Distribution Grid Line Outage Detection with Privacy Data

Chenhan Xiao, Yizheng Liao, Yang Weng
Department of Electrical, Computer and Energy Engineering
Arizona State University, Tempe, Arizona, United States
Email: {cxiao20, yzliao, yang.weng}@asu.edu

***Abstract*—Change point detection is important for many real-world applications. While sensor readings enable line outage identification, they bring privacy concerns by allowing an adversary to divulge sensitive information such as household occupancy and economic status. In this paper, to preserve privacy, we develop a decentralized randomizing scheme to ensure no direct exposure of each user's raw data. Brought by the randomizing scheme, the trade-off between privacy gain and degradation of change point detection performance is quantified via studying the differential privacy framework and the Kullback–Leibler divergence. Furthermore, we propose a novel statistic to mitigate the impact of randomness, making our detection procedure both privacy-preserving and have optimal performance. The results of comprehensive experiments show that our proposed framework can effectively find the outage with privacy guarantees.**

I. Introduction

Distribution grid line outage detection is important for efficient system monitoring and system control in smart grids, for restoring the network stability [1] and reducing financial loss. Recently, smart meters with advanced metering infrastructure (AMI) and fault location, isolation, and service restoration (FLISR) systems were installed to report the outage when there is a loss of power [2]. But these methods are limited when customers still receive power after the line outage from distributed energy resources, which are increasingly penetrated nowadays. To detect these kinds of line outages, real-time grid readings are utilized, including voltage magnitudes, phasor angles and load estimates [3]–[8].

However, the utilization of grid readings brings privacy concerns, i.e., the leakage of sensitive information. For instance, given user's time-series grid readings, an untrusted third-party can discern the usage of appliances [9], and divulge the household occupancy and economic status [10]–[12] through the non-intrusive load monitoring technique. Thus, it calls for protecting such readings from being directly exposed to a third party while maintaining the performance of outage detection.

We consider the outage detection procedure from our previous work [7], [8], which has theoretical guarantees for the detection performance but requires the user's meter readings. Specifically, the voltage magnitudes are collected to find the line outage. The increment of voltage magnitudes in the distribution grid was proven to follow multivariate Gaussian distributions before and after the line outage. Then, the outage can be identified by detecting the change in the data distribution under the change point detection framework. This framework aims to find the change in data distribution as quickly as possible under the constraint of false alarm tolerance [13]. It has been widely used in line outage and fault detection in transmission and distribution grids [14]–[16].

In this paper, nevertheless, the smart meter readings may be exposed to an untrusted third party and serve to invade people's privacy [17]. Specifically, the popular linear coupled power flow model indicates that voltage magnitude is coupled with power consumption. Thus, the voltage magnitude can reflect the energy consumption or production billing information to a certain extent, causing privacy leakage issues [18].

To protect data privacy in change point detection, randomizing schemes are developed to "encrypt" the data, hiding sensitive information from potential attackers. For example, [19] applied the report noisy max algorithm by adding noise to partial log-likelihood ratios to estimate the change time. [20] introduced noise to the test statistic and privately estimated the change points using the Mann-Whitney test. However, these works "encrypted" the test statistic after raw user data was collected, which is crucial to user privacy in distribution grids. Another challenge of existing work is the compromised detection performance despite the privacy guarantee. To our best knowledge, protecting the privacy of raw user data without compromising the detecting performance still lies out of reach of existing theory, which is the focus of our paper.

To guarantee no exposure of raw grid readings, we design a decentralized scheme to directly "encrypt" each user's raw readings. Then, the proposed scheme is shown to satisfy the differential privacy [21], which is a commonly used framework to evaluate the privacy gain. Despite the privacy guarantee, detection performance is also degraded due to randomized data. Specifically, we show a prolonged detection delay is induced by studying the Kullback–Leibler divergence between data distributions before and after the line outage. These analytical studies allow us to answer the question: given a desired level of privacy, how much detection performance will be degraded? Finally, to mitigate the degradation of detection performance, we propose a novel statistic by considering an unbiased estimation of the noise-free optimal statistic. The proposed statistic is shown to have a detection delay proximate to the noise-free optimal case with the constraint of the false alarm rate. In doing so, our detecting procedure can be both privacy-preserving and have comparable detection performance as the optimal case.

The rest of the paper is organized as follows. Section II models the line outage detection problem via voltage data. Section III proposes a randomizing scheme to protect raw user data. Section IV evaluates the proposed method using representative grid systems and real residential load profiles.

II. System Model: Line Outage Detection

For showing our decentralized approach to protect each user's privacy, we model the distribution grid as a graph $\mathcal{G} := \{1, 2, \cdots, M\}$ containing $M > 0$ buses (users). As mentioned earlier, we consider the outage detection procedure from our previous work [7], [8], which utilizes the voltage magnitudes from each bus $i \in \mathcal{G}$ to detect the outage. The reason for using this detection procedure is threefold. First, it has an analysis framework and theoretical guarantees for detection performance, enabling us to develop a privacy-preserving approach on its foundation. Second, the voltage magnitude from smart meters is easy-to-acquire since over 107 million smart meters were deployed by 2021, covering 75% of U.S. households. Third, the increment of voltage data are shown to follow Gaussian distributions. It allows us to quantify the privacy gain and performance degradation once we add Gaussian noise to the voltage increments data.

The voltage magnitude from each bus $i \in \mathcal{G}$ is modeled as a random variable V_i, and the collection of them in the entire grid is modeled as a random vector $\mathbf{V}_{\mathcal{G}} := [V_1, V_2, \cdots, V_M]^T$. Since $\mathbf{V}_{\mathcal{G}}$ usually do not follow a common distribution [7], we further model the increment change of voltage data as $\Delta\mathbf{V}_{\mathcal{G}}$. As a time-series, its realization at time step n is given by $\Delta\mathbf{v}[n] = \mathbf{v}[n] - \mathbf{v}[n-1] \in \mathcal{R}^M$. For the sake of simplicity, we also use the notation $\Delta\mathbf{v}^{1:N} = \{\Delta\mathbf{v}[1], \cdots, \Delta\mathbf{v}[N]\}$ to represent the data up to time N. Given the voltage data $\Delta\mathbf{v}^{1:N}$, we aim to find the outage time while protecting the raw data.

As shown in [7], [22], the increment of voltage data $\Delta\mathbf{V}_{\mathcal{G}}$ follows Gaussian distribution, and such distribution is sensitive to the line outages. Suppose the outage occurs at time $\lambda \in \mathbb{N}$, we can write the sequence of voltage increments as

$$\begin{cases}\Delta\mathbf{v}[n] \overset{i.i.d}{\sim} g : \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0), & n = \quad 1, 2, \cdots, \lambda - 1,\\ \Delta\mathbf{v}[n] \overset{i.i.d}{\sim} f : \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), & n = \quad \lambda, \lambda + 1, \cdots, N,\end{cases} \tag{1}$$

where g denotes the pre-outage distribution and f denotes the post-outage distribution. In these distributions, $\boldsymbol{\mu}_0, \boldsymbol{\mu}_1$ are mean vectors, and $\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}_1$ serve as covariance matrices. Even when the post-outage distribution parameters $(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ are unknown, we have data-driven techniques to learn them [7].

Detecting outage time λ is equivalent to finding when the distribution changes. To achieve this, at each time step N, we compute the posterior probability ratio $\frac{\mathbb{P}(\lambda \leq N | \Delta\mathbf{v}^{1:N})}{\mathbb{P}(\lambda > N | \Delta\mathbf{v}^{1:N})}$ as

$$\Lambda(\Delta\mathbf{v}^{1:N}) = \sum_{k=1}^{N} \frac{\pi(k)}{\sum_{s=N+1}^{\infty} \pi(s)} \prod_{n=k}^{N} \frac{f(\Delta\mathbf{v}[n])}{g(\Delta\mathbf{v}[n])}, \tag{2}$$

where π denotes the prior distribution of λ. We declare the outage once the ratio (2) surpasses a predefined threshold. By the Shiryaev-Roberts-Pollaks procedure [13], [23], the following threshold in Theorem 1 optimally considers the trade-off between the false alarm and the detection delay.

Theorem 1. *When λ follows a geometric prior Geo(ρ), we declare the outage once the posterior probability ratio $\Lambda(\Delta\mathbf{v}^{1:N})$ surpasses the threshold $B_{\rho,\alpha} = (1-\alpha)/(\rho\alpha)$ as*

$$\tau = \inf\{N \in \mathbb{N} : \Lambda(\Delta\mathbf{v}^{1:N}) \geq B_{\rho,\alpha}\}, \tag{3}$$

where the false alarm rate $\mathbb{P}(\tau < \lambda)$ is upper bounded by the given α. As $\alpha \to 0$, τ in (3) is asymptotically optimal for minimizing the average detection delay $\mathbb{E}[\tau - \lambda | \tau \geq \lambda]$ as

$$\inf_{\mathbb{P}(\tau^* \leq \lambda) \leq \alpha} \mathbb{E}[\tau^* - \lambda | \tau^* \geq \lambda] = \frac{|\log \alpha|}{-\log(1-\rho) + D_{KL}(f||g)}, \tag{4}$$

where $D_{KL}(f||g)$ is the Kullback–Leibler divergence.

III. Outage Detection with Privacy Guarantee

In the aforementioned outage identification procedure (3), the increments of voltage magnitude data are critical. However, such readings may also be used to infer the household occupancy, e.g., when the owner arrives or leaves home. To protect raw voltage data of users, at each time step n when data $\Delta\mathbf{v}[n]$ is received, we apply a randomizing scheme

$$\Delta\tilde{\mathbf{v}}[n] = \Delta\mathbf{v}[n] + \mathbf{e}[n], \tag{5}$$

where $\mathbf{e}[n] \in \mathcal{R}^M$ is a random noise vector. The noise $\mathbf{e}[n]$ has to be sufficiently large to hide the characteristics of the raw data while not being too large to impact the detection performance. In determining an appropriate amount of noise, we quantify the privacy gain under the differential privacy framework in Section III-A, and analyze how the detection performance is degraded in Section III-B. We show that in general, the noise added to data makes it harder to distinguish whether the data comes from distribution g or f, leading to a prolonged detection delay. Integrating these analyses, we propose a new statistic in Section III-C (to replace (2)) such that the new detection procedure is both privacy-preserving and has comparable detection performance as the optimal case.

For making the scheme (5) satisfy the differential privacy, we generate noise from the same distribution family (Gaussian) as raw data, i.e., $\mathbf{e}[n] \sim \mathcal{N}(\mathbf{0}, \boldsymbol{D}_e)$. The covariance matrix is designed to be diagonal, i.e., $\boldsymbol{D}_e = \text{diag}(\sigma_e^2, \cdots, \sigma_e^2)$ where variance σ_e^2 represents the noise level or amount of noise. A diagonal covariance indicates that each element in the noise vector is independent. In doing so, the scheme (5) is equivalent to adding a random noise scalar to each dimension of the data vector, ensuring that each user's raw data is "encrypted" before sending to any third party (see Fig. 1). Notice, unlike some works that add noise to the statistics (e.g., $\Lambda(\Delta\mathbf{v}^{1:N})$) [19], [20] after raw data is collected, our approach ensures no direct exposure of the raw data.

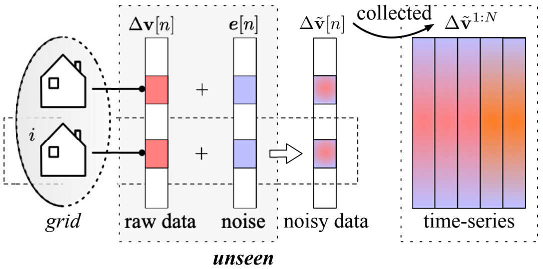

Fig. 1: The decentralized randomizing scheme (5) to protect privacy, where the detection procedure was performed on noisy data $\Delta\tilde{\mathbf{v}}^{1:N}$.

A. Differential Privacy of the Randomizing Scheme

Applying the randomizing scheme (5), the detection procedure was performed on the noisy data $\Delta\tilde{\mathbf{v}}^{1:N}$ to find the outage time. In this subsection, we quantify how much privacy is preserved w.r.t. the noise level σ_e^2. To achieve this, we prove that our scheme (5) satisfies the classic (ε, δ)-differential privacy mechanism [21]. A differential privacy scheme indicates that by looking at the observed data $\Delta\tilde{\mathbf{v}}[n]$, an adversary struggles to tell whether any user's real data $\Delta\mathbf{v}_i[n]$ is included or not.

Since our noise term $\mathbf{e}[n]$ is Gaussian as raw data, the randomized data $\Delta\tilde{\mathbf{v}}[n]$ also follows Gaussian, allowing us to detour the proof of classic differential privacy by the tool of Gaussian differential privacy [24], which extends differential privacy to broader distributions. Specifically, a G_μ-(Gaussian) differential private scheme indicates that distinguishing if any user's real data $\Delta \mathbf{v}_i[n]$ is included in noisy data $\Delta \tilde{\mathbf{v}}[n]$ is more difficult than distinguishing distributions $\mathcal{N}(0,1)$ and $\mathcal{N}(\mu,1)$. To measure the "difficulty", the authors use the trade-off function between type I and type II errors in distinguishing two distributions. Type I error indicates a false inference of a fake distribution, while type II error indicates a failed inference of a real distribution. The particular trade-off function $T(\mathcal{N}(0,1),\mathcal{N}(\mu,1))$ is also denoted as G_μ.

Theorem 2. *The randomizing scheme* (5) *is G_{s/σ_e}- Gaussian differential private where $s := \sup_{n,\Delta\mathbf{v}[n],\Delta\mathbf{v}'[n]} \|\Delta\mathbf{v}[n] - \Delta\mathbf{v}'[n]\|$ is the sensitivity of raw voltage data.*

Proof. The noisy data $\Delta\tilde{\mathbf{v}}[n]$ and its neighboring data $\Delta\tilde{\mathbf{v}}'[n]$ (i.e., they differ in exactly one element) both follow Gaussian distribution as $\Delta\tilde{\mathbf{v}}[n] \sim \mathcal{N}(\Delta\mathbf{v}[n], \boldsymbol{D}_e)$ and $\Delta\tilde{\mathbf{v}}'[n] \sim \mathcal{N}(\Delta\mathbf{v}'[n], \boldsymbol{D}_e)$. Then, we have

$$\begin{aligned} T\left(\Delta\tilde{\mathbf{v}}[n], \Delta\tilde{\mathbf{v}}'[n]\right) &= T(\mathcal{N}(\Delta\mathbf{v}[n], \boldsymbol{D}_e), \mathcal{N}(\Delta\mathbf{v}'[n], \boldsymbol{D}_e)) \\ &= G_{\|\Delta\mathbf{v}[n]-\Delta\mathbf{v}'[n]\|/\sigma_e} \geq G_{\frac{s}{\sigma_e}}, \end{aligned} \tag{6}$$

where $T(\Delta\tilde{\mathbf{v}}[n], \Delta\tilde{\mathbf{v}}'[n])$ is defined as the trade-off function between type I and II errors in differentiating data $\Delta\tilde{\mathbf{v}}[n]$ and $\Delta\tilde{\mathbf{v}}'[n]$. The inequality is due to the definition of sensitivity, i.e., $\|(\Delta\mathbf{v}[n] - \Delta\mathbf{v}'[n])/\sigma_e\| \leq s/\sigma_e$. □

In Theorem 2, the sensitivity of voltage magnitude can be calculated according to its standard operating limit (between 0 $p.u.$ and 1.1 $p.u.$). Provided the Gaussian differential privacy, our proposed scheme (5) satisfies the corresponding $(\varepsilon, \delta(\varepsilon))$-differential privacy where $\delta(\varepsilon) = \Phi(-\frac{\varepsilon\sigma_e}{s} + \frac{s}{2\sigma_e}) - e^{\varepsilon}\Phi(-\frac{\varepsilon\sigma_e}{s} - \frac{s}{2\sigma_e})$ [24]. This ensures that an adversary can not easily determine if the data he observes ($\Delta\tilde{\mathbf{v}}[n]$) is real user data, thus preserving privacy. Moreover, we can control the level of noise to achieve any desired level of privacy guarantee.

B. Analysis of Detection Performance Degradation

Despite the privacy protection brought by the randomizing scheme (5), we need to understand how much detection performance is degraded of using noisy data $\Delta\tilde{\mathbf{v}}[n]$.

To study the performance degradation, it is important to notice that the noisy data $\Delta\tilde{\mathbf{v}}[n]$ follows a distribution that is combined from the distribution of raw data and the distribution of noise. We denote such "noisy" distribution by $g_e \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0 + \boldsymbol{D}_e)$ before the line outage and by $f_e \sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1 + \boldsymbol{D}_e)$ after the line outage. In Theorem 3, we show that after adding noise, distributions g_e and f_e become "closer" than g and f by evaluating the corresponding KL divergence. The "closer" the distributions are, the more difficult to distinguish them when detecting the outage.

Theorem 3. *The randomizing scheme (5) reduces the KL divergence between pre- and post-outage distributions:*

$$\begin{aligned} &KL_\Delta := D_{KL}(f\|g) - D_{KL}(f_e\|g_e) \geq 0, \\ &KL_\Delta \leq \mathcal{O}(\sigma_e^2)(\|\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1\|_2^2 + \frac{(tr(\boldsymbol{\Sigma}_1) - tr(\boldsymbol{\Sigma}_0))^2}{tr(\boldsymbol{\Sigma}_1)}). \end{aligned} \tag{7}$$

Proof. For showing $\mathrm{KL}_\Delta \geq 0$, we have

$$\begin{aligned} \mathrm{KL}_\Delta &= \frac{1}{2}(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)^T[(\boldsymbol{\Sigma}_0)^{-1} - (\boldsymbol{\Sigma}_0^e)^{-1}](\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1) \\ &+ \frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_0|}{|\boldsymbol{\Sigma}_1|}\frac{|\boldsymbol{\Sigma}_1^e|}{|\boldsymbol{\Sigma}_0^e|} + \frac{1}{2}\mathrm{tr}\{(\boldsymbol{\Sigma}_0)^{-1}(\boldsymbol{\Sigma}_1) - (\boldsymbol{\Sigma}_0^e)^{-1}(\boldsymbol{\Sigma}_1^e)\} \\ &\geq \frac{1}{2}\sum\nolimits_{i=1}^{M}\left[(\nu_i - \log\nu_i) - (\xi_i - \log\xi_i)\right], \end{aligned}$$

where $\boldsymbol{\Sigma}_i^e = \boldsymbol{\Sigma}_i + \boldsymbol{D}_e$ for $i = 0, 1$. $\nu_1, \cdots, \nu_M$ and $\xi_1, \cdots, \xi_M$ are the eigenvalues of $(\boldsymbol{\Sigma}_0)^{-1}\boldsymbol{\Sigma}_1$ and $(\boldsymbol{\Sigma}_0^e)^{-1}\boldsymbol{\Sigma}_1^e$, respectively. The inequality is due to that matrix $(\boldsymbol{\Sigma}_0)^{-1} - (\boldsymbol{\Sigma}_0^e)^{-1}$ is positive semi-definite. Moreover, since $|\xi_i - 1| \leq |\nu_i - 1|, \forall i = 1, \cdots, M$, we obtain $\mathrm{KL}_\Delta \geq 0$. Aside from the lower bound as zero, an upper bound of KL_Δ is further derived as

$$\mathrm{KL}_\Delta \leq \frac{\sigma_e^2}{2(\nu_0^{\min})^2}(\|\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1\|_2^2 + M\frac{(\nu_1^{\max} - \nu_0^{\min})^2}{\nu_1^{\max}}),$$

where $\nu_0^{\min}$ is the smallest eigenvalue of $\boldsymbol{\Sigma}_0$, and $\nu_1^{\max}$ is the largest eigenvalue of $\boldsymbol{\Sigma}_1$. □

As a corollary of $\mathrm{KL}_\Delta \geq 0$ in Theorem 3, the lower bound of detection delay in (4) is increased when the privacy-preserving scheme (5) is utilized, resulting in a performance degradation. Theorem 3 not only indicates a strict performance degradation but also infers the magnitude of this degradation by deriving the upper bound of KL_Δ. That is, we know approximately how much extra delay is brought by the noise at an amount of σ_e^2 as noise variance.

C. A New Statistic to Mitigation Noise Impact

In this subsection, we further propose a noise-mitigation technique to obtain a comparable detection performance as the noise-free case while still preserving privacy. To achieve this, we design a new posterior probability ratio $\Lambda^*(\Delta\tilde{\mathbf{v}}^{1:N})$ given the observed noisy data $\Delta\tilde{\mathbf{v}}^{1:N}$. The new statistic should provide an accurate approximation of the noise-free statistic $\Lambda(\Delta\mathbf{v}^{1:N})$ in (2) since the latter is shown in Theorem 1 to have optimal detection performance.

For designing a statistic $\Lambda^*(\Delta\tilde{\mathbf{v}}^{1:N})$ to approximate $\Lambda(\Delta\mathbf{v}^{1:N})$, we notice that although noise is randomly generated, the pattern of noise, i.e., its distribution parameter σ_e, is known. Given this knowledge, it motivates us to calculate the expectation of the noise-corresponding terms in $\Lambda(\Delta\mathbf{v}^{1:N})$. We can replace these terms with their expectations that provide an unbiased estimation. Following this motivation, the new design of the statistic $\Lambda^*(\Delta\tilde{\mathbf{v}}^{1:N})$ is given in (8):

$$\Lambda^*(\Delta\tilde{\mathbf{v}}^{1:N}) = \sum_{k=1}^{N}\frac{\pi_k}{\sum_{s=N+1}^{\infty}\pi_s}\prod_{n=k}^{N}\frac{\sqrt{|\boldsymbol{\Sigma}_0|}\exp\left(\beta_1[n]\right)}{\sqrt{|\boldsymbol{\Sigma}_1|}\exp\left(\beta_0[n]\right)}, \tag{8}$$

where $\beta_i[n] := -\frac{1}{2}(\Delta\mathbf{v}[n] - \boldsymbol{\mu}_i)^T(\boldsymbol{\Sigma}_i)^{-1}(\Delta\mathbf{v}[n] - \boldsymbol{\mu}_i) + \frac{1}{2}\sigma_e \cdot \mathrm{tr}(\boldsymbol{\Sigma}_i^{-1})$ for $i = 0, 1$. To show that our proposed statistic $\Lambda^*(\Delta\tilde{\mathbf{v}}^{1:N})$ enables comparable detection performance as the noise-free case, we argue that $\Lambda^*(\Delta\tilde{\mathbf{v}}^{1:N})$ "best" approximates the noise-free statistic $\Lambda(\Delta\mathbf{v}^{1:N})$. It is verified in Lemma 1.

Lemma 1. *The statistic $\Lambda^*(\Delta\tilde{\mathbf{v}}^{1:N})$ serves as an unbiased estimation of the optimal noise-free statistic $\Lambda(\Delta\mathbf{v}^{1:N})$.*

Proof. We show that $\beta_i[n]$ is an unbiased estimation of the corresponding term in the statistic $\Lambda(\Delta\mathbf{v}^{1:N})$, i.e., $\mathbb{E}_{\mathbf{e}\sim\mathcal{N}(\mathbf{0},\boldsymbol{D}_e)}\beta_i[n] = -\frac{1}{2}(\Delta\mathbf{v}[n]-\boldsymbol{\mu}_i)^T(\boldsymbol{\Sigma}_i)^{-1}(\Delta\mathbf{v}[n]-\boldsymbol{\mu}_i)$. Based on $\mathbb{E}_{\mathbf{e}[n]\sim\mathcal{N}(\mathbf{0},\boldsymbol{D}_e)}\tilde{\beta}_i[n] = \beta[n], \forall i = 0,1, \forall n$, we can derive $\mathbb{E}_{\mathbf{e}[n]\sim\mathcal{N}(\mathbf{0},\boldsymbol{D}_e)}\Lambda^*(\Delta\tilde{\mathbf{v}}^{1:N}) = \Lambda(\Delta\mathbf{v}^{1:N})$. □

Compared to statistic $\Lambda(\Delta\mathbf{v}^{1:N})$ and $\Lambda(\Delta\tilde{\mathbf{v}}^{1:N})$, the new statistic $\Lambda^*(\Delta\tilde{\mathbf{v}}^{1:N})$ has two advantages (illustrated in Fig. 2): (1) since it also uses noisy data to perform the outage detection, the raw data is still protected from exposure, and (2) since it provides an unbiased estimation of the noise-free statistic, we can achieve comparable lower bound of detection delay. Using the statistic $\Lambda^*(\Delta\tilde{\mathbf{v}}^{1:N})$, we can both preserve privacy while not sacrificing detection performance. The detection procedure is summarized in Algorithm 1.

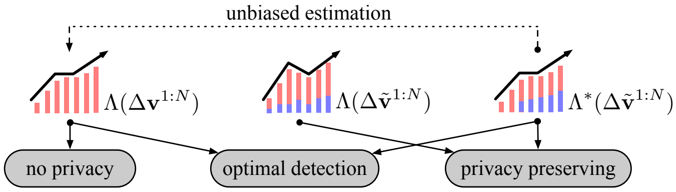
unbiased estimation
$\Lambda(\Delta\mathbf{v}^{1:N})$
$\Lambda(\Delta\tilde{\mathbf{v}}^{1:N})$
$\Lambda^*(\Delta\tilde{\mathbf{v}}^{1:N})$
no privacy
optimal detection
privacy preserving

Fig. 2: The noise mitigation design to achieve both optimal detection performance and privacy-preserving effect.

Algorithm 1 Private Outage Detection (POD) with privacy guarantee. It takes as input a stream of voltage data $\Delta\mathbf{v}[1], \Delta\mathbf{v}[2], \ldots$ and is parameterized by noise variance σ_e^2

```
Input: New observation $\Delta\mathbf{v}[N]$
Apply noise to protect raw data
$\Delta\tilde{\mathbf{v}}[N] = \Delta\mathbf{v}[N] + \mathbf{e}[N],\ \mathbf{e}[N] \sim \mathcal{N}(\mathbf{0}, \mathrm{diag}(\sigma_e^2, \cdots))$
Calculate statistic $\Lambda^*(\Delta\tilde{\mathbf{v}}^{1:N})$ in (8)
if $\Lambda^*(\Delta\tilde{\mathbf{v}}^{1:N}) \geq B_{\rho,\alpha}$ then
    report outage time $N$
end if
```

IV. Numerical Simulations and Results

This section shows how our proposed method performs in distribution grids with real-world data. Various experiments are designed on the balanced distribution systems, including IEEE 8-bus and IEEE 123-bus networks, a 36-bus network in the urban area, and a 115-bus network in the suburban area from Europe. We also consider a 19-bus unbalanced radial distribution system. The time-series voltage increments data are simulated by the MATLAB Power System Simulation Package, and empirically drawn to verify the Gaussianality. To obtain realistic voltage data, we use the residential power profile from Duquesne Light Company in Pittsburgh, PA, USA. This profile contains anonymized and secure hourly (and 15-minute) smart meter readings for more than 5,000 houses in the year 2016. In the subsequent experiments, we compare our private outage detection (POD) framework in Algorithm 1 to the benchmark in [7] with optimal detection performance but no privacy. Each experiment is conducted by the Monte Carlo simulation with over 1,000 replications. In every replication, we randomly simulate outage time λ through geometric prior with parameter $\rho = 0.04$, as we assume that outage can happen in any time step independently with an equal probability ρ.

A. Visualization of Privacy Guarantee

As proved in Theorem 2, our scheme (5) satisfies a $G_{\frac{s}{\sigma_e}}$-Gaussian differential privacy. To visualize this privacy guarantee, we plot the trade-off functions $T(\mathcal{N}(0,1), \mathcal{N}(\frac{s}{\sigma_e}, 1))$ at different levels of noise variance. As shown in the left half of Fig. 3, the privacy scheme with $\sigma_e^2 = 0.05$ makes distinguishing the noisy data from real data harder than distinguishing distributions $\mathcal{N}(0,1)$ from $\mathcal{N}(3,1)$. Increasing the amount of noise, i.e., $\sigma_e^2 = 0.2$, we arrive at better protection of privacy: distinguishing the noisy data from real user data becomes harder than distinguishing $\mathcal{N}(0,1)$ and $\mathcal{N}(1,1)$.

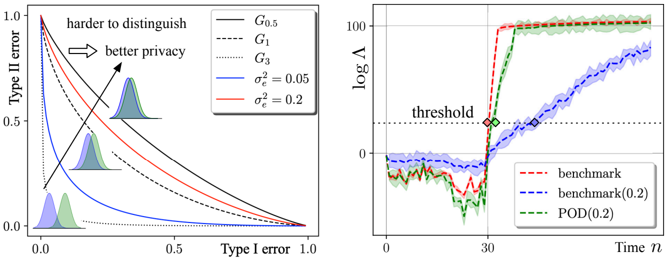

Fig. 3: **Left:** The trade-off functions of unit-variance Gaussian distributions and our randomizing schemes applied to IEEE 8-bus outage dataset. **Right:** The logarithm posterior probability ratio under different methods in IEEE 8-bus system with $\lambda = 30$.

B. Evaluation of the Noise-Mitigation Design

To evaluate the noise-mitigation effect of our proposed statistic in (8), we plot it with the benchmark statistic $\Lambda(\Delta\mathbf{v}^{1:N})$ in the right half of Fig. 3. The benchmark statistic under noise-free case increases dramatically after the outage time $\lambda = 30$, resulting in a near-zero detection delay. The benchmark statistic with noisy data $\Lambda(\Delta\tilde{\mathbf{v}}^{1:N})$ ($\sigma_e^2 = 0.2$) postpones the increment of posterior probability ratio, which leads to a larger detection delay, matching the conclusion in Theorem 3. On the contrary, once we implement the new calculation of posterior probability ratio as in (8), the postponed effect is mitigated as the new statistic approximates the noise-free ratio very well.

C. Average Detection Delay and Empirical False Alarm Rate

We evaluate the detection performance of using (8) to handle randomized data of $\sigma_e^2 = 0.2$. To validate the asymptotic optimality of the detection delay in Theorem 1, we plot in Fig. 4 the average delay $\mathbb{E}(\tau - \lambda | \tau \geq \lambda)$ divided by $|\log\alpha|$ and the theoretical lower bound $-\log(1-\rho) + D_{\mathrm{KL}}(f||g)$. The detection delay of the noise-free benchmark and that of the POD both achieve the optimal lower bound asymptotically, while the delay of benchmark statistic under noisy data is higher, which matches our quantification of performance degradation.

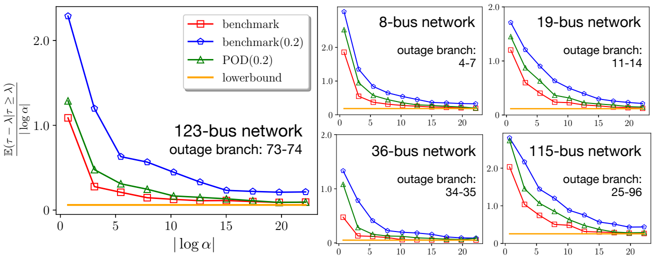

Fig. 4: The average detection delay in various systems

The detection rule is also expected to restrict the false alarm rate below α. To verify this, we calculate the empirical false alarm rate $\mathbb{P}(\tau < \lambda)$ and compare it against the upper bound α, as shown in Fig. 5. Our proposed method has a similar performance compared to the benchmark since the empirical false alarm is mainly below the upper bound α (especially when $\alpha \to 0$). These observations indicate that our proposed algorithm could quickly detect line outages with a low false alarm rate in both balanced and unbalanced distribution systems, even when randomized data are utilized.

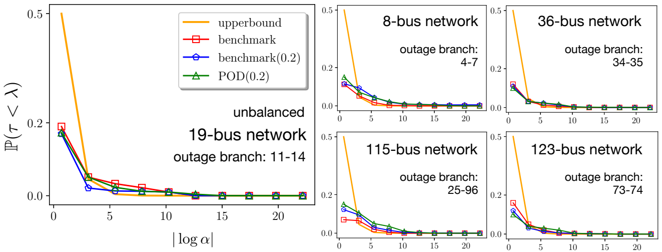

Fig. 5: The false alarm rate in various systems.

D. Sensitivity Analysis to Meter Coverage

Not every household in the distribution grid has installed smart meters. Thus, we test our method under different levels of meter coverage by selecting a ratio of buses to provide its voltage data for outage detection. We also "encrypt" the voltage data using the scheme (5) with $\sigma_e^2 = 0.2$ to preserve user privacy. Fig. 6 demonstrates both the average detection delay and the false alarm rate of our method under different levels of coverage ratio. Compared to the scenario where all buses have smart meters, 2 more samples are needed to detect the outage given noisy data when the smart meter coverage ratio is 75%, and 6 more samples are needed when the coverage ratio is only 50%. The false alarm rate increases from 0.9% to 3.1% when only 75% of buses have smart meters.

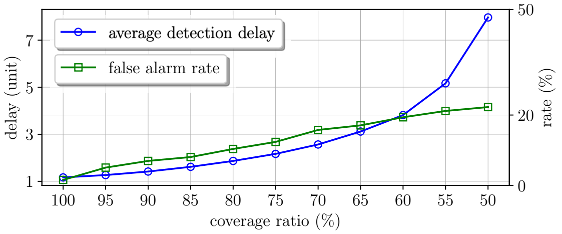

Fig. 6: Average detection delay (unit) and false alarm rate (%) under different meter coverages in IEEE 123-bus system, $\alpha = 0.01$.

V. Conclusion

This paper proposes a novel approach to detect line outages in distribution grids with privacy guarantees. To preserve privacy, a decentralized randomizing scheme is developed to protect users' raw data from exposure. By quantifying the privacy gain and detection performance degradation, we know how much performance is sacrificed given a desired level of privacy. To mitigate the performance degradation, we design a new statistic that serves as an unbiased estimation of the optimal statistic. Then, our detection procedure becomes both privacy-preserving and has comparable performance to the optimal case. The numerical results on both balanced and unbalanced distribution systems show that our method is suitable for real-world outage detection with privacy guarantees.

References

[1] J. Yuan and Y. Weng, "Physically invertible system identification for monitoring system edges with unobservability," *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)*, 2022.

[2] F. Location, "Isolation and service restoration technologies reduce outage impact and duration," *Smart Grid Investment Grant Program*, 2014.

[3] M. He and J. Zhang, "Fault detection and localization in smart grid: A probabilistic dependence graph approach," in *2010 First IEEE International Conference on Smart Grid Communications*.

[4] M. Babakmehr, F. Harirchi, A. Al-Durra, S. Muyeen, and M. G. Simões, "Compressive system identification for multiple line outage detection in smart grids," *IEEE Transactions on Industry Applications*, 2019.

[5] R. A. Sevlian, Y. Zhao, R. Rajagopal, A. Goldsmith, and H. V. Poor, "Outage detection using load and line flow measurements in power distribution systems," *IEEE Transactions on Power Systems*, 2017.

[6] M. Soleymani and A. Safdarian, "Unsupervised learning for distribution grid line outage and electricity theft identification," in *2019 Smart Grid Conference (SGC)*. IEEE, 2019, pp. 1–5.

[7] Y. Liao, Y. Weng, C.-W. Tan, and R. Rajagopal, "Quick line outage identification in urban distribution grids via smart meters," *CSEE Journal of Power and Energy Systems*, 2021.

[8] Y. Liao, C. Xiao, and Y. Weng, "Quickest line outage detection with low false alarm rate and no prior outage knowledge," in *2022 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, 2022, pp. 1–5.

[9] A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar, "Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey," *Sensors*, vol. 12, no. 12, pp. 16 838–16 866, 2012.

[10] G. Wood and M. Newborough, "Dynamic energy-consumption indicators for domestic appliances: environment, behaviour and design," *Energy and buildings*, vol. 35, no. 8, pp. 821–841, 2003.

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

[12] G. Kalogridis, R. Cepeda, S. Z. Denic, T. Lewis, and C. Efthymiou, "Elecprivacy: Evaluating the privacy protection of electricity management algorithms," *IEEE Transactions on Smart Grid*, 2011.

[13] A. N. Shiryaev, "On optimum methods in quickest detection problems," *Theory of Probability & Its Applications*, vol. 8, no. 1, pp. 22–46, 1963.

[14] C. Wei, A. Wiesel, and R. S. Blum, "Change detection in smart grids using errors in variables models," in *2012 IEEE 7th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, 2012, pp. 17–20.

[15] Y. C. Chen, T. Banerjee, A. D. Domínguez-García, and V. V. Veeravalli, "Quickest line outage detection and identification," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 749–758, 2016.

[16] K. Gajula, V. Le, X. Yao, S. Zou, and L. Herrera, "Quickest detection of series arc faults on dc microgrids," in *2021 IEEE Energy Conversion Congress and Exposition (ECCE)*. IEEE, 2021, pp. 796–801.

[17] M. R. Asghar, G. Dán, D. Miorandi, and I. Chlamtac, "Smart meter data privacy: A survey," *IEEE Communications Surveys & Tutorials*, 2017.

[18] H. Maaß, H. K. Cakmak, F. Bach, R. Mikut, A. Harrabi, W. Süß, W. Jakob, K.-U. Stucky, U. G. Kühnapfel, and V. Hagenmeyer, "Data processing of high-rate low-voltage distribution grid recordings for smart grid monitoring and analysis," *EURASIP Journal on Advances in Signal Processing*, vol. 2015, no. 1, pp. 1–21, 2015.

[19] R. Cummings, S. Krehbiel, Y. Mei, R. Tuo, and W. Zhang, "Differentially private change-point detection," *Advances in neural information processing systems*, vol. 31, 2018.

[20] R. Cummings, S. Krehbiel, Y. Lut, and W. Zhang, "Privately detecting changes in unknown distributions," in *International Conference on Machine Learning*. PMLR, 2020, pp. 2227–2237.

[21] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of cryptography conference*. Springer, 2006, pp. 265–284.

[22] Y. Liao, Y. Weng, C.-W. Tan, and R. Rajagopal, "Urban distribution grid line outage identification," in *International Conference on Probabilistic Methods Applied to Power Systems*. IEEE, 2016, pp. 1–8.

[23] A. G. Tartakovsky and V. V. Veeravalli, "General asymptotic bayesian theory of quickest change detection," *Theory of Probability &amp; Its Applications*, vol. 49, no. 3, pp. 458–497, 2005.

[24] J. Dong, A. Roth, and W. J. Su, "Gaussian differential privacy," *arXiv preprint arXiv:1905.02383*, 2019.