Poster: Privacy Preserving Smart Meter Streaming Against Inference Attacks

Yuan Hong
University at Albany, SUNY
Albany, New York 12222
Email: hong@albany.edu

Wen Ming Liu
Concordia University
Montreal, Canada
Email: l_wenmin@ciise.concordia.ca

Lingyu Wang
Concordia University
Montreal, Canada
Email: wang@ciise.concordia.ca

***Abstract*—Smart meter reading streams would pose severe privacy threats to the consumers. In this paper, we first quantitatively measure the risks that adversaries infer about specific appliances' status from a sequence of smart meter readings, and define a novel privacy notion to bound such inference threats in the reading streams. In addition, we propose a series of novel lightweight privacy preserving streaming algorithms for smart meters to promptly output safe readings, which satisfy the privacy notion while ensuring superior utility, such as 0 billing error and 0 aggregation error.**

1. Introduction

Smart meters frequently transmit fine-grained meter readings to the electric utility, e.g., as frequent as 15 minutes. Such reading streams greatly benefit the utilities (e.g., electricity transmission scheduling) and the energy consumers (e.g., optimizing electricity usage and cutting down the bill). However, recent studies show that such features may also lead to serious breaches of consumers' privacy [2], [4], [5]. To prevent adversaries from compromising energy consumers' personal privacy, three major categories of privacy-preserving techniques were proposed recently: injecting noise into the original or aggregated meter readings (e.g., [2]), encrypting the meter readings with cryptographic primitives (e.g., [4]) and only reporting the temporally or geographically aggregated readings, or attaching batteries for households to mask the meter readings (e.g., [5]).

However, most of the privacy preserving smart metering techniques (e.g., [2], [4], [5]) only consider that *any meter reading is sensitive* and protect such time series data in general – it is unclear that which reading is sensitive and vulnerable. To the best of our knowledge, the privacy in terms of the "status of specific appliances at different times" (which is the *root cause* of various privacy concerns in smart meter readings [2], [3], [4], [5]) has not been formally defined and quantitatively measured in literature. In this paper, we investigate the *inference threats* by linking the meter readings to the status of specific appliances, and define a novel privacy notion ("$[\epsilon, \delta]$-Uncertainty") to bound such inference threats in any reading stream.

Specifically, we study two problems: (1) what exactly can adversaries infer about the status of appliances from the reading stream? and (2) how to efficiently modify the way smart meters stream the output readings such that the defined privacy notion is ensured with good output utility? To this end, we propose lightweight privacy preserving algorithms to stream the output readings *without any aggregation* while guaranteeing rigorous privacy and excellent utility. Therefore, the outputs can support most smart metering services, e.g., billing, regional statistics, load monitoring.

2. Problem Formulation

We denote a smart meter's associated *appliance set* as $A = \{a_1, \ldots, a_m\}$ with m appliances, and use $|a_1|$, $\ldots$, $|a_m|$ to represent their labeled consumption rates (in watts). In addition, we define *reading frequency* as ϕ to represent the time interval between two adjacent readings (e.g., 15 minutes). Any reading can be converted into a consumption rate, and vice-versa.

Given A, $|a_1|, \ldots, |a_m|$, we first define a function $h(\cdot)$ to calculate the overall consumption rate of any subset of A (which is a combination of appliances, denoted by E). Then, the consumption rate of E is referred as $h(E)$.

On the contrary, for any consumption rate ω, we can find all the combinations of appliances whose overall consumption rate equals ω – the "Candidate Appliance Set": $c(\omega) = \{E : E \subseteq 2^A, h(E) = \omega\}$.

Inference Attack. We then measure the probability that *an appliance can be inferred as "ON" from a given overall consumption rate* ω. Denoting the size of $c(\omega)$ as $|c(\omega)|$, we can represent ω's candidate appliance set as $c(\omega) = \{c(\omega)_1, c(\omega)_2, \ldots, c(\omega)_{|c(\omega)|}\}$. Then, there are $|c(\omega)|$ combinations of appliances corresponding to the consumption rate ω, and thus the probability of each combination equals $\frac{1}{|c(\omega)|}$ (based on the prior knowledge of appliances). To sum up, given a consumption rate ω, we can calculate the *inference probability* of any appliance $\forall i \in [1, m], a_i \in A$ w.r.t. ω as: $Pr[\omega \rightarrow a_i] = \sum_{j=1}^{|c(\omega)|} \frac{I_{ij}}{|c(\omega)|}$ where $\forall j \in [1, |c(\omega)|], I_{ij} \in \{0, 1\}$ and if $a_i \in c(\omega)_j$ then $I_{ij} = 1$; otherwise $I_{ij} = 0$.

In real world, the adversaries can easily obtain the background knowledge of a wide variety of appliances (e.g., a common set of appliances in each household A) and their consumption rates (e.g., available in [1]) as well as the reading frequency ϕ. Given a reading stream $\vec{R_{in}}$, $\forall r \in \vec{R_{in}}$

(consumption rate $\omega = \frac{r}{\phi}$), the adversaries can derive ω's candidate appliance set $c(\omega)$ and thus derive the inference probabilities of all the appliances in $c(\omega)$.

Privacy Notion. We first define a privacy notion to quantify and bound such inference risk in any single reading:

Definition 1 ($[\epsilon, \delta]$-Uncertainty). A meter reading r and the corresponding consumption rate $\omega = \frac{r}{\phi}$ satisfy $[\epsilon, \delta]$-Uncertainty if $\forall a_i \in c(\omega), Pr[\omega \rightarrow a_i] \in [\epsilon, \delta]$, where $0 \leq \epsilon \leq \delta \leq 1$.

Thus, if any given reading r satisfies $[\epsilon, \delta]$-Uncertainty (or say r is $[\epsilon, \delta]$-Uncertain), the inference probabilities of all the possible appliances in ω's candidate appliance set $c(\omega)$ are bounded in the predefined range $[\epsilon, \delta]$. Furthermore, we extend a privacy notion for a streaming algorithm:

Definition 2 ($[\epsilon, \delta] \otimes \alpha$-Uncertainty). A streaming algorithm achieves $[\epsilon, \delta] \otimes \alpha$-Uncertainty, if for any input reading stream $\vec{R_{in}}$, the algorithm ensures that the output reading stream $\vec{R_{out}}$ satisfies

$$\frac{\left| r : r \in \vec{R_{out}} \wedge r \text{ satisfies } [\epsilon, \delta]\text{-Uncertainty} \right|}{\left| \vec{R_{out}} \right|} \geq \alpha$$

where $\left| \vec{R_{out}} \right|$ represents the number of readings in the output (which equals $|\vec{R_{in}}|$) and $0 \leq \alpha \leq 1$.

3. Privacy Preserving Streaming Algorithms

For preventing inference attacks, we propose a two-phase technique to ensure $[\epsilon, \delta]$-Uncertainty for readings in smart metering streams, assuming that adversaries may have background knowledge on appliances:

Offline Phase: the smart meter is initialized with the appliance set A and privacy parameters $[\epsilon, \delta]$ to generate the safe candidate rate set G' (each consumption rate in G' and the corresponding readings satisfy $[\epsilon, \delta]$-Uncertainty). This phase is a one-time offline process (for small or medium appliance set, the exact G' can be derived; for large appliance set, the exponential number of appliance combinations can be approximated using heuristics).

Online Phase: the smart meter continuously converts the original reading stream to safe readings (the closest safe reading derived from G'), and sends the safe readings immediately to the electric utility. We propose three different lightweight streaming algorithms (complexity $O(K)$): (1) Cyclic Reading Conversion (CRC) – the remainder of every single reading conversion is rolled over to the last reading of the stream, (2) Dynamic Reading Conversion (DRC) – the remainder of every single reading conversion is rolled over to the next reading, and (3) Tariff-Aware Reading Conversion (TARC) – following either CRC or DRC for streaming output readings while locally computing the billed amount using the original readings by the smart meter rather than the electric utility. Note that TARC is proposed for ensuring 0 billing error if dynamic energy pricing policies are adopted, e.g., time-of-use plan.

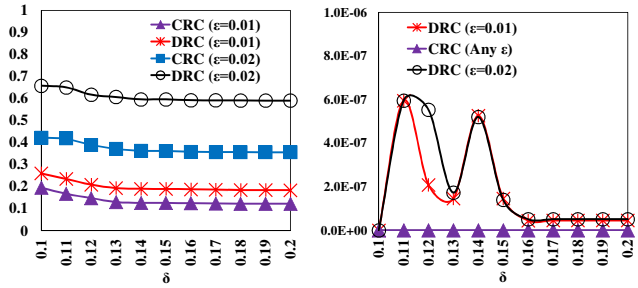
Lemma 1. CRC, DRC and TARC satisfy $[\epsilon, \delta] \otimes \frac{K-1}{K}$-Uncertainty, $[\epsilon, \delta] \otimes 1$-Uncertainty and $[\epsilon, \delta] \otimes \frac{K-2}{K}$-Uncertainty respectively where K is the number of readings in the stream (since $\frac{K-1}{K}$ and $\frac{K-1}{K} \approx 1$, all the output readings are $[\epsilon, \delta]$-Uncertain).

Lemma 2. Given the appliance set $A = \{a_1, \ldots, a_m\}$, the probability of identifying any sequential pattern $a_i \rightarrow a_j$ within k consecutive $[\epsilon, \delta]$-Uncertain readings is bounded in the range $[\epsilon(1-(1-\epsilon)^k), \delta]$.

Due to space limitation, we skip the details of three algorithms and the privacy analysis here. Table 1 summarizes three different utilities (billing error in a billing cycle, error of aggregated readings over a period, and sum of all the reading errors), where both constant tariff ("Standard") and time-of-use plan ("Dynamic") are considered. We conducted experiments on real datasets to validate such utility performance. For example, Figure 1 demonstrates the reading and billing error rates on varying ϵ and δ respectively.

TABLE 1. UTILITY OF THE ALGORITHMS

Algorithms	Billing Error		Aggregation Error	Reading Error
	Standard	Dynamic		
CRC	0	Medium/Low	0	Low
DRC	≈ 0	Medium/Low	≈ 0	Medium
TARC	0	0	0	Medium/Low



(a) Reading Error Rate vs. δ (b) Billing Error Rate vs. δ

Figure 1. CRC and CRC (Reading and Billing Error Rate)

References

[1] *http://energy.gov/*.

[2] G. Ács and C. Castelluccia. I have a dream! (differentially private smart metering). In *Information Hiding*, pages 118–132, 2011.

[3] Y. Hong, S. Goel, and W. M. Liu. An efficient and privacy-preserving scheme for p2p energy exchange among smart microgrids. *International Journal of Energy Research*, 40(3):313–331, 2016.

[4] C. Rottondi, G. Verticale, and C. Krauss. Distributed privacy-preserving aggregation of metering data in smart grids. *IEEE Journal on Selected Areas in Communications*, 31(7):1342–1354, 2013.

[5] W. Yang, N. Li, Y. Qi, W. H. Qardaji, S. E. McLaughlin, and P. McDaniel. Minimizing private data disclosures in the smart grid. In *ACM Conference on Computer and Communications Security*, pages 415–427, 2012.