Preserving Occupancy in Smart Grid

Maryam Kamgarpour and Christos Dimitrakakis

August 12, 2016

Smart meters obtain data from individual households for the purpose of grid monitoring, grid stability control and real-time billing of customers. There are several concerns over smart meter data infringing privacy of households. The aim of our work is to develop an approach that can ensure privacy of household occupancy, while ensuring accuracy of customers' bills.

Let the data that smart meters obtain from each household be the energy consumption in kilowatt-hour (KWhr) every Δ units of time, for example $\Delta = 15$ minutes. Denote this data by $x_t^i \in [0, \bar{x}]$, where i represents the household and t is the time index. Let occupancy of household i at time t be denoted by $z_t^i \in \{0, 1\}$. Let the utility function be denoted by $U^i(x^i, y^i) = |\sum_{t=1}^{T}(x_t^i - y_t^i)p_t|$, where p_t denotes the price of electricity at time t. This ensures that the customer bill remains close to its true value. Our problem is to maximise the utility, while ensuring occupancy remains private. We formalise this problem as follows. Design $M_t : \mathbb{R}_+ \to \mathbb{R}_+$ and communicate $y_t^i = M_t(x_t^i)$ through the smart meters to utility, with the objectives of minimising probability of finding $\{z_t^i\}_{\tau_1}^{\tau_2}$ from $\{y_t^i\}_{t_1}^{t_2}$ while maximising utility. MK: reasonable choice of t_1, t_2, τ_1, τ_2? As all the discussion below corresponds to data from an individual household, from now on, we will drop the index i.

Let occupancy evolve according to a two-state Markov chain, with time-dependent transition probability matrix $P_t^z(z_{t+1}|z_t)$. Let electricity consumption evolve according to a Markov process $P_t(x_{t+1}|x_t, z_t)$. For example, x_t can have a truncated normal distribution with mean determined by occupancy and time of day. MK: what choices of distribution are accurate representations and enable efficient computation?.

Motivated by the Pufferfish privacy framework [1], let the secret S denote occupancy of a household at a given time $\tau \in [\tau_1, \tau_2]$. Let the secret to hide be represented by $\mathcal{Q} = \{z_{\tau,0}, z_{\tau,1}\}$, where 0, 1 indices indicate no occupancy, occupancy at time τ, respectively. Let $Z = (Z_{\tau_1}, Z_{\tau_1+1}, \ldots, Z_{\tau_2})$. Let Θ denote all probability distributions consistent with the Markov transition kernel P_t defined above. Let $X = (x_{t_1}, x_{t_1+1}, \ldots, x_{t_2})$, $Y = (y_{t_1}, y_{t_1+1}, \ldots, y_{t_2})$. we call a mechanism $M : X \to Y$ ϵ-Pufferfish private if $\forall W \subset \text{Range}(M)$, and datasets $X \sim \theta$, with $\theta \in \Theta$

$$e^{-\epsilon} \leq \frac{P(M(X) \in W | Z_\tau = 0, \theta)}{P(M(X) \in W | Z_\tau = 1, \theta)} \leq e^{\epsilon}.$$

We need the above to hold for any $\tau \in [\tau_1, \tau_2]$. MK: Is this definition good?

In the Pufferfish paper, this class of problems are addressed by adding Laplacian noise to data, where the scale of the noise is dependent on the maximum Wasserstein distance W, between two densities in Θ. MK: Does this approach map to our setting? If so, how do we compute W? Are there other ways to ensure Pufferfish privacy? What is the loss in utility?

References

[1] Y. Wang, S. Song, and K. Chaudhuri. Privacy-preserving Analysis of Correlated Data. European Control Conference, 2016.