Differentially Private State Estimation
in Distribution Networks with Smart Meters

Henrik Sandberg, György Dán, and Ragnar Thobaben

Abstract— State estimation is routinely being performed in high-voltage power transmission grids in order to assist in operation and to detect faulty equipment. In low- and medium-voltage power distribution grids, on the other hand, few real-time measurements are traditionally available, and operation is often conducted based on predicted and historical data. Today, in many parts of the world, smart meters have been deployed at many customers, and their measurements could in principle be shared with the operators in real time to enable improved state estimation. However, customers may feel reluctance in doing so due to privacy concerns. We therefore propose state estimation schemes for a distribution grid model, which ensure differential privacy to the customers. In particular, the state estimation schemes optimize different performance criteria, and a trade-off between a lower bound on the estimation performance versus the customers' differential privacy is derived. The proposed framework is general enough to be applicable also to other distribution networks, such as water networks.

I. INTRODUCTION

State estimation has long been an essential component of electric power system monitoring and control systems [1]. In transmission systems it is typically based on near-real time measurements of voltages, power flows, and more recently, voltage phasors. In low- and medium-voltage distribution systems it is, however, often based on predicted loads in lack of a dedicated communication infrastructure and measurement devices [2], [3]. Due to the proliferation of volatile distributed renewable generation (*e.g.,* residential PV systems), predicted loads will soon no longer be sufficient to achieve a good enough state estimate that would allow efficient and safe operation of the distribution grid, *e.g.,* voltage conservation and VAR control (VVC), or fault location, isolation and restoration (FLIR), and real-time measurements will therefore become necessary.

While a dedicated real-time monitoring infrastructure for distribution grids may be too expensive, automated and smart meters, which are being deployed world-wide for billing purposes, could be used for providing the operators frequent measurements of the loads, thereby enabling dynamic state estimation. Nonetheless, collecting frequent measurement data about residential or industrial customers' loads can be used to invade the customers' privacy in a variety of ways [4].

To protect customers' privacy but at the same time allow frequent smart meter measurements for system operation, a number of privacy-preserving aggregation schemes have been proposed recently [5], [6], [7]. Aggregation can be done via a trusted third party [5] or by using cryptographic solutions such as homomorphic encryption [6], but these solutions increase system complexity significantly. An alternative aggregation scheme that does not increase system complexity relies on adding random noise, typically Gaussian or Laplacian, to the measurement data submitted by the customers. Combined with careful protocol design, aggregation with random noise has been shown to provide differential privacy [7], a probabilistic notion of privacy that is widely used in statistical databases [8], [9] and has recently found application in filtering and control [10], [11]. While adding noise may help preserve privacy, a fundamental question is how it would affect the quality of state estimation, and ultimately the power system applications that rely on it.

In this paper, we address this question by providing a characterization of the trade-off between differential privacy and the mean distortion of the state estimate in a single feeder of a distribution network based on load measurements from smart meters. We characterize the ϵ- and the (ϵ, δ)-differential privacy under Laplacian and Gaussian additive random noise, respectively. We express the maximum a posteriori state estimate as the solution of a convex programming problem, and provide a closed-form expression for the linear minimum mean square error estimate. We show that a customer by adding random noise may only have to give up a little bit extra of privacy, and can still dramatically improve the state estimation performance, which could enable optimizing financial incentives for trading estimation performance and customer privacy.

The trade-off between privacy and estimation quality has been considered for inter-area state estimation in [12], where the trade-off between the mutual information and the expected distortion was investigated in an information-theoretic framework. The authors in [13] studied the trade-off between the distortion of the estimate of a random variable modeling an electric load and the information leakage quantified by the mutual information. Instead of mutual information, in this paper we use differential privacy, which allows us to evaluate privacy when different users' data are non-stationary, and even without a statistical model of the data. In [10], the authors study the differential privacy of the inputs to a dynamical system when releasing the outputs to a potential adversary, and design a differentially private Kalman filter that protects the input variables. In this paper, the setup is

somewhat similar but we do not consider a dynamical system and we do not consider a dynamical system and we do analyze the case when the filter has access to side information, beyond the control of the inputs. In [11], the authors evaluate the cost of differential privacy in a system of agents that try to estimate their environment in a distributed manner. In our work, the estimation is done by a central entity instead of the agents themselves.

The rest of the paper is organized as follows. Section II describes our model and the problem formulation. Section III defines differential privacy in the context of state estimation. Section IV provides optimal state estimation algorithms under various assumptions and bounds on their accuracy. Section V analyzes the trade-off between privacy and estimation error, and Section VI concludes the paper.

II. MODELING AND PROBLEM FORMULATION

Operators of low- and medium-voltage power distribution networks today typically only have access to a relatively small number of real-time measurements [2], [3]. It is then not possible to run standard state estimation algorithms as is commonly used in high-voltage transmission grids [1]. In this work, we therefore analyze how low-level measurements of the customers' loads can be integrated in state estimation schemes, while ensuring the customers' privacy.

We consider a single distribution line that is connected to a substation under the control of an operator, see Fig. 1. Since distribution grids are typically radial, this is not a severe restriction. There are N service drops or laterals along the line; at each service drop location j along the line, there are several loads connected in parallel, which draw a total load current L_j. We will here investigate the "value" of a measurement of L_j to the operator. Such a measurement could be realized in a modern distribution power grid by using individual customers' smart meters. For example, the individual customers at location j could aggregate their measurements and send them directly to the operator, in real time. However, for privacy reasons, it may not be desirable for an individual customer to share its instantaneous load. We denote such a customer by C in Fig. 1, and will next investigate how the customer can contribute with measurements and still maintain a quantifiable amount of privacy.

Remark 1: We consider current flows on the lines in this paper, rather than active power flows. The reason is that resistive losses are non-negligible in a distribution grid, and so we choose the currents which are conserved. Also note that the dynamics in distribution grids are very fast, and we only consider the steady state in this first study.

Remark 2: It is not necessary to have a trusted agent at location j to perform the aggregation of the measured load currents, to form the measurement of L_j. In [7], it is shown how such computations can be securely decentralized among the customers.

1) Modeling: In what we call the *Base Scenario*, the operator has only access to a real-time measurement Z_0 of the total current flow I_0 that is being supplied from the
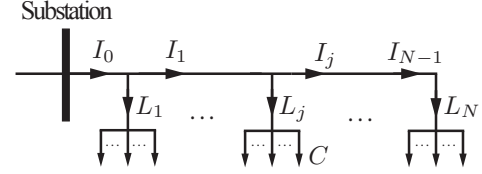

Fig. 1. The considered distribution line. The grid operator can only directly measure the current I_0 at the substation. We analyze how the operator can benefit in the state estimation from communicated measurements of the loads $L_1, L_2, \ldots, L_N$, while maintaining individual customers' privacy (customer denoted by C).

substation,

$$Z_0 = I_0 + W_0, \tag{1}$$

where $W_0 \sim \mathcal{N}(0, R_0)$ is zero-mean Gaussian measurement noise, of variance $R_0 \equiv \sigma_0^2$ (σ_0 is the standard deviation). In the Base Scenario, we also assume the operator has access to a statistical model of the loads $L_1, L_2, \ldots, L_N$ along the line. In the power systems literature, sometimes such data are referred to as *pseudo measurements*. The load vector L is assumed to have a multivariate Gaussian distribution, $L \sim \mathcal{N}(m, P)$, with mean and covariance denoted by

$$m = \begin{pmatrix} m_1 \\ \vdots \\ m_N \end{pmatrix} := \mathbf{E}[L] = \begin{pmatrix} \mathbf{E}[L_1] \\ \vdots \\ \mathbf{E}[L_N] \end{pmatrix}$$

$$P = \begin{pmatrix} P_{11} & \cdots & P_{1N} \\ \vdots & \ddots & \vdots \\ P_{N1} & \cdots & P_{NN} \end{pmatrix} := \mathbf{E}[(L-m)(L-m)^T].$$

The Gaussian distribution of the loads can be justified if we assume each load L_j is an aggregate of a multitude of individual customers as indicated already in Fig. 1. The operator could estimate this statistical model based on historical load data. The following notation will also be used,

$$\begin{pmatrix} P_1 \\ \vdots \\ P_N \end{pmatrix} := \begin{pmatrix} P_{11} + \ldots + P_{1N} \\ \vdots \\ P_{N1} + \ldots + P_{NN} \end{pmatrix} = P\mathbf{1},$$

where $\mathbf{1}$ is column vector of ones.

In this paper, the sole physical model used is based on the *conservation of currents*, *i.e.,*

$$I_j = \sum_{k>j} L_k, \quad j = 0, 1, \ldots, N-1, \tag{2}$$

where I_j are the (unknown) currents along the line. It is clear that the operator may benefit from a more detailed physical model in the state estimation to follow, but we leave this option for future work.

Remark 3: Note that flow conservation models similar to (2) apply to other distribution networks, such as water networks, where mass is a conserved quantity. Hence, the results that follow are applicable also in these scenarios.

Based on the model (2), it follows that the total current I_0 has a Gaussian distribution, $I_0 \sim \mathcal{N}(m_0, P_0)$, where

$$m_0 := m_1 + m_2 + \ldots + m_N, \quad P_0 := \mathbf{1}^T P \mathbf{1}.$$

In what we call the *Smart Meter Scenario*, the operator is also provided real-time measurements Z_j of the loads $L_1, L_2, \ldots, L_N$,

$$Z_j = L_j + W_j, \quad j = 1, 2, \ldots, N, \tag{3}$$

where $W_j \sim \text{Lap}(b_j)$ are independent zero-mean random noise variables with Laplacian distributions of variance $R_j = 2b_j^2$. (The probability density function (PDF) of a zero-mean Laplacian random variable W is $p_W(w) = \frac{1}{2b}e^{-|w|/b}$.) The reason for the Laplacian noise assumption will become clear in Section III.

2) Problem Formulation: Based on the measurement Z_0 and the statistical model of L, it is possible for the operator in the Base Scenario to estimate the loads $L_1, L_2, \ldots, L_N$ and currents $I_1, I_2, \ldots, I_{N-1}$ along the line, see Section IV. With such estimates, one can monitor the status of the line and detect faults in real time. However, when the load uncertainty is large (the covariance P is large), the quality of this estimate is low.

The main problem considered in this paper is how to integrate the load measurements $Z_1, Z_2, \ldots, Z_N$ in the state estimation in the Smart Meter Scenario, and to quantify the improvement of the estimate and the *loss of privacy* that the customers experience.

III. DIFFERENTIAL PRIVACY FOR THE DISTRIBUTION LINE LOADS

In this section, the goal is to quantify the loss of privacy that the customers along the distribution line experience through the introduction of the load measurements $Z_0, Z_1, \ldots, Z_N$.

A. Differential Privacy

To quantify the level of privacy enjoyed by the customers along the line, we employ the notion of *differential privacy*, see [8], [9], [10], [11], for example. We first need to introduce some concepts: Suppose that data from m customers are stored in a vector $d \in \mathbb{R}^m$. We call the vector d a *data vector*. We say two data vectors $d, d' \in \mathbb{R}^m$ are *adjacent*, $\text{Adj}(d, d')$, if and only if they differ in one entry:

for some k, $d_k \neq d_k'$, and $d_l = d_l'$ for all $l \neq k$.

A concrete example is that one customer turns off his smart meter and does not participate in the data collection, transforming d into d'. We also note that $\text{Adj}(\cdot, \cdot)$ is a symmetric binary relation.

We say a measurement $M(\cdot, W) : \mathbb{R}^m \to \mathbb{R}$, where W is a random variable, is (ϵ, δ)-*differentially private* if for all adjacent data vectors $d, d' \in \mathbb{R}^m$, and possible events $E \subseteq \text{Supp}(M(d, W)) \cup \text{Supp}(M(d', W))$, it holds

$$\Pr[M(d, W) \in E] \leq e^\epsilon \Pr[M(d', W) \in E] + \delta. \tag{4}$$

If $\epsilon \geq 0$ and $\delta \geq 0$ are small, this means that the changes in the statistics of the output of M are very small when applied to any adjacent data vectors. Put differently, with access to only the output of M, it is practically impossible to decide whether a single customer has participated in

the measurement, or not, and (ϵ, δ) provides a quantitative certificate of privacy to the customer. If (4) holds with $\delta = 0$, we say M is ϵ-*differentially private*. More details on the interpretation of differential privacy can be found in [8], [9], [10], [11], for example.

To design (ϵ, δ)-differentially private measurements M, we use Theorems 2 and 3 in [10]. First, we suppose the measurements can be written in the general form

$$M(d, W) = q(d) + W,$$

where $q : \mathbb{R}^m \to \mathbb{R}$ is a deterministic *query* that depends on the data vector only. Next, define the *sensitivity* of a query q by

$$S(q) := \sup_{d, d' : \text{Adj}(d, d')} |q(d) - q(d')|.$$

That is, $S(q)$ measures the worst-case change in output of $q(d)$ over adjacent data vectors d. The following proposition then holds.

Proposition 1 ([10]): The measurement $M(d, W) = q(d) + W$ is:

(a) ϵ-differentially private if the random variable W has a Laplace distribution, $W \sim \text{Lap}(b)$, with zero mean and scale parameter $b \geq \frac{S(q)}{\epsilon}$;

(b) (ϵ, δ)-differentially private if the random variable W has a Gaussian distribution, $W \sim \mathcal{N}(0, \sigma^2)$, with zero mean and standard deviation $\sigma \geq \frac{S(q)}{2\epsilon}(K + \sqrt{K^2 + 2\epsilon})$, where $K := K(\delta) = Q^{-1}(\delta)$ and $Q(x) := \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-u^2/2} du$.

Hence, by either adding Laplacian noise or Gaussian noise to a query we obtain differential privacy. Note that Laplacian noise yields better privacy ($\delta = 0$) because its PDF has fatter tails. Thus if customers can choose to add noise, Laplacian noise is the natural selection and justifies the model (3). The noise in (1) is modeled as being Gaussian, because we assume it arises in the physical meter owned by the operator. There is of course also similar physical noise in the customers' smart meters, but for simplicity we assume this is negligible. In any case, introducing more noise sources in the customers' measurements would only increase their privacy.

B. Privacy in the Base Scenario

By providing measurements of the loads L_j, it is clear that the customers will loose some privacy. However, the operator can already measure the total current I_0 by means of (1), beyond the control of the customers. We should first therefore evaluate the loss of privacy through this measurement. Because Z_0 is subjected to Gaussian noise, we use Proposition 1-(b).

The query in this case is the total current, $q_0 = l_1 + l_2 + \ldots + l_N$, where $l_1, l_2, \ldots, l_N$ are realizations of the random variables $L_1, L_2, \ldots, L_N$. An adjacent load pattern is obtained if a single customer C at an arbitrary location j changes his or her load, so that the load goes from l_j to l_j'. If we assume a (uniform) bound on the difference between the

maximum and minimum load current of a single customer is Δ, it easily follows that the sensitivity of q_0 is

$$S(q_0) = \Delta.$$

The differential privacy in the Base Scenario is stated next.

Lemma 1: The measurement Z_0 gives (ϵ_0, δ_0)-differential privacy to the customers, where

$$\epsilon_0 = \frac{\Delta K}{\sigma_0} + \frac{\Delta^2}{2\sigma_0^2}, \quad K = K(\delta_0) = Q^{-1}(\delta_0).$$

Proof: Verify that ϵ_0 is a solution to $\sigma_0 = \frac{\Delta}{2\epsilon_0}(K + \sqrt{K^2 + 2\epsilon_0})$ from Proposition 1-(b). ∎

A typical choice of δ_0 is 0.05, which gives $K \approx 1.64$, or $\delta_0 = 0.01$, which gives $K \approx 2.33$. We note that if the customer size Δ is small compared to the noise σ_0, we have $\epsilon_0 \approx \Delta K/\sigma_0$, which is used in Section V.

Remark 4: Note that differential privacy is independent of the underlying probability distribution of the load pattern L (the data). Hence, no matter if it is really Gaussian, or not, the level of privacy is guaranteed. The statistical model of the data is needed for reliable state estimation, however.

C. Privacy in the Smart Meter Scenario

In the Smart Meter Scenario, the customers provide the measurements Z_j, $j = 1, 2, \ldots, N$ to the operator, which causes a loss of privacy compared to the Base Scenario. In this case, the queries are of the type $q_j = l_j$, where l_j is a realization of L_j. Under the same assumptions as in the previous subsection, we have that

$$S(q_j) = \Delta, \quad j = 1, 2, \ldots, N,$$

where Δ is the uniform bound on a single customer's maximum load current change.

Remark 5: Note that the measurement Z_j sent to the operator is the sum (aggregate) of all the customers' load currents at location j, plus the chosen Laplacian noise. We assume that Δ is a bound that works for every single customer at location j.

By Proposition 1-(a), it is clear that the measurement Z_j alone gives ϵ-differential privacy to every single customer, where $\epsilon = \Delta/b_j$. Hence, by choosing the scaling parameter b_j for the Laplacian noise properly, the customers can tune the level of privacy. However, there is also a loss of privacy through Z_0, and for the composite measurement (Z_0, Z_j) we have the following result.

Proposition 2: The composite measurement (Z_0, Z_j) gives a customer at location j a $(\epsilon_0 + \epsilon, \delta_0)$-differential privacy, when Z_0 is (ϵ_0, δ_0)-differentially private and Z_j is ϵ-differentially private.

Proof: The result follows by a minor adaptation of the proof of Theorem 1 in [14]. ∎

That there is not a dramatic loss of privacy under compositions of differentially private measurements is in fact one of the main advantages of differential privacy, see [14], [9] for more general treatments.

IV. State Estimation of the Distribution Line

In this section, we present algorithms for computing (optimal) state estimates, and we bound their accuracies. To simplify the presentation, we here only consider the estimates of an arbitrary load L_j, and not currents I_j. Because of the linear relation (2), it should be clear how to obtain estimates of the currents from the load estimates.

A. Optimal State Estimation in the Base Scenario

In the Base Scenario, the operator has only access to the measurement Z_0 and the statistical model of the loads L. Since these have a multivariate Gaussian distribution, the optimal state estimate (minimum mean square error (MMSE) estimate) is given by the conditional mean of L_j [15, Example 3.4],

$$\hat{L}_j^0 := \mathbf{E}[L_j|Z_0] = m_j + \frac{P_j}{P_0 + R_0}(Z_0 - m_0), \quad (5)$$

where $P_j = P_{j1} + \ldots + P_{jj} + \ldots + P_{jN}$. The variance of the estimation error is

$$Q_j^0 := \mathbf{E}[(\hat{L}_j^0 - L_j)^2] = P_{jj} - \frac{P_j^2}{P_0 + R_0}, \quad (6)$$

and provides a baseline accuracy indicator for state estimation. It is clear that if the variance of the load L_j (P_{jj}) is small in comparison to the variance of the measurement Z_0, the operator gains very little information through this measurement.

B. MAP State Estimation with Smart Meters

When the load model and the measurements all have a jointly Gaussian distribution, the MMSE estimate and the maximum a posteriori probability (MAP) estimate (see [16], for example) coincide, and are equal to $\mathbf{E}[L_j|Z_0]$. In the Smart Meter Scenario, the random noise $W_1, W_2, \ldots, W_N$ follow a Laplacian distribution, and therefore this is no longer the case, in general. It is well known [15, Theorem 3.1] that the MMSE estimate is still given by the conditional mean $\hat{L}_j^\star := \mathbf{E}[L_j|Z_0, Z_1, \ldots, Z_N]$, but a simple closed-form expression for $\hat{L}_j^\star$ similar to (5) is in this case not known to us. The MAP estimate also does not have a simple closed-form solution, but it is possible to compute it online by means of a simple convex program, as we show next.

The PDF for the load vector L takes the form

$$p_L(l) = \frac{(2\pi)^{-N/2}}{\sqrt{\det P}} e^{-\|P^{-1/2}(l-m)\|_2^2/2},$$

and the conditional PDF for the measurements are

$$p_{Z_0, Z|L}(z_0, z|l) = \frac{1}{\sqrt{2\pi}\sigma_0 (2b)^N} e^{-(z_0 - l_0)^2/(2\sigma_0^2)} e^{-\|z - l\|_1/b},$$

where $Z := (Z_1, Z_2, \ldots, Z_N)^T$ and $l_0 := \mathbf{1}^T l$. Note that $Z_0, Z_1, \ldots, Z_N$ are conditionally independent given L. Given the received measurements z_0 and z, the MAP estimate of L_j can be obtained as

$$\hat{L}^{\text{MAP}}(z_0, z) := \arg\max_l p_{\bar{Z}|L}(z_0, z|l) p_L(l). \quad (7)$$

The following proposition states an alternative, and more explicit form of the estimate.

Proposition 3: The MAP estimate of L given the measurements z_0, z is

$$\hat{L}^{\text{MAP}}(z_0, z) = \arg\min_l \frac{(z_0 - \mathbf{1}^T l)^2}{2\sigma_0^2}$$
$$+ \frac{1}{2}\|P^{-1/2}(l - m)\|_2^2 + \frac{1}{b}\|z - l\|_1.$$

Proof: Follows by a straightforward insertion of the PDFs into (7). ∎

The optimization problem in Proposition 3 is convex since the objective is a sum of 1-norm and 2-norm expressions. Hence, it can easily be solved using standard optimization software. Although this result is useful for an implementation of the state estimator, we are not aware of an accuracy indicator similar to Q_j^0. Because of this, we next turn to an optimal linear estimator, where we can characterize the accuracy in general.

Remark 6: Recently [17] investigated dynamical state estimation problems where the noise was a mix of Gaussian and Laplacian random variables, although the motivation for that work was not related to privacy.

C. LMMSE State Estimation with Smart Meters

For arbitrarily distributed measurements, it turns out that if we limit ourselves to only *linear* estimators when minimizing the mean square error (the LMMSE), we obtain simple closed-form expressions. We will use the following result.

Proposition 4 (Theorem 2.1 [15]): Let the random variable (X, Z) have mean (m_x, m_z) and covariance $\left[\begin{smallmatrix} \Sigma_{xx} & \Sigma_{xz} \\ \Sigma_{yz} & \Sigma_{zz} \end{smallmatrix}\right]$. Then the LMMSE estimate is given by

$$\mathbf{E}^{\text{lin}}[X|Z] := m_x + \Sigma_{xz}\Sigma_{zz}^{-1}(Z - m_z),$$

with estimation error covariance $\Sigma_{xx} - \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}$.

In the Smart Meter Scenario, using Proposition 4 one easily obtains the LMMSE estimate of L_j given all the measurements as

$$\hat{L}_j^{0:N} := \mathbf{E}^{\text{lin}}[L_j|Z_0, Z]$$
$$= m_j + \begin{bmatrix} P_j \\ P_{:j} \end{bmatrix}^T \begin{bmatrix} P_0 + R_0 & \mathbf{1}^T P \\ P\mathbf{1} & P + R \end{bmatrix}^{-1} \begin{bmatrix} Z_0 - m_0 \\ Z - m \end{bmatrix},$$

where $P_{:j}$ is the j-th column of the covariance matrix P. The accuracy of the LMMSE estimate is

$$Q_j^{0:N} := \mathbf{E}[(L_j - \hat{L}_j^{0:N})^2]$$
$$= P_{jj} - \begin{bmatrix} P_j \\ P_{:j} \end{bmatrix}^T \begin{bmatrix} P_0 + R_0 & \mathbf{1}^T P \\ P\mathbf{1} & P + R \end{bmatrix}^{-1} \begin{bmatrix} P_j \\ P_{:j} \end{bmatrix}. \quad (8)$$

It should be clear that the accuracy is not as good as the MMSE estimate $\hat{L}_j^\star$, in general, since we have only optimized over the linear estimators.

To get an even simpler closed-form expression, an LMMSE estimate based on only a few of the measurements is constructed.

Proposition 5: The LMMSE estimate of L_j given the measurements Z_0, Z_j is

$$\hat{L}_j^{0,j} := \mathbf{E}^{\text{lin}}[L_j|Z_0, Z_j]$$
$$= \hat{L}_j^0 + K_j \left[(Z_j - m_j) - \frac{P_j}{R_0 + P_0}(Z_0 - m_0)\right],$$

where $\hat{L}_j^0$ is the Base Scenario MMSE estimate in (5), and

$$K_j = \frac{(R_0 + P_0)P_{jj} - P_j^2}{(R_0 + P_0)(P_{jj} + R_j) - P_j^2} \in [0, 1]. \quad (9)$$

The accuracy of $\hat{L}_j^{0,j}$ is

$$Q_j^{0,j} := \mathbf{E}[(L_j - \hat{L}_j^{0,j})^2] = Q_j^0(1 - K_j) \leq Q_j^0, \quad (10)$$

where Q_j^0 is the accuracy of the Base Scenario MMSE estimate in (6).

Proof: Applying the Schur complement formula to the inverse covariance matrix in the LMMSE estimate

$$\hat{L}_j^{0,j} := \mathbf{E}^{\text{lin}}[L_j|Z_0, Z_j]$$
$$= m_j + \begin{bmatrix} P_j \\ P_{jj} \end{bmatrix}^T \begin{bmatrix} P_0 + R_0 & P_j \\ P_j & P_{jj} + R_j \end{bmatrix}^{-1} \begin{bmatrix} Z_0 - m_0 \\ Z_j - m_j \end{bmatrix}$$

yields the desired expression. After some manipulations of (8), with only the measurements Z_0 and Z_j, the expression in (10) is obtained. ∎

It is seen from Proposition 5 that the variable K_j can both be seen as the gain one should use to fuse in the additional measurement Z_j to the estimate of L_j, and as a measure of the relative improvement in the estimate as compared to the Base Scenario, *i.e.,* $K_j = \frac{Q_j^0 - Q_j^{0,j}}{Q_j^0}$.

The relation between the accuracies of the different estimators is summarized next. It easily follows since more measurements can only improve the quality of an optimal estimation scheme.

Proposition 6: For $j = 1, 2, \ldots, N$, it holds that

$$Q_j^\star \leq Q_j^{0:N} \leq Q_j^{0,j} \leq Q_j^0,$$

where $Q_j^\star := \mathbf{E}[(L_j - \hat{L}_j^\star)^2]$ and $\hat{L}_j^\star$ is the MMSE estimate given measurements Z_0, Z.

The reason we are interested in $\hat{L}_j^{0,j}$ and $Q_j^{0,j}$ is because they are easily computed and provide a simple non-trivial expression for the improvement given by a single load measurement Z_j. The point is that if a more complicated estimate is constructed, such as $\hat{L}_j^\star$ or $\hat{L}_j^{0:N}$, their accuracy will be at least as good as indicated by K_j in (10). This will prove useful in the next section where some insights related to the trade-off between privacy and estimation quality are obtained.

V. TRADE-OFF: ESTIMATION PERFORMANCE VS. PRIVACY LOSS

A customer in the Smart Meter Scenario has $(\epsilon + \epsilon_0, \delta_0)$-differential privacy, following Proposition 2. The customer can increase his or her privacy by increasing the smart meter noise level R_j, but this comes at the expense of the state estimation performance. In this section, we investigate this

trade-off a bit closer, and in particular identify situations where only a minor loss of privacy can lead to a large improvement in the state estimate.

We start by introducing the dimensionless quantities

$$\eta_j := \frac{\Delta^2}{P_{jj}}, \quad \zeta_j := \frac{P_{jj}}{P_0 + R_0}.$$

The variable η_j normalizes the customer's variability with the total load variability at location j. Hence, it is a measure of how large player the customer is in comparison with other loads at location j. The variable ζ_j measures how large the total load variability at location j is in comparison with the overall current variability in the distribution line. Thus if both η_j and ζ_j are small, it means the customer is a relatively small player at location j, and location j is overall responsible for only a small part of the variability on the line.

In the following, we for simplicity assume that the loads are uncorrelated, i.e., that $P_j = P_{jj}$. We use that the variance of the Laplacian measurement noise added to L_j is

$$R_j = \frac{2\Delta^2}{\epsilon^2} = \frac{2\eta_j P_{jj}}{\epsilon^2}.$$

We can now express the relative improvement of the state estimate as (see (9)–(10))

$$
\begin{aligned}
K_j &= \frac{(R_0 + P_0)P_{jj} - P_j^2}{(R_0 + P_0)(P_{jj} + R_j) - P_j^2} \\
&= \frac{1}{1 + \frac{R_j(R_0+P_0)}{(R_0+P_0)P_{jj} - P_{jj}^2}} \\
&= \frac{1}{1 + \frac{2\eta_j}{\epsilon^2(1-\zeta_j)}} \approx \frac{\epsilon^2(1-\zeta_j)}{2\eta_j},
\end{aligned}
\tag{11}
$$

where the last approximation holds for small ϵ. The total differential privacy for the customer is $(\epsilon + \epsilon_0, \delta_0)$ and ϵ is a measure of how much extra privacy the customer at location j gives up by sharing his or her load measurement. The customer's differential privacy can be no better than (ϵ_0, δ_0), and using Lemma 1 this lower bound is

$$\epsilon_0^2 \approx \frac{\Delta^2 K^2}{R_0} = \eta_j \zeta_j K^2 \left(1 + \frac{P_0}{R_0}\right). \tag{12}$$

Not surprisingly, small η_j and ζ_j assure a high level of privacy to the customer in the Base Scenario because his or her actions are barely noticeable in Z_0.

In Figs. 2 and 3, we plot the lower bound on relative improvement (11) as a function of the total loss of privacy, for fixed ζ_j and η_j, respectively. The values of η_j and ζ_j are chosen to give comparable values for the lower bounds ϵ_0 (indicated by circles in Figs. 2 and 3). It is interesting to see that the improvement curves increase faster for small values of ϵ_0, especially as the customer's variability η_j decreases. This is also clearly seen in the quadratic approximation (11), whose curvature is $(1-\zeta_j)/(2\eta_j)$. Combining (11) and (12),

$$K_j \approx \frac{\epsilon^2}{2\eta_j}(1 - \zeta_j) \approx \frac{K^2}{2}\left(1 + \frac{P_0}{R_0}\right)\zeta_j(1 - \zeta_j)\left(\frac{\epsilon}{\epsilon_0}\right)^2 \tag{13}$$
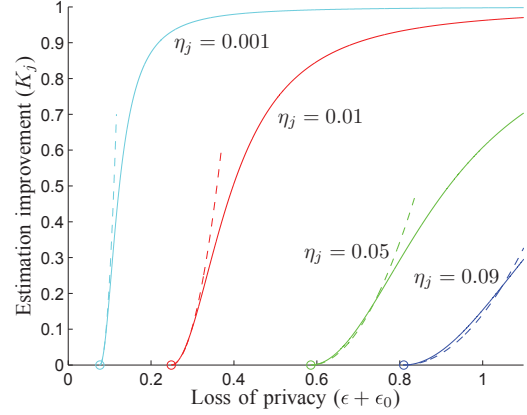

Fig. 2. Lower bound on the relative improvement of the state estimate in the Smart Meter Scenario (K_j), as a function of total loss of customer privacy, when $P_0 = 1$, $R_0 = 0.05$, $\delta_0 = 0.05$, $\zeta_j = 0.1$, and varying customer variability η_j. It is seen that the curvature dramatically increases for customers who have a high level of privacy in the Base Scenario (small ϵ_0).
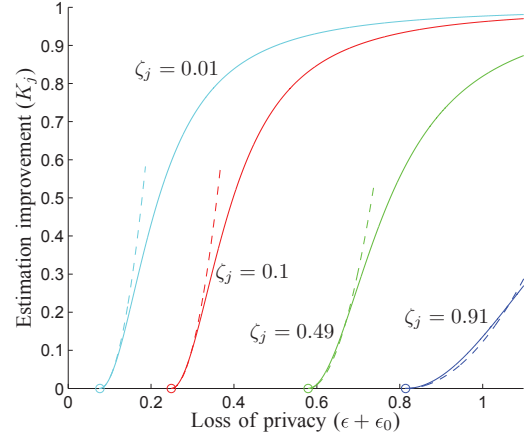

Fig. 3. Lower bound on the relative improvement of the state estimate in the Smart Meter Scenario (K_j), as a function of total loss of customer privacy, when $P_0 = 1$, $R_0 = 0.05$, $\delta_0 = 0.05$, $\eta_j = 0.01$, and varying load variability ζ_j. It is seen that the curvature increases for customers who have a higher level of privacy in the Base Scenario (small ϵ_0).

we also directly see that a small ϵ_0 increases the curvature. (The quadratic approximations are indicated by dashed lines in Figs. 2 and 3.) The importance of this observation is that a customer who enjoys a high level of differential privacy in the Base Scenario can dramatically improve the state estimation performance by only giving up a little bit extra of privacy. For example, in Fig. 2, we see that if a customer is willing to increase $\epsilon_0 + \epsilon$ from 0.25 to 0.35 when $\eta_j = 0.01$, there is at least around 30% improvement in estimation quality. On the other hand, the smart meter measurements of a customer who already has a low level of privacy in the Base Scenario are of much lower value. The reason is that such customers are already possible to estimate quite accurately using the measurement Z_0 alone.

VI. Conclusion

We have proposed a modeling framework to analyze the differential privacy for customers equipped with smart meters in a radial distribution power grid. Several different state estimation schemes were proposed, and an interesting trade-off between the utility for the operator and the loss of privacy for the customer was identified. The analysis revealed that aggregated measurements from small customers can significantly improve the operator's state estimate at a small loss of privacy.

It should be noted that in the current framework it is not possible for the customer to influence the baseline (ϵ_0, δ_0)-differential privacy. This would be possible if the customer has access to a local controllable energy storage, see [18], for example. Open problems for future research also include more detailed modeling of the distribution grid and quantification of the accuracy of the MAP estimate.

References

[1] A. Abur and A. Exposito, *Power System State Estimation: Theory and Implementation*. Marcel-Dekker, 2004.

[2] W. H. Kersting, *Distribution System Modeling and Analysis, Second Edition (Electric Power Engineering Series)*, 2nd ed. CRC Press, Nov. 2006.

[3] A. Abdel-Majeed and M. Braun, "Low voltage system state estimation using smart meters," in *Universities Power Engineering Conference (UPEC), 2012 47th International*, Sept 2012, pp. 1–6.

[4] A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet, and D. Irwin, "Private memoirs of a smart meter," in *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, ser. BuildSys '10, 2010, pp. 61–66.

[5] J. Bohli, C. Sorge, and O. Ugus, "A privacy model for smart metering," in *Communications Workshops (ICC), 2010 IEEE International Conference on*. IEEE, 2010, pp. 1–5.

[6] F. Li, B. Luo, and P. Liu, "Secure and privacy-preserving information aggregation for smart grids," *Int. J. Secur. Netw.*, vol. 6, no. 1, pp. 28–39, Apr. 2011.

[7] G. Ács and C. Castelluccia, "I have a DREAM! (DiffeRentially privatE smArt Metering)," in *Information Hiding*, ser. Lecture Notes in Computer Science, T. Filler, T. Pevný, S. Craver, and A. Ker, Eds. Springer Berlin Heidelberg, 2011, vol. 6958, pp. 118–132. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-24178-9_9

[8] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of Cryptography, Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006, Proceedings*, 2006, pp. 265–284.

[9] C. Dwork, G. N. Rothblum, and S. P. Vadhan, "Boosting and differential privacy," in *51th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2010, October 23-26, 2010, Las Vegas, Nevada, USA*, 2010, pp. 51–60.

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

[11] Z. Huang, Y. Wang, S. Mitra, and G. E. Dullerud, "On the cost of differential privacy in distributed control systems," in *Proceedings of the 3rd International Conference on High Confidence Networked Systems*, ser. HiCoNS '14. New York, NY, USA: ACM, 2014, pp. 105–114.

[12] L. Sankar, S. Kar, R. Tandon, and H. V. Poor, "Competitive privacy in the smart grid: An information-theoretic approach," in *Smart Grid Communications (SmartGridComm), 2011 IEEE International Conference on*, Oct 2011, pp. 220–225.

[13] L. Sankar, S. Rajagopalan, S. Mohajer, and H. V. Poor, "Smart meter privacy: A theoretical framework," *Smart Grid, IEEE Transactions on*, vol. 4, no. 2, pp. 837–846, June 2013.

[14] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, "Our data, ourselves: Privacy via distributed noise generation," in *Advances in Cryptology (EUROCRYPT 2006)*, ser. Lecture Notes in Computer Science, vol. 4004. Saint Petersburg, Russia: Springer Verlag, May 2006, pp. 486–503.

[15] B. D. O. Anderson and J. B. Moore, *Optimal Filtering (Dover Books on Engineering)*. Dover Publications, Jan. 2005.

[16] S. Saha, P. Mandal, A. Bagchi, Y. Boers, and J. Driessen, "Particle based smoothed marginal MAP estimation for general state space models," *Signal Processing, IEEE Transactions on*, vol. 61, no. 2, pp. 264–273, Jan 2013.

[17] L. Carrillo, W. Russell, J. Hespanha, and G. Collins, "State estimation of multiagent systems under impulsive noise and disturbances," *Control Systems Technology, IEEE Transactions on*, vol. 23, no. 1, pp. 13–26, Jan 2015.

[18] O. Tan, D. Gunduz, and H. V. Poor, "Increasing smart meter privacy through energy harvesting and storage devices," *Selected Areas in Communications, IEEE Journal on*, vol. 31, no. 7, pp. 1331–1341, July 2013.