Privacy in the Smart Grid: Information, Control & Games*

Vince Poor Princeton University

Joint work with

V. Belmega, D. Gündüz, S. Kar, S. Mohajer, L. Sankar, et al.

Supported in part by NSF Grants CMMI-14-35778 and ECCS-1549881

* To appear in Information Theoretic Security and Privacy of Information Systems (CUP)

Outline

- I. Motivation
- 2. Information: A General Formalism
- 3. Control: Smart Meter Privacy
- 4. Games: Competitive Privacy

- The smart grid cyber layer generates considerable electronic data:
 - Power flow sensors, phasor measurement units, smart meters, etc.





- The smart grid cyber layer generates considerable electronic data:
 - Power flow sensors, phasor measurement units, smart meters, etc.





• This data can leak information that should be kept secure, or private.

- The smart grid cyber layer generates considerable electronic data:
 - Power flow sensors, phasor measurement units, smart meters, etc.





- This data can leak information that should be kept secure, or private.
- But, the utility of this data depend on its accessibility.

- The smart grid cyber layer generates considerable electronic data:
 - Power flow sensors, phasor measurement units, smart meters, etc.





- This data can leak information that should be kept secure, or private.
- But, the utility of this data depend on its accessibility.
- How can we characterize this fundamental tradeoff?

Information:

A General Formalism

[Sankar-Rajagopalan-Poor, T-IFS' | 3]

Data Source Model

• A sequence of *n* i.i.d. observations of a vector random variable $\mathbf{X} = (X_1 X_2 \dots X_K)$ with a joint distribution:

$$p_{\mathbf{X}}(\mathbf{x}) = p_{X_1 X_2 \dots X_K}(x_1, x_2, \dots, x_K)$$

Data Source Model

• A sequence of *n* i.i.d. observations of a vector random variable $\mathbf{X} = (X_1 X_2 \dots X_K)$ with a joint distribution:

$$p_{\mathbf{X}}(\mathbf{x}) = p_{X_1 X_2 \dots X_K}(x_1, x_2, \dots, x_K)$$

• Variables can be divided into public (revealed) and private (hidden) variables, typically not disjoint:

$$\left(\begin{array}{c} \mathbf{X}_{h,k} \\ \mathbf{X}_{r,k} \\ \vdots \\ \mathbf{revealed} \end{array} \right) \longrightarrow k^{th} \text{ entry} \\ \vdots \\ \mathbf{X}_{k} = \left(\begin{array}{c} \mathbf{X}_{r,k} \\ \mathbf{X}_{h,k} \\ \end{array} \right)$$

• How can we characterize the tradeoff between utility and privacy in such a setting?

- How can we characterize the tradeoff between utility and privacy in such a setting?
 - Measure utility by distortion of the public variables as revealed by the data source; and

- How can we characterize the tradeoff between utility and privacy in such a setting?
 - Measure utility by distortion of the public variables as revealed by the data source; and
 - Measure privacy by equivocation of the private variables in information revealed by the source. (Can also use other leakage measures.)

- How can we characterize the tradeoff between utility and privacy in such a setting?
 - Measure utility by distortion of the public variables as revealed by the data source; and
 - Measure privacy by equivocation of the private variables in information revealed by the source. (Can also use other leakage measures.)
- Then the distortion-equivocation region describes the tradeoff.

Encoder:
$$\mathbf{X}^{n} \rightarrow \mathcal{W} = \left\{ QDS_{1}, QDS_{2}, \dots, QDS_{M} \right\}$$

Encoder:
$$\mathbf{X}^{n} \rightarrow \mathcal{W} = \left\{ QDS_{1}, QDS_{2}, \dots, QDS_{M} \right\}$$



Encoder:
$$\mathbf{X}^{n} \rightarrow \mathcal{W} = \left\{ QDS_{1}, QDS_{2}, \dots, QDS_{M} \right\}$$



Encoder:
$$\mathbf{X}^{n} \rightarrow \mathcal{W} = \left\{ QDS_{1}, QDS_{2}, \dots, QDS_{M} \right\}$$



Encoder:
$$\mathbf{X}^{n} \rightarrow \mathcal{W} = \left\{ QDS_{1}, QDS_{2}, \dots, QDS_{M} \right\}$$



Utility-Privacy/RDE Regions



(a): Rate-Distortion-Equivocation Region

(b): Utility-Privacy Tradeoff Region

Example: Categorical Data

- Categorical data: finite alphabet data
 - e.g.: SSN, zipcode, etc.



Example: Categorical Data

- Optimal input to output mapping: reverse 'water-filling'
 - Only x with $p(x) > \lambda$ revealed (λ : water-level).



- Eliminates samples with low probabilities (relative to level λ)
 - Equivalent to outlier aggregation/suppression
 - Such samples reveal the most information
- As D^{\uparrow} , λ^{\uparrow} , revealing fewer samples

Summary (General Formalism)

- A data source is divided into private and public variables
 - Leads to an equivocation-distortion characterization
 - Adding rate: a rate-distortion problem with an equivocation constraint

Summary (General Formalism)

- A data source is divided into private and public variables
 - Leads to an equivocation-distortion characterization
 - Adding rate: a rate-distortion problem with an equivocation constraint
- We can also consider
 - multiple sources (side information)
 - other measures of privacy and/or utility

Control: Smart-Meter Privacy

[Sankar-Rajagopolan-Mohajer-Poor, T-SG'13] [Tan-Gündüz-Poor, JSAC: SG Series'13] [Yang-Chen-Zhang-Poor, T-SG'15]

Smart Meter Utility & Privacy

• Smart meter data is useful for price-aware usage, load balancing.



Smart Meter Utility & Privacy

- Smart meter data is useful for price-aware usage, load balancing.
- But, it leaks information about in-home activity.



A Source-Coding Approach

[Sankar-Rajagopolan-Mohajer-Poor, T-SG'13]

<u>Model</u>:

- hidden Gauss-Markov
- hidden state is in {continuous, intermittent}
- encoding of the meter readings

A Source-Coding Approach

[Sankar-Rajagopolan-Mohajer-Poor, T-SG'13]

<u>Model</u>:

- hidden Gauss-Markov
- hidden state is in {continuous, intermittent}
- encoding of the meter readings

distortion of usage

Tradeoff:

versus information leakage about the intermittent state

A Source-Coding Approach

[Sankar-Rajagopolan-Mohajer-Poor, T-SG'13]

<u>Model</u>:

- hidden Gauss-Markov
- hidden state is in {continuous, intermittent}
- encoding of the meter readings

distortion of usage

Tradeoff:

versus information leakage about the intermittent state

Solution:

a type of "reverse water-filling" (i.e., rate-minimizing source coding for Gaussian sources)

Reverse Water-Filling



A Control Approach

[Tan-Gündüz-Poor, JSAC: SG Series' 13]

• Consider situations with energy harvesting (e.g., solar or wind) and rechargeable storage devices (e.g., electric vehicle):



Energy Management Policies

Tradeoff: wasted energy rate:
$$P_W^n = \frac{1}{n} \sum_{i=1}^n (Z_i + Y_i - X_i)$$

information leakage rate: $I^n = \frac{1}{n} I(X^n; Y^n)$

Energy Management Policies

wasted energy rate:
$$P_W^n = \frac{1}{n} \sum_{i=1}^n (Z_i + Y_i - X_i)$$

versus
information leakage rate: $I^n = \frac{1}{n} I(X^n; Y^n)$

Policy: transition probabilities: $P(Y_i | X_i, Z_i, B_{i-1})$

Energy Management Policies

Tradeoff: wasted energy rate:
$$P_W^n = \frac{1}{n} \sum_{i=1}^n (Z_i + Y_i - X_i)$$

information leakage rate: $I^n = \frac{1}{n} I(X^n; Y^n)$

transition probabilities: $P(Y_i | X_i, Z_i, B_{i-1})$

Policy:



- battery introduces memory: closed form expressions are elusive
- numerically compute mutual information



With/ Without a Battery Vs. EH Rate


With No Energy Harvesting



Privacy vs. battery capacity

Tradeoff vs. battery capacity (allow wasted grid energy)

Summary (Smart Meter Privacy)

- Two approaches to smart meter privacy:
 - source coding at the meter (reverse water filling)
 - control with storage and local supply

Summary (Smart Meter Privacy)

- Two approaches to smart meter privacy:
 - source coding at the meter (reverse water filling)
 - control with storage and local supply
- We can also consider [Yang-Chen-Zhang-Poor, T-SG'15]:
 - adaptive control
 - jointly consider privacy and cost (exploit price variations)



Competitive Privacy

[Belmega-Sankar-Poor, JSTSP'15]

• N.A. Grid: interconnected regional transmission organizations (RTOs)



- N.A. Grid: interconnected regional transmission organizations (RTOs) which
 - need to share state measurements for reliability of state estimation (utility)



- N.A. Grid: interconnected regional transmission organizations (RTOs) which
 - need to share state measurements for reliability of state estimation (utility)
 - wish to withhold information for economic competitiveness (privacy)



- N.A. Grid: interconnected regional transmission organizations (RTOs) which
 - need to share state measurements for reliability of state estimation (utility)
 - wish to withhold information for economic competitiveness (privacy)



• Leads to a problem of competitive privacy

• Noisy measurements at RTO k:



• Noisy measurements at RTO k:



• Cooperation leads to inevitable leakage of state information.

• Noisy measurements at RTO k:



- Cooperation leads to inevitable leakage of state information.
- Utility for RTO k: mean-square error for its own state X_k

• Noisy measurements at RTO k:



- Cooperation leads to inevitable leakage of state information.
- Utility for RTO k: mean-square error for its own state X_k
- Privacy for RTO k: leakage of information about X_k to other RTOs

n i.i.d. observations at each RTO:

$$Y_{1,i} = X_{1,i} + \alpha X_{2,i} + Z_{1,i}, \ i = 1,...,n$$
$$Y_{2,i} = \beta X_{1,i} + X_{2,i} + Z_{2,i}, \ i = 1,...,n$$

Stochastic model:

$$X_{j,i} \sim N(0,1); Z_{2,i} \sim N(0,\sigma_j^2);$$
all indep.

n i.i.d. observations at each RTO:

$$Y_{1,i} = X_{1,i} + \alpha X_{2,i} + Z_{1,i}, \ i = 1,...,n$$
$$Y_{2,i} = \beta X_{1,i} + X_{2,i} + Z_{2,i}, \ i = 1,...,n$$

Stochastic model:

Tradeoff parameters:

$$X_{j,i} \sim N(0,1); Z_{2,i} \sim N(0,\sigma_j^2);$$
 all indep.

$$D_{j} = \frac{1}{n} \sum_{i=1}^{n} E\left[\left(X_{j,i} - \hat{X}_{j,i}\right)^{2}\right]$$
$$L_{j} = \frac{1}{n} I\left(X_{j}^{n}; J_{j}, Y_{3-j}^{n}\right)$$

n i.i.d. observations at each RTO:

$$Y_{1,i} = X_{1,i} + \alpha X_{2,i} + Z_{1,i}, \ i = 1,...,n$$
$$Y_{2,i} = \beta X_{1,i} + X_{2,i} + Z_{2,i}, \ i = 1,...,n$$

Stochastic model:

$$X_{j,i} \sim N(0,1); Z_{2,i} \sim N(0,\sigma_j^2);$$
 all indep.

$$D_{j} = \frac{1}{n} \sum_{i=1}^{n} E\left[\left(X_{j,i} - \hat{X}_{j,i}\right)^{2}\right]$$
$$L_{j} = \frac{1}{n} I\left(X_{j}^{n}; J_{j}, Y_{3-j}^{n}\right)$$

Tradeoff parameters:

<u>Theorem</u>: Wyner-Ziv coding maximizes privacy (i.e., minimizes L_1 and L_2) for a desired utility at each agent (fixed D_1 and D_2).

n i.i.d. observations at each RTO:

$$Y_{1,i} = X_{1,i} + \alpha X_{2,i} + Z_{1,i}, \ i = 1,...,n$$
$$Y_{2,i} = \beta X_{1,i} + X_{2,i} + Z_{2,i}, \ i = 1,...,n$$

Stochastic model:

$$X_{j,i} \sim N(0,1); Z_{2,i} \sim N(0,\sigma_j^2);$$
 all indep.

$$D_{j} = \frac{1}{n} \sum_{i=1}^{n} E\left[\left(\boldsymbol{X}_{j,i} - \hat{\boldsymbol{X}}_{j,i}\right)^{2}\right]$$
$$L_{j} = \frac{1}{n} I\left(\boldsymbol{X}_{j}^{n}; \boldsymbol{J}_{j}, \boldsymbol{Y}_{3-j}^{n}\right)$$

Tradeoff parameters:

<u>Theorem</u>: Wyner-Ziv coding maximizes privacy (i.e., minimizes L_1 and L_2) for a desired utility at each agent (fixed D_1 and D_2).

But, L_i depends on D_{3-i} (not D_i), so how should each agent choose to behave?

n i.i.d. observations at each RTO:

$$Y_{1,i} = X_{1,i} + \alpha X_{2,i} + Z_{1,i}, \ i = 1,...,n$$
$$Y_{2,i} = \beta X_{1,i} + X_{2,i} + Z_{2,i}, \ i = 1,...,n$$

$$X_{j,i} \sim N(0,1); Z_{2,i} \sim N(0,\sigma_j^2);$$
 all indep

$$D_{j} = \frac{1}{n} \sum_{i=1}^{n} E\left[\left(X_{j,i} - \hat{X}_{j,i}\right)^{2}\right]$$
$$L_{j} = \frac{1}{n} I\left(X_{j}^{n}; J_{j}, Y_{3-j}^{n}\right)$$

Tradeoff parameters:

Stochastic model:

<u>Theorem</u>: Wyner-Ziv coding maximizes privacy (i.e., minimizes L_1 and L_2) for a desired utility at each agent (fixed D_1 and D_2).

But, L_i depends on D_{3-i} (not D_i), so how should each agent choose to behave?

We can study this issue via game theory [Belmega-Sankar-Poor].

Rate and Privacy Leakage (Illustration) $\alpha = 1, \beta = 8, \sigma_1^2 = 0.05, \sigma_2^2 = 1$



The action a_j of agent j is the distortion caused at agent 3-j, maximally \overline{D}_{3-j} .

The action a_j of agent j is the distortion caused at agent 3-j, maximally \overline{D}_{3-j} .

The **payoff** for agent *j*:

$$u_{j}(a_{j}, a_{3-j}) = -w_{j}L(a_{j}) + w_{j}' \log\left(\frac{\overline{D}_{j}}{a_{3-j}}\right)$$

The action a_j of agent j is the distortion caused at agent 3-j, maximally \overline{D}_{3-j} .

The **payoff** for agent *j*:

$$u_{j}(a_{j}, a_{3-j}) = -w_{j}L(a_{j}) + w_{j}' \log\left(\frac{\overline{D}_{j}}{a_{3-j}}\right)$$

Nash equilibrium (prisoner's dilemma):

$$(a_1^*, a_2^*) = (\overline{D}_2, \overline{D}_1)$$

The action a_j of agent j is the distortion caused at agent 3-j, maximally \overline{D}_{3-j} .

 $u_{j}(a_{j}, a_{3-j}) = -w_{j}L(a_{j}) + w'_{j}\log\left(\frac{\bar{D}_{j}}{a_{3-j}}\right)$

Nash equilibrium (prisoner's dilemma):

$$(a_1^*, a_2^*) = (\overline{D}_2, \overline{D}_1)$$

Add pricing:

The payoff for agent *j*:

$$\tilde{u}_{j}(a_{j}, a_{3-j}) = u_{j}(a_{j}, a_{3-j}) + p_{j} \log\left(\frac{\overline{D}_{3-j}}{a_{j}}\right)$$

The action a_j of agent j is the distortion caused at agent 3-j, maximally \overline{D}_{3-j} .

$$u_{j}(a_{j}, a_{3-j}) = -w_{j}L(a_{j}) + w'_{j}\log\left(\frac{\overline{D}_{j}}{a_{3-j}}\right)$$

Nash equilibrium (prisoner's dilemma):

$$(a_1^*, a_2^*) = (\overline{D}_2, \overline{D}_1)$$

Add pricing:

The payoff for agent *j*:

$$\tilde{u}_{j}(a_{j}, a_{3-j}) = u_{j}(a_{j}, a_{3-j}) + p_{j}\log\left(\frac{\overline{D}_{3-j}}{a_{j}}\right)$$

Any behavior can then be incentivized within the limits of the model.

A Common-Goal Game

A common payoff:

$$u_{sys}(a_1, a_2) = -L(a_1) - L(a_2) + \frac{q}{2} \log\left(\frac{\overline{D}_1 + \overline{D}_2}{a_1 + a_2}\right)$$

A Common-Goal Game

A common payoff:

$$u_{sys}(a_1, a_2) = -L(a_1) - L(a_2) + \frac{q}{2} \log\left(\frac{\overline{D}_1 + \overline{D}_2}{a_1 + a_2}\right)$$

Enables cooperation in a non-cooperative setting (a potential game).

A Common-Goal Game

A common payoff:

$$u_{sys}(a_1, a_2) = -L(a_1) - L(a_2) + \frac{q}{2} \log\left(\frac{\overline{D}_1 + \overline{D}_2}{a_1 + a_2}\right)$$

Enables cooperation in a non-cooperative setting (a potential game).

Nontrivial equilibria exist; the nature of these depends on the value of q.

A Multi-Stage Game

T-stage game, with a discounted payoff for agent *j*:



A Multi-Stage Game

T-stage game, with a discounted payoff for agent *j*:



With $T < \infty$, the only Nash equilibrium (subgame perfect equilibrium):

 $(a_1^{(t)*}, a_2^{(t)*}) = (\overline{D}_2, \overline{D}_1), \forall t$

A Multi-Stage Game

T-stage game, with a discounted payoff for agent *j*:



With $T < \infty$, the only Nash equilibrium (subgame perfect equilibrium):

$$(a_1^{(t)^*}, a_2^{(t)^*}) = (\overline{D}_2, \overline{D}_1), \forall t$$

But, with $T = \infty$, any (D_1^*, D_2^*) satisfying the condition below is also a subgame perfect equilibrium for large enough $\rho < 1$:

$$u_{j}(D_{j}^{*}, D_{3-j}^{*}) > u_{j}(\overline{D}_{j}, \overline{D}_{3-j}); j = 1, 2$$

Minimal Discount Factor for Sustaining Non-trivial Equilibria

$$\alpha = 0.9, \beta = 0.5, \sigma_1^2 = \sigma_2^2 = 0.1, w_j' = 5w_j$$



Summary (Competitive Privacy)

• An additional dimension to privacy vs. utility tradeoff is added when there are multiple competing agents.

Summary (Competitive Privacy)

- An additional dimension to privacy vs. utility tradeoff is added when there are multiple competing agents.
- Wyner-Ziv coding gives optimal information exchange.

Summary (Competitive Privacy)

- An additional dimension to privacy vs. utility tradeoff is added when there are multiple competing agents.
- Wyner-Ziv coding gives optimal information exchange.
- Game theory can help in modeling and understanding this problem:
 - one-shot games: prisoner's dilemma/pricing
 - multi-stage games: finite vs. infinite time window
 - common-goal games: enables cooperation

Summary

• Motivation: privacy-utility tradeoff

Summary

- Motivation: privacy-utility tradeoff
- General Formalism: information theoretic formulation

Summary

- Motivation: privacy-utility tradeoff
- General Formalism: information theoretic formulation
- Smart Meter Privacy: source coding & control approaches
Summary

- Motivation: privacy-utility tradeoff
- General Formalism: information theoretic formulation
- Smart Meter Privacy: source coding & control approaches
- Competitive Privacy: game theoretic approach

Summary

- Motivation: privacy-utility tradeoff
- General Formalism: information theoretic formulation
- Smart Meter Privacy: source coding & control approaches
- Competitive Privacy: game theoretic approach
- Information-, control- and game-theoretic ideas allow fundamental examination of privacy issues in smart grid.

Basic P-U Tradeoff: Other Potential Applications

Biometric Systems: tradeoff between security & privacy







Social Networks: tradeoff between sharing & privacy

E-Commerce: tradeoff between economic benefit & privacy



Competitive Privacy: Other Potential Applications

Other Networks of Interacting Agents, e.g.:

- resource localization in competitive environments
- joint sensing with untrustworthy allies



References

S. Sankar, S. R. Rajagopalan, HVP, "Utility-Privacy Tradeoffs in Databases: An Informationtheoretic Approach," IEEE Trans. Inform. Forensics & Security 8 (6) 838-852, 2013

S. Sankar, S. R. Rajagopalan, S. Mohajer, HVP, "Smart Meter Privacy: A Theoretical Framework," IEEE Trans. Smart Grid **4** (2) 837-846, 2013

O. Tan, D. Gündüz, HVP, "Increasing Smart Meter Privacy Through Energy Harvesting and Storage Devices," IEEE JSAC: Smart Grid Communications **31** (7) 1331-1341, 2013

L. Yang, X. Chen, J. Zhang, HVP, "Cost-Effective and Privacy-Preserving Energy Management for Smart Meters," IEEE Trans. Smart Grid 6 (1) 486-495, 2015

V. Belmega, L. Sankar, HVP, "Enabling Data Exchange in Interactive State Estimation under Privacy Constraints," IEEE J. Select. Topics in Signal Process. **9** (7) 1285-1297, 2015

HVP, "Privacy in the Smart Grid: Information, Control & Games," in Information Theoretic Security and Privacy of Information Systems. (Cambridge UP), to appear.

