence denoted as

$$F_m : \mathcal{Y}_m^n \times \mathcal{J}_m^1 \times \dots \times \mathcal{J}_m^K \rightarrow \hat{\mathcal{X}}_m^n, \quad m = 1, 2. \quad (4)$$

Let $J_m \equiv \{J_m^k\}_{k=1}^K$, denote the set of all indices communicated by the m^{th} RTO, $m = 1, 2$, such that M_1 and M_2 denote the size of J_1 and J_2 , respectively. The expected distortion D_m at the m^{th} RTO is given by

$$D_m = \frac{1}{n} \mathbb{E} \left[\sum_{i=1}^n \left(X_{m,i} - \hat{X}_{m,i} \right)^2 \right], \quad m = 1, 2, \quad (5)$$

and the privacy leakage, L_1 , about state 1 at RTO 2 is given by

$$L_1 = \frac{1}{n} I(X_1^n; J_1, J_2, Y_2^n), \quad (6)$$

and the privacy leakage, L_2 , about state 2 at RTO 1 is given by

$$L_2 = \frac{1}{n} I(X_2^n; J_1, J_2, Y_1^n). \quad (7)$$

The communication rate of the m^{th} RTO is denoted by

$$R_m = n^{-1} \log_2 M_m, \quad m = 1, 2. \quad (8)$$

Definition 1: The utility-privacy tradeoff region \mathcal{T} is the set of all (D_1, D_2, L_1, L_2) for which there exists a coding scheme given by (2)-(4) with parameters $(n, K, M_1, M_2, D_1 + \epsilon, D_2 + \epsilon, L_1 + \epsilon, L_2 + \epsilon)$ for n sufficiently large such that $\epsilon \rightarrow 0$ as $n \rightarrow \infty$.

III. MAIN RESULTS

The coding and communication protocol described in the previous section is constrained only by the fidelity requirements of the desired state estimate at each RTO and the privacy leakage of the state at the other RTO. Furthermore, from Definition 1, we note that the utility-privacy tradeoff region \mathcal{T} is also dependent on size of the encoding indices M_1 and M_2 , i.e., for every $(D_1, D_2, L_1, L_2) \in \mathcal{T}$, there can be several achievable coding schemes each corresponding to some communication rate pair (R_1, R_2) . Thus, it suffices to determine the set of all rate-distortion-leakage $(R_1, R_2, D_1, D_2, L_1, L_2)$ tuples to determine the utility-privacy tradeoff region.

A. Rate-Distortion-Leakage (RDL) Tradeoff

Recall that $Y_1 = X_1 + \alpha X_2 + N_1$ and $Y_2 = \beta X_1 + X_2 + N_2$, where $X_m \sim \mathcal{N}(0, 1)$, $m = 1, 2$, and $N_m \sim \mathcal{N}(0, \sigma_m^2)$, $m = 1, 2$, are all mutually independent. We first state the range of allowable distortion pairs (D_1, D_2) as

$$D_m \in [D_{\min, m}, D_{\max, m}], \quad m = 1, 2,$$

where

$$D_{\min, 1} = 1 - \frac{(\beta^2 V_1 + V_2 - 2\beta E)}{(V_1 V_2 - E^2)} \quad (9a)$$

$$D_{\min, 2} = 1 - \frac{(V_1 + \alpha^2 V_2 - 2\alpha E)}{(V_1 V_2 - E^2)}, \quad (9b)$$

and

$$D_{\max, 1} = 1 - \frac{1}{V_1}, \quad D_{\max, 2} = 1 - \frac{1}{V_2}, \quad (10a)$$

for

$$V_1 \equiv 1 + \alpha^2 + \sigma_1^2, \quad (11a)$$

$$V_2 \equiv 1 + \beta^2 + \sigma_2^2, \quad \text{and} \quad (11b)$$

$$E \equiv \alpha + \beta. \quad (11c)$$

The lower bounds $D_{\min, 1}$ and $D_{\min, 2}$ on D_1 and D_2 in (9), respectively, are obtained by considering an enhanced system, in which both observations (Y_1, Y_2) can be jointly used for estimating X_1 and X_2 . For such a system, the minimum mean square error (MMSE) estimator minimizes the quadratic distortion in (5). Therefore, $D_{\min, 1}$ and $D_{\min, 2}$ are given by

$$D_{\min, 1} = \mathbb{E} [\text{var}(X_1 | Y_1, Y_2)] \quad (12)$$

$$D_{\min, 2} = \mathbb{E} [\text{var}(X_2 | Y_1, Y_2)] \quad (13)$$

where for two random variables A and B , $\text{var}(A|B)$ is the conditional variance of A conditioned on B . On the other hand, $D_{\max,1}$ and $D_{\max,2}$ correspond to the fidelity criterion achievable using only the locally available measurements at each RTO such that

$$D_{\max,1} = \mathbb{E}[\text{var}(X_1|Y_1)] \quad (14)$$

$$D_{\max,2} = \mathbb{E}[\text{var}(X_2|Y_2)]. \quad (15)$$

The range of feasible leakage values L_1 and L_2 are

$$L_1 \in [I(X_1; Y_2), I(X_1; Y_1, Y_2)], \quad \text{and} \quad (16)$$

$$L_2 \in [I(X_2; Y_1), I(X_2; Y_1, Y_2)]. \quad (17)$$

The lower bound on L_1 , i.e., $I(X_1; Y_2)$ is guaranteed *only* for the case in which there is no public communication. This scenario can arise if D_1 and D_2 can be achieved at each RTO with its own measurements (corresponding to $D_{\max,1}$ and $D_{\max,2}$, respectively). At the other extreme, the maximum value of leakage L_1 can go up to $I(X_1; Y_1, Y_2)$. This corresponds to an infinite rate transmission from RTO 1 to RTO 2 to satisfy the distortion requirement $D_2 = D_{\min,2}$. A similar upper bounding argument holds for L_2 .

Finally, we exploit the Gaussian nature of the system variables to write

$$E[X_1|Y_1, Y_2] = k_1 Y_1 + k_2 Y_2 \quad (18a)$$

$$E[X_2|Y_1, Y_2] = l_1 Y_1 + l_2 Y_2 \quad (18b)$$

where

$$k_1 = \frac{V_2 - \beta E}{V_1 V_2 - E^2}, \quad k_2 = \frac{\beta V_1 - E}{V_1 V_2 - E^2}, \quad \text{and} \quad (19a)$$

$$l_1 = \frac{\alpha V_2 - E}{V_1 V_2 - E^2}, \quad l_2 = \frac{V_1 - \alpha E}{V_1 V_2 - E^2}. \quad (19b)$$

Using these results, we now summarize the rate-distortion-leakage tradeoff for the problem under consideration in the following theorem.

Theorem 2: The rate-distortion-leakage tradeoff $(R_1, R_2, D_1, D_2, L_1, L_2)$ is given as follows.

If $D_2 \in [D_{\min,2}, D_{\max,2}]$, then

$$R_1 = \frac{1}{2} \log \left(\frac{(V_1 V_2 - E^2) l_1^2}{V_2 (D_2 - D_{\min,2})} \right), \quad \text{and} \quad (20)$$

$$L_1 = \frac{1}{2} \log \left(\frac{l_1^2}{l_1^2 D_{\min,1} + k_1^2 (D_2 - D_{\min,2})} \right). \quad (21)$$

If $D_2 \geq D_{\max,2}$, then

$$R_1 = 0, \quad \text{and} \quad (22)$$

$$L_1 = \frac{1}{2} \log (V_2 / (V_2 - \beta)). \quad (23)$$

If $D_1 \in [D_{\min,1}, D_{\max,1}]$,

$$R_2 = \frac{1}{2} \log \left(\frac{(V_1 V_2 - E^2) k_2^2}{V_1 (D_1 - D_{\min,1})} \right), \quad \text{and} \quad (24)$$

$$L_2 = \frac{1}{2} \log \left(\frac{k_2^2}{k_2^2 D_{\min,2} + l_2^2 (D_1 - D_{\min,1})} \right). \quad (25)$$

If $D_1 \geq D_{\max,1}$, then

$$R_2 = 0, \quad \text{and} \quad (26)$$

$$L_2 = \frac{1}{2} \log (V_1 / (V_1 - \alpha)), \quad (27)$$

where $D_{\min,m}$, $D_{\max,m}$, V_m , E , and l_m , k_m , $m = 1, 2$, are defined in (9), (10a), (11), and (19).

Remark 3: Theorem 2 shows that the optimal rate-distortion-leakage tradeoff for the Gaussian case admits a decoupling with respect to the distortion pair (D_1, D_2) . By decoupling, we mean that the leakage of source X_1 and the transmission rate of RTO 1 depends *solely* on the distortion D_2 desired by RTO 2. Similarly, the leakage of X_2 and the transmission rate of RTO 2 depends *solely* on the distortion D_1 desired by RTO 1. The reason for this decoupling can be attributed to the no rate-loss property of jointly Gaussian sources, which can be described as follows: if RTO 2 is interested in reconstructing X_2 at a distortion D_2 , then the minimal rate of transmission required by RTO 1 remains unchanged even if the source Y_2 is available at RTO 1.

We briefly sketch the proof below; we first derive lower bounds on the rate and privacy leakage and present a coding scheme that achieves them.

Proof: Lower bound on R_1, R_2 : Recall that we write J_1 to denote the message from RTO 1 over all K communication rounds. Given a distortion pair (D_1, D_2) , we prove the lower bound on R_1 as follows:

$$nR_1 = H(J_1) \quad (28a)$$

$$\geq I(Y_1^n; J_1 | Y_2^n) \quad (28b)$$

$$\geq nh(Y_1 | Y_2) - \sum_{t=1}^n h(Y_{1t} | J_1, Y_2^n) \quad (28c)$$

$$\geq nh(Y_1 | Y_2) - \sum_{t=1}^n \frac{1}{2} \log (2\pi e \text{Var}(Y_{1t} | J_1, Y_2^n)) \quad (28d)$$

$$\geq nh(Y_1 | Y_2) - \frac{n}{2} \log \left(2\pi e \sum_{t=1}^n \text{Var}(Y_{1t} | J_1, Y_2^n) / n \right) \quad (28e)$$

$$\geq nh(Y_1 | Y_2) - \frac{n}{2} \log (2\pi e (D_2 - D_{\min,2}) / l_1^2) \quad (28f)$$

$$= \frac{n}{2} \log (2\pi e (V_1 V_2 - E^2) / V_2) - \frac{n}{2} \log (2\pi e (D_2 - D_{\min,2}) / l_1^2) \quad (28g)$$

$$= \frac{n}{2} \log \left(\frac{(V_1 V_2 - E^2) l_1^2}{V_2 (D_2 - D_{\min,2})} \right) \quad (28h)$$

where the inequalities in (28) follow from the chain rule, the fact that conditioning does not increase entropy, the fact that Gaussian distribution maximizes the differential entropy for a given variance, and the concavity of the log function. The variance of Y_{1t} conditioned on J_1 and Y_2^n can be computed using (18) and (19) and is omitted due to space limitations. A lower bound on R_2 can be obtained similarly. Therefore, R_1 and R_2 can be lower bounded by the expressions in (20) and (24), respectively.

Lower bounds on (L_1, L_2) : Given an arbitrary code that achieves a certain distortion pair (D_1, D_2) , we derive lower bound on the leakages for both RTOs as follows:

$$L_1 \geq \frac{1}{n} I(X_1^n; J_1, J_2, Y_2^n) \quad (29a)$$

$$\geq h(X_1) - \frac{1}{n} h(X_1^n | J_1, J_2, Y_2^n) \quad (29b)$$

$$\geq h(X_1) - \frac{1}{n} h(X_1^n | J_1, Y_2^n) \quad (29c)$$

$$\geq h(X_1) - \frac{1}{n} h(X_1^n | J_1, Y_2^n) \quad (29d)$$

$$\geq h(X_1) - \frac{1}{n} \sum_{t=1}^n h(X_{1t} | J_1, Y_2^n) \quad (29e)$$

$$\geq h(X_1) - \frac{1}{n} \sum_{t=1}^n \frac{1}{2} \log(2\pi e \text{Var}(X_{1t} | J_1, Y_2^n)) \quad (29f)$$

$$\geq h(X_1) - \frac{1}{2} \log \left(2\pi e \frac{1}{n} \sum_{t=1}^n \text{Var}(X_{1t} | J_1, Y_2^n) \right) \quad (29g)$$

$$\geq h(X_1) - \frac{1}{2} \log(2\pi e (D_{\min,1} + k_1^2 (D_2 - D_{\min,2}) / l_1^2)) \quad (29h)$$

where the inequalities in (29) follow from the chain rule, the fact that conditioning does not increase entropy, the fact that Gaussian distribution maximizes the differential entropy for a given variance, and the concavity of the log function. A lower bound on L_2 can be obtained similarly. Simplifying further, we can lower bound L_1 and L_2 with the expressions in (21) and (25), respectively.

Upper bounds on (R_1, R_2) and (L_1, L_2) via an achievable coding scheme: The lower bounds derived above can be achieved using the Wyner-Ziv source coding scheme with decoder side information [12] at each RTO. We briefly describe the encoding scheme at RTO 1 and the resulting decoding at RTO 2; the coding scheme for RTO 2 follows analogously. The encoding is such that RTO 1 generates a set of M_1 sequences $U_1^n(j_1)$, $j_1 = 1, 2, \dots, M_1$, where $M_1 = 2^{n(I(U_1; Y_1) + \epsilon)}$. However, to exploit the fact that RTO 2 has correlated measurements, RTO 1 further bins its M_1 sequences into S_1 bins chosen at random where $S_1 = 2^{n(I(U_1; Y_1) - I(U_1; Y_2) + \epsilon)}$. Upon observing a measurement sequence y_1^n , the encoder at RTO 1 searches for a $U_1^n(j_1)$ sequence such that y_1^n and $u_1^n(j_1)$ are jointly typical (as defined in [13, Chap. 10]) where the choice of M_1 ensures that there exists at least one such j_1 . Using the fact that RTO 2 has its own measurements Y_2^n as side information, the encoder at RTO 1 sends only $b(j_1)$ where $b(j_1)$ is the bin index of the $u_1^n(j_1)$ sequence.

For Gaussian distributed state and measurements, the sequence U_1^n is chosen such that the ‘test channel’ from U_1 to Y_1 yields $Y_1 = U_1 + Q_1$, where Q_1 is a Gaussian random variable, independent of U_1 , with variance S_1 chosen to satisfy (5) and

$$\hat{X}_1 = \mathbb{E}[X_1 | U_2, Y_1]. \quad (30)$$

Using (5) and (30), we see that the achievable distortion D_1 is simply the conditional variance of X_1 given (Y_1, U_2) .

From (9a), since $D_{\min,1}$ is the conditional variance of X_1 given (Y_1, Y_2) , we have that D_1 is given by the expression in (9a) with V_2 replaced by $V_2 + S_2$; thus, the variance of the quantization noise S_2 is a function of D_1 . One can similarly obtain D_2 from (9b) by replacing V_1 by $V_1 + S_1$ and observe that S_1 depends on D_2 but not D_1 . The resulting rate R_1 is simply the Wyner-Ziv rate R_{WZ} between encoder 1 and decoder 2 with side information Y_2 and is given by

$$R_1 = R_{WZ} = I(Y_1; U_1 | Y_2) \quad (31)$$

$$= \frac{1}{2} \log \left(\frac{\text{var}(Y_1 | Y_2)}{\text{var}(Y_1 | U_1 Y_2)} \right). \quad (32)$$

Writing S_2 in terms of D_2 and $D_{\min,2}$ and simplifying the expressions, one can verify that R_1 simplifies to the expression in Theorem (20). Since U_1^n is a function of (J_1, Y_2^n) , the expression for L_1 defined in (7) simplifies to $I(X_1; U_1 | Y_2)$ which in turn further simplifies to the expression in (21). Finally, one can similarly show that Wyner-Ziv encoding is rate and leakage optimal for RTO 2. ■

Remark 4: The leakage of X_1 at RTO 2 depends only on the desired distortion of X_2 and vice-versa, or alternatively, the feasibility of the distortion desired by RTO 2 depends on the leakage permissible by RTO 1. Thus, cooperation to achieve a desired fidelity at the other RTO inevitably results in a proportional privacy leakage.

Remark 5: Practical implementations of the Wyner-Ziv coding schemes have been well studied and can be applied to the problem at hand.

B. Illustration

We illustrate our results for $\alpha = 1$, $\beta = 8$, $\sigma_1^2 = 0.05$, and $\sigma_2^2 = 1$. The parameters so chosen demonstrate the need for communication between the RTOs since at each RTO, there is interference from the measurements at the other RTO as well as noise. In Fig. 3, R_1 and L_1 are plotted as functions of D_2 while in Fig. 4, the effect of changing the distortion D_1 from its minimum to its maximum value on the rate R_2 and privacy leakage L_2 for RTO 2 are shown. The asymmetry in the interference and noise levels is captured in the two plots.

IV. CONCLUDING REMARKS

In this paper, we have formalized the conflicting objectives of estimation accuracy and competitive privacy in smart-grid operations in terms of a class of information theoretic multi-terminal source coding problems. For a scalar Gaussian source model with two RTOs, we have completely characterized the set of optimal privacy-utility tuples via the rate-distortion-leakage tradeoff for a mean-squared estimation criterion. We have shown that the RDL tradeoff at each RTO depends only on the fidelity requirement at the other RTO with a higher fidelity requirement leading to a higher rate and higher leakage and vice-versa. While the results here also extend to vector Gaussian sources and a mean-square fidelity criterion, we intend to investigate the RDL trade-off for general non-Gaussian source models and multiple RTOs in a forthcoming paper.

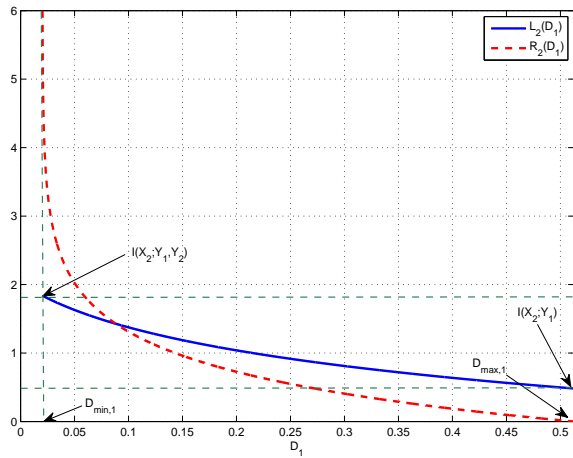


Fig. 3. Plots of R_2 and L_2 vs. D_2 .

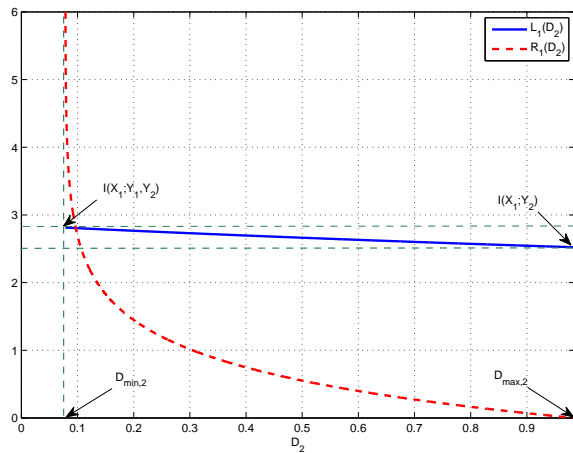


Fig. 4. Plots of R_1 and L_1 vs. D_2 .

Finally, we comment on some practical implementation issues associated with the proposed distributed formulation. While the optimal encoding strategies are necessarily block based, in practice it may not be feasible to accumulate the incoming data for long at each RTO and subsequently encode the entire block. A potential research direction would be the investigation of heuristic real-time encoding schemes that yield close to optimal performance with much less encoding complexity. Finally, a practical approach for the distributed and decentralized information broadcast in such multi-agent networks is via gossip protocols (see, for example, [14]).

REFERENCES

- [1] A. Bose, "Smart transmission grid applications and their supporting infrastructure," *IEEE Trans. Smart Grid*, vol. 1, no. 1, pp. 11–19, Jun. 2010.
- [2] F. F. Wu, K. Moslehi, and A. Bose, "Power system control centers: Past, present, and future," *Proc. IEEE*, vol. 93, no. 11, pp. 1890–1908, Nov. 2005.
- [3] T. V. Cutsem, J. L. Horward, and M. Ribbens-Pavella, "A two-level static state estimator for electric power systems," *IEEE Trans. Power Apparatus and Systems*, vol. 100, no. 8, pp. 3722–3732, Aug. 1981.

- [4] E. L. Quinn, "Privacy and the new energy infrastructure," *Social Science Research Network*, Feb. 2009.
- [5] *Guidelines for Smart Grid Cyber Security: Vol. 2, Privacy and the Smart Grid*, National Institutes of Standards and Technology, Aug. 2010, <http://csrc.nist.gov/publications/>.
- [6] C. Efthymiou and G. Kalogridis, "Smart grid privacy via anonymization of smart metering data," in *Proc. IEEE 1st Intl. Conf. Smart Grid Comm.*, Gaithersburg, MD, USA, Oct. 2010, pp. 238–243.
- [7] G. Kalogridis, C. Efthymiou, S. Z. Denic, T. A. Lewis, and R. Cepeda, "Privacy for smart meters: Towards undetectable appliance load signatures," in *Proc. IEEE 1st Intl. Conf. Smart Grid Comm.*, Gaithersburg, MD, Oct. 2010, pp. 232–237.
- [8] H. Y. Lam, G. S. K. Fung, and W. K. Lee, "A novel method to construct taxonomy of electrical appliances based on load signatures," *IEEE Trans. Consumer Electronics*, vol. 53, no. 2, pp. 653–660, May 2007.
- [9] S. Papadimitriou, F. Li, G. Kollios, and P. S. Yu, "Time series compressibility and privacy," in *Proc. 33rd Intl. Conf. Very Large Databases*, ser. VLDB '07, Vienna, Austria, 2007, pp. 459–470.
- [10] R. C. W. Wong and E. Lo, "Competitive privacy: Secure analysis on integrated sequence data," in *Proc. Database Syst. Ad. Apps.*, Tsukuba, Japan, Apr. 2010, pp. 168–175.
- [11] V. Prabhakaran, D. Tse, and K. Ramachandran, "Rate region of the quadratic Gaussian CEO problem," in *Proc. Intl. Symp. Inform. Theory*, Jun. 28-Jul. 2 2004, p. 119.
- [12] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.
- [13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.
- [14] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inform. Theory*, vol. 52, pp. 2508–2530, Jun. 2006.