Balancing Privacy and Access to Smart Meter Data

Dr Fei Teng, Mr Saurab Chhachhi
Imperial College London
Email: f.teng@imperial,ac.uk

Questions?

- Does the privacy concern prevent the consumers from sharing the data?
- What private information can be leaked by smart meter data? How is affected by other sources of data Do the consumers know? How does this affect consumer choice?
- What are the alternative approaches to protect privacy? How to choose?
- What are the costs of privacy protections (Infrastructure investment? impact on the utility of the data)?
- How to evaluate data under privacy protection?
- How to incentive data sharing by balancing privacy and utility?

Consumer Privacy Concerns

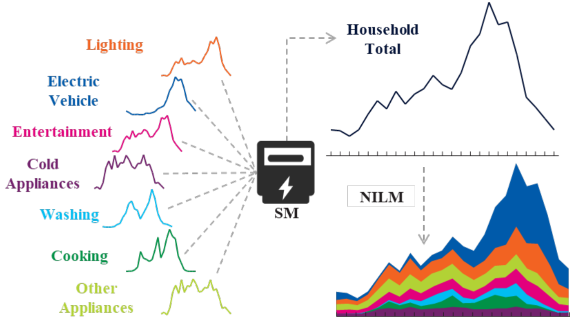
- Majority of consumers are willing to share their smart meter data, but it varies among customers
- Smart meter data is considered less sensitive than other types of personal data (financial, location, medical etc.)[1].
- Yet many consumers are not providing half-hourly data (49%) or simply do not know (37%) what their data sharing options[2]. WHY?
- **Are consumers currently making informed decisions?**
  - The options for data sharing
  - What are the implications of smart meter data sharing?

Willingness to Share: Customer-Facing Use Cases[3]

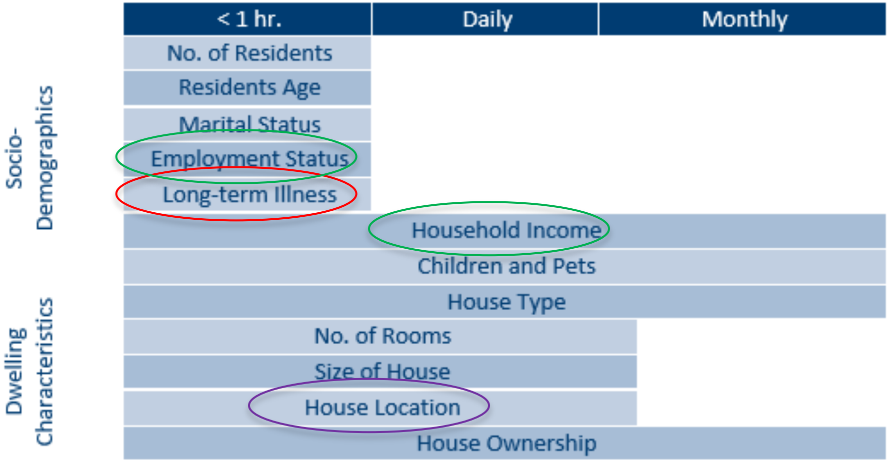
Willingness to Share: Market Operation Use Cases[3]

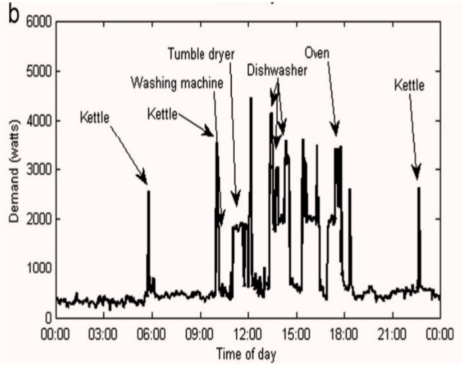
1. Skatova, A., McDonald, R. L., Ma, S., & Maple, C. (2019). Unpacking Privacy: Willingness to pay to protect personal data. https://doi.org/10.31234/osf.io/ahwe4
2. Citizen's Advice. (2019). *Clear and in control*.
3. Knight, A. (2018). *Consumer views on sharing half-hourly settlement data*. https://www.ofgem.gov.uk/publications-and-updates/consumer-research-datasets

Smart Meter Data – What information can be inferred?

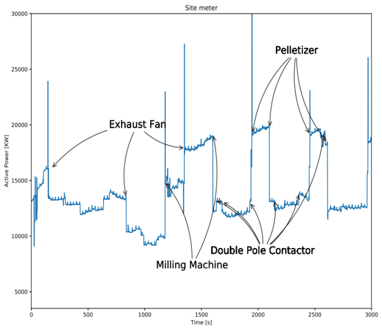


Medical
Financial
Location





Typical domestic load profile in the UK[2]



Typical load in a poultry feed factory[3]

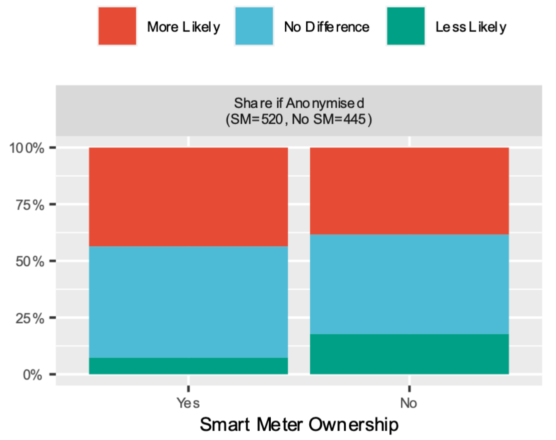
1. **Teng, F.**, Chhachhi, S., Ge, P., Prof, J. G., & Gunduz, D. (2022). Balancing privacy and access to smart meter data: an Energy Futures Lab briefing paper. *64*. https://doi.org/10.25561/96974
2. McKenna, E., Richardson, I., & Thomson, M. (2012). Smart meter data: Balancing consumer privacy concerns with legitimate applications. *Energy Policy*, *41*, 807–814. https://doi.org/10.1016/j.enpol.2011.11.049
3. Martins, P. B. et. al. (2018). Application of a Deep Learning Generative Model to Load Disaggregation for Industrial Machinery Power Consumption Monitoring. *2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, 1–6. https://doi.org/10.1109/SmartGridComm.2018.8587415

Can we provide a truly privacy-preserving mechanism to promote energy data sharing?

Attitudes to Sharing Smart Meter Data - Options

- Majority are not fully aware of the information that can be inferred from smart meter data.
- When provided with information on the implications of data sharing concerns increase.
- Consumers would be more inclined to share smart meter data if it were privacy-preserved.

Contract decisions once privacy implications known[1]



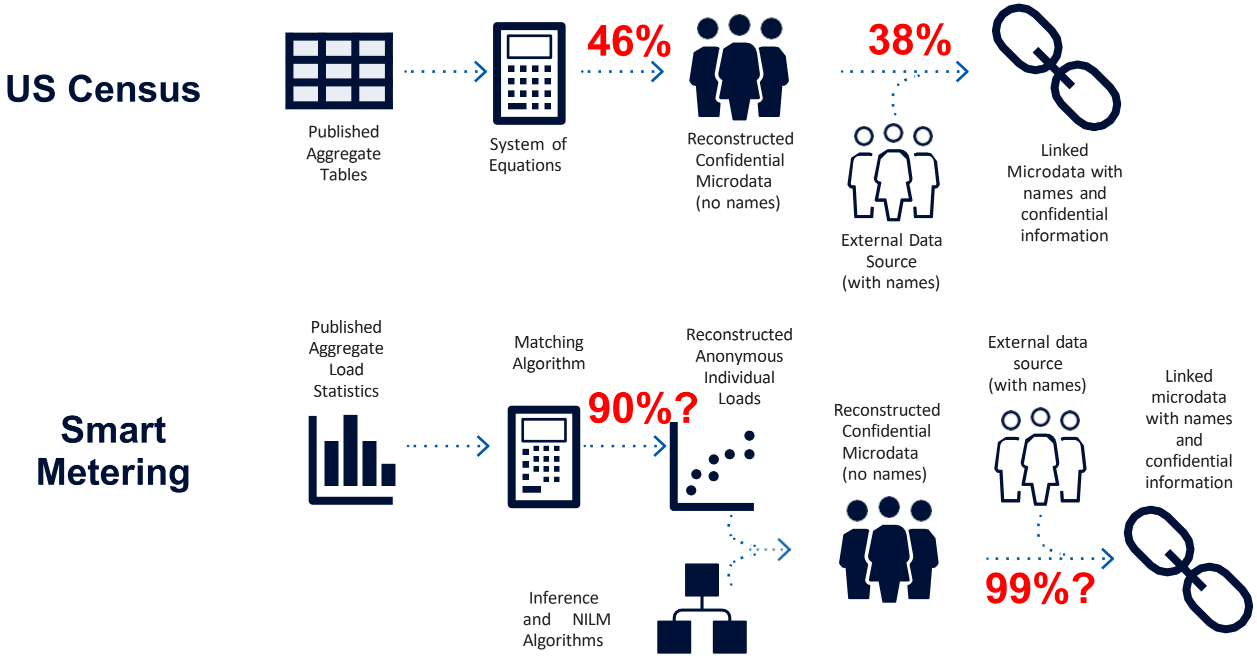
1. Jakobi, T., Patil, S., Randall, D., Stevens, G., & Wulf, V. (2019). It is about what they could do with the data: a user perspective on privacy in smart metering. ACM Transactions on Computer-Human Interaction, 26(1). https://doi.org/10.1145/3281444

Privacy-Preserving Technologies (PPTs)

- **Pseudonymisation**: Replacing identifiable features with consistent unique identifiers.
- **Aggregation:** Aggregating consumption data across multiple periods of time (temporal) or households(spatial).
- **Differential Privacy**: Carefully tuned noise addition to 'hide' individual contributions.
- **Homomorphic Encryption**: Perform arithmetic operations (e.g., addition, multiplication) on encrypted data without having to first decrypt it.
- **User Demand Shaping**: Altering actual consumption patterns using flexible assets to hide appliance characteristics.
- **Distributed Data Processing (Federated Learning)**: Distributed learning technique in which model parameters are shared but raw data kept locally.

Teng, F., Chhachhi, S., Ge, P., Prof, J. G., & Gunduz, D. (2022). Balancing privacy and access to smart meter data: an Energy Futures Lab briefing paper. *64*. https://doi.org/10.25561/96974

Vulnerabilities of Traditional Techniques



Teng, F., Chhachhi, S., Ge, P., Prof, J. G., & Gunduz, D. (2022). Balancing privacy and access to smart meter data: an Energy Futures Lab briefing paper. *64*. https://doi.org/10.25561/96974

Suitability

		Pseudo-nymisation	Aggregation	Homo-morphic Encryption	User Demand Shaping	Differential Privacy	Federated Learning
Privacy Guarantees	Anonymity	★	★	★		✓	
	Invulnerable to Linking					✓	
	Invulnerable to Inference				✓	✓	
	Minimise Impact of Data Breaches		★	✓	✓	✓	✓
Desirable Properties	Individual Level Data	✓			✓	★	★
	No Trusted Third-Party			✓	✓	✓	✓
	Easily Integrated	✓	✓			✓	
	Preserve Data Utility	✓	★	★	★		★
	Preference Heterogeneity	✓			✓	✓	✓

Orange stars indicate properties that the privacy-preserving technique is purported to have and which, in some cases, may have for practical purposes. However, these properties are not evidenced by theoretical guarantees.

Teng, F., Chhachhi, S., Ge, P., Prof, J. G., & Gunduz, D. (2022). Balancing privacy and access to smart meter data: an Energy Futures Lab briefing paper. *64*. https://doi.org/10.25561/96974

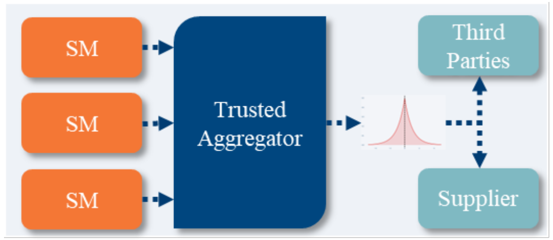
Is DP applicable for Energy Data? Energy Data is time series and is continuously released.

DP introduces a trade-off between privacy and accuracy. Implications for Energy Data?

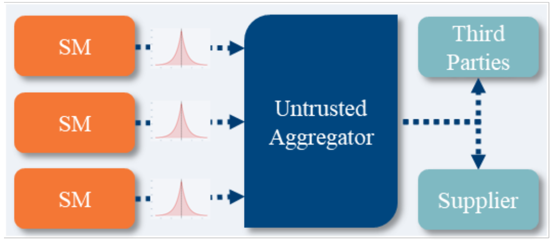
Centralised Framework – Differential Privacy

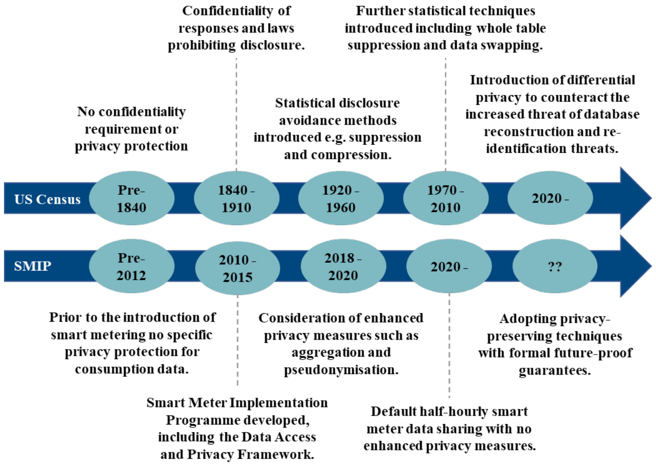
Differential Privacy (DP) – adding noise to mathematically ensure individuals cannot be identified from a dataset.

Global DP

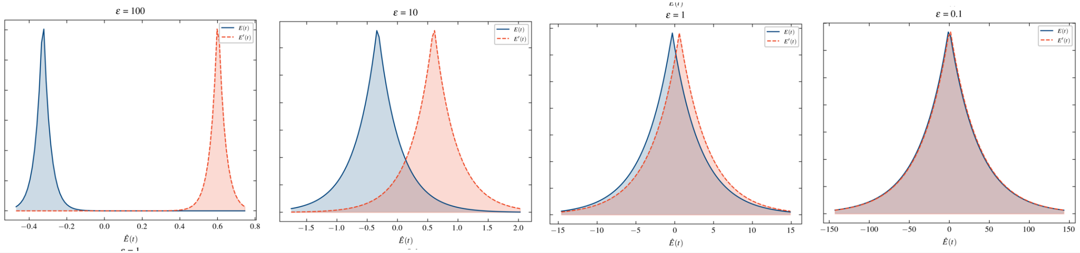


Local DP





$$\hat{E}(t)|E(t)$$
$$\hat{E}(t)|E'(t)$$



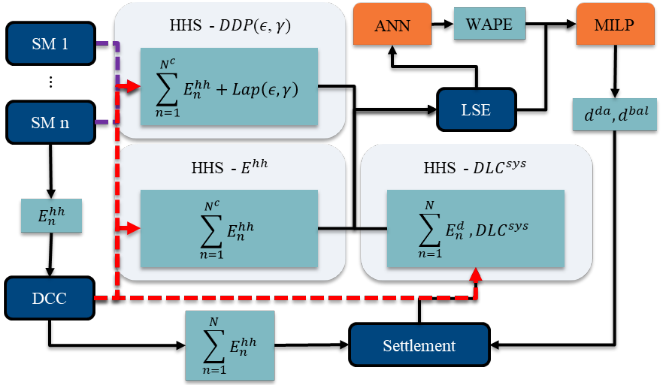
Teng, F., Chhachhi, S., Ge, P., Prof, J. G., & Gunduz, D. (2022). Balancing privacy and access to smart meter data: an Energy Futures Lab briefing paper. *64*. https://doi.org/10.25561/96974

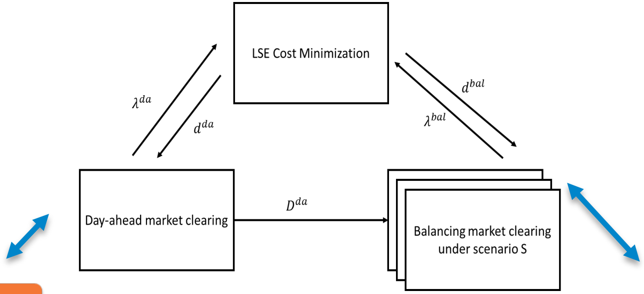
What is the right level of DP for each individual?

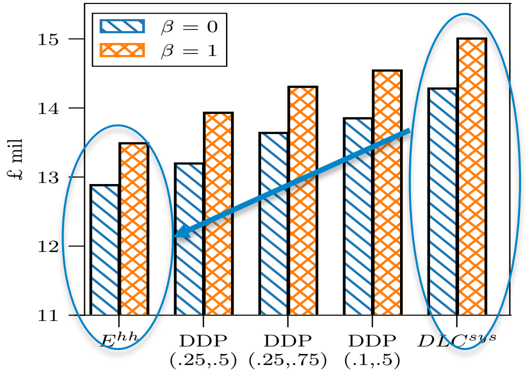
Imperial College London

Differentially Private Load Forecasting and Energy Procurement

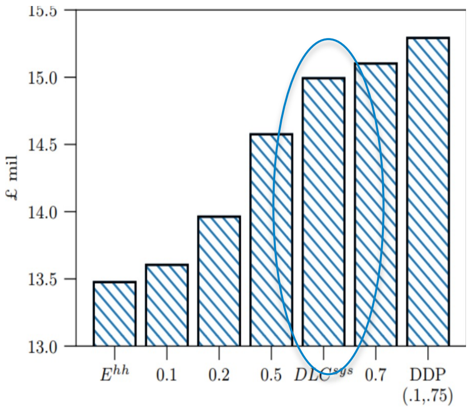
Project "Consumer-centric privacy protection scheme for energy consumption data", Supergen Energy Networks







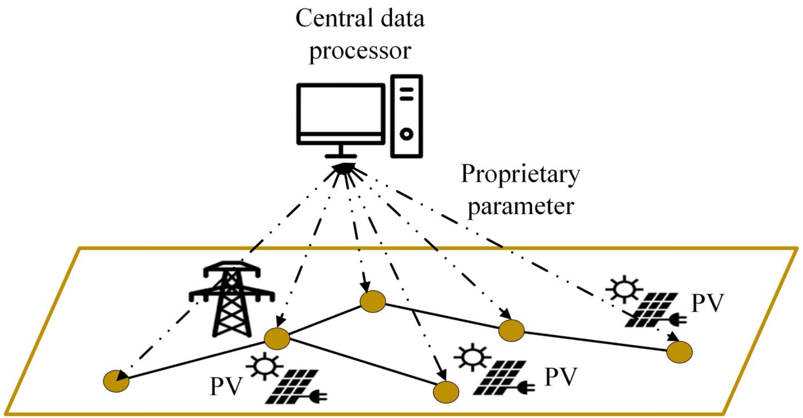
Leveraging Preference Heterogeneity?

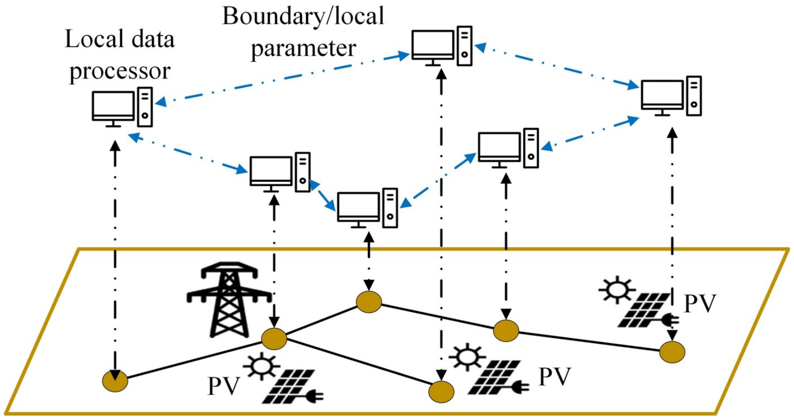


Chhachhi, S., & **Teng, F.** (2021). *Market Value of Differentially-Private Smart Meter Data*. 1–5. https://doi.org/10.1109/isgt49243.2021.9372228

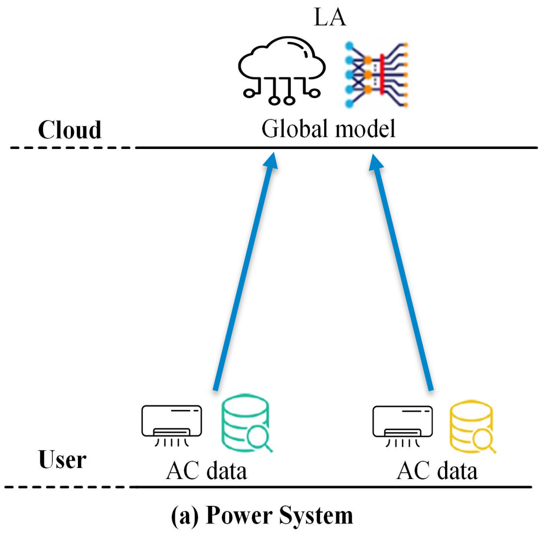
Imperial College London

Distributed Data Processing Framework for Energy Systems by Utilizing Edge-device





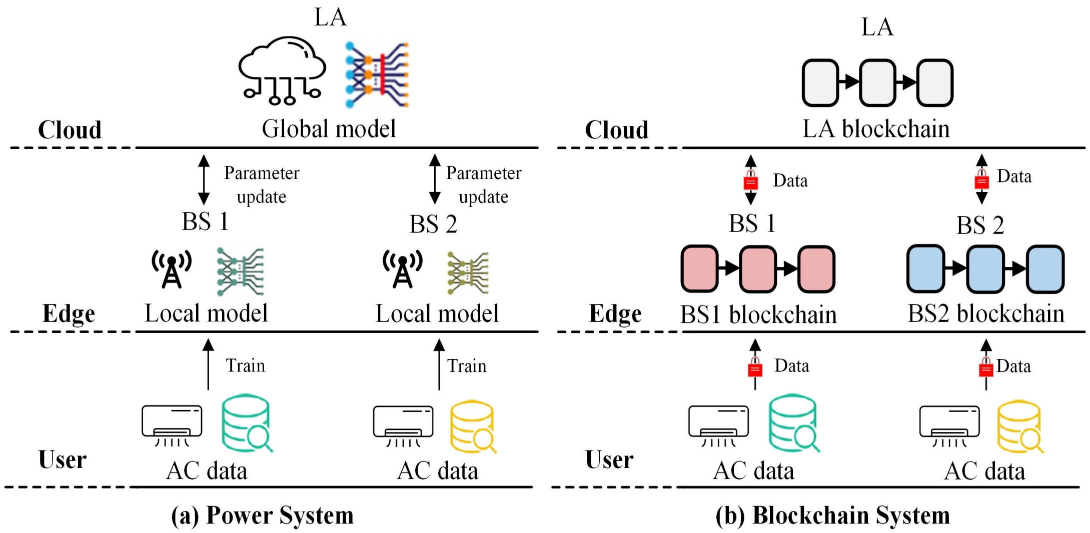
Federated Learning



Project "Blockchain-enabled cloud-edge coordination for demand side management", EPSRC-SIEMENS

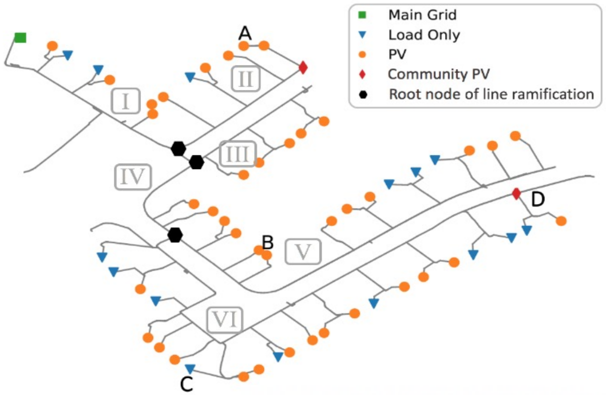
Wang, Y., Bennani, I. L., Liu, X., Sun, M., & Zhou, Y. (2021). Electricity Consumer Characteristics Identification: A Federated Learning Approach. *IEEE Transactions on Smart Grid*, *12*(4), 3637–3647. https://doi.org/10.1109/TSG.2021.3066577

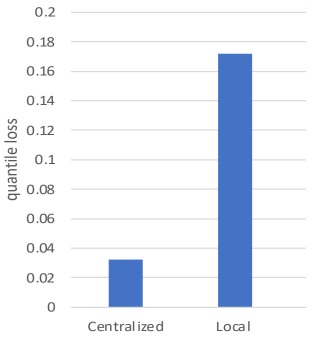
Federated Learning

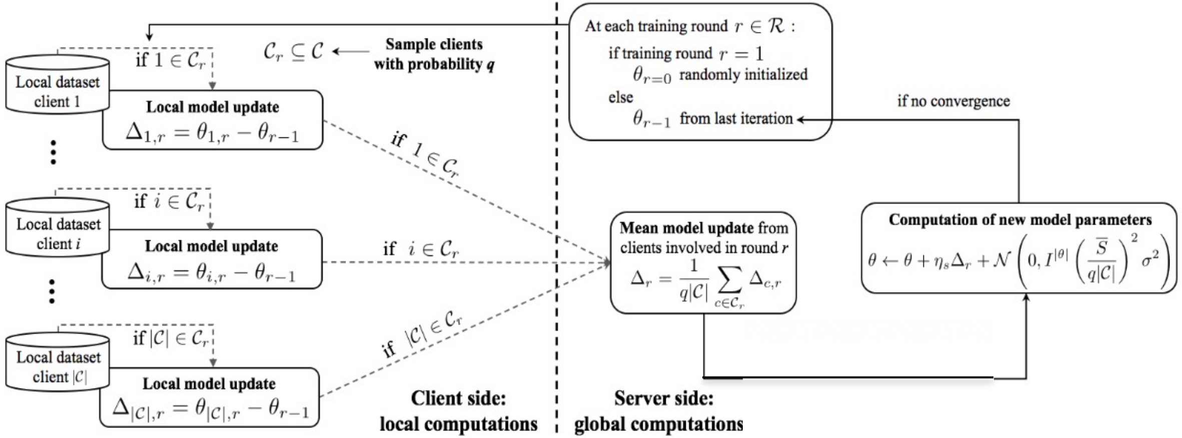


Project "Blockchain-enabled cloud-edge coordination for demand side management", EPSRC-SIEMENS

Federated Learning in Voltage Forecasting



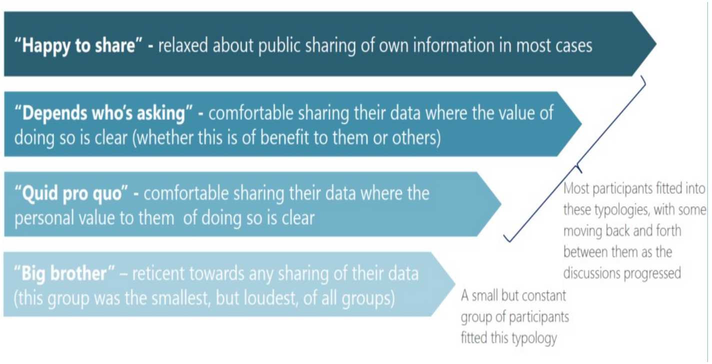


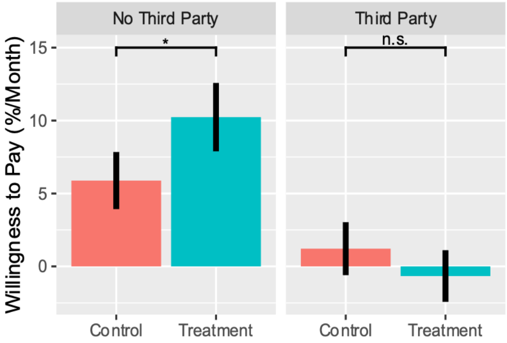


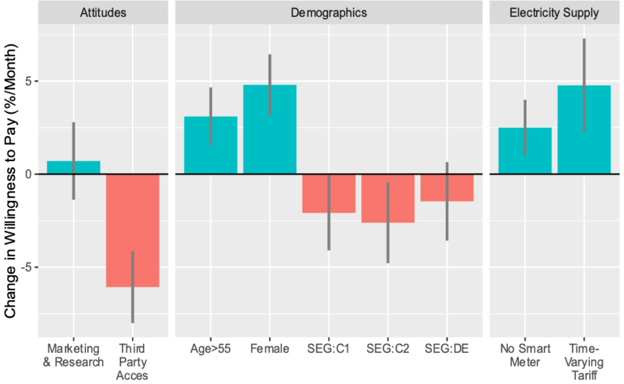
JF Toubeau, **F. Teng,** T. Morstyn, L. Krannichfeldt and Y. Wang "Privacy-Preserving Probabilistic Voltage Forecasting in Local Energy Communities", ArXiv

Attitudes to Sharing Smart Meter Data - Incentives

- A big portion of consumers were happy to share their data only if details on how it may benefit the system as well as benefit them personally is provided[1].
- Consumers are aware that their data has value and demand compensation for it (when given the choice).
- This increases when consumers are aware of the inferable information embedded within smart meter data.
- Significant heterogeneity across socio-demographic, contractual and attitudinal characteristics.







1. Dickman, A., & Aslaksen, A. P. (2017). *Consumer attitudes to DNO access to half hourly electricity consumption data*. https://www.ipsos.com/ipsos-mori/en-uk/data-privacy-and-smart-meters

Markets for Differentially-Private Energy Data

Data Valuation Mechanism

- Data value dependent on:
  - Task: model, evaluation metric
  - Context: other (public) data
  - Quality: noise and quantity
- Privacy concerns and ownership rights warrant valuation prior to data access.
- Data re-used repeatedly and for different tasks motivating a model agnostic ‘intrinsic’ valuation mechanism.

Data Market Mechanism

- Budget Feasibility: Data owners should be compensated commensurate with data value
- Individual rationality: Compensation should cover owners' own perceptions of value.
- Incentive compatibility: Privacy concerns and owners’ valuations should be truthful.
- Dependent Privacy: Payment-dependent privacy preferences.

Joint Energy & Data Market

- Data has direct effect on uncertainty in the energy markets (load, flexibility, prices).
- Inherent coupling and decision-dependent structure requires joint optimisation across energy and data markets.
- To ensure budget feasibility gains in energy market must cover improvements in, for example, procurement costs.

How do we value data in a model agnostic manner while preserving privacy i.e. without first accessing it?

✓ - shown,
✓ - possible with minimal extension

Data Valuation – Overview

1. **Model Error/Performance:** Reduction in, for example, mean square error for linear regression[1].
2. **Dual Variables/Shadow Prices:** Sensitivity of an optimization problem to a particular input/constraint (e.g. dual variables of Wasserstein DRO[2]).
3. **Regression Coefficients:** Regression performed with regularization determined by data owners' willingness-to-sell. (e.g. LASSO penalty parameters[3])
4. **Composite metrics + Transfer functions:** Metrics which incorporate quality and quantity metrics (e.g. Shannon entropy x non-noise ratio[4]) are fitted to model evaluation metrics using (synthetic) data.
5. **Statistical distances:** Measures which compute the similarity between two distributions (e.g. Kullback-Liebler Divergence or Wasserstein Distance[5]).

	Model Agnostic	No Data Access	Efficient Calculation	DP Effect
1				✓ (possible with minimal extension)
2			✓	✓
3			✓	
4		✓		✓ (possible with minimal extension)
5	✓	✓	✓	✓

1. Goncalves, C, et.al. (2021). Towards Data Markets in Renewable Energy Forecasting. *IEEE Transactions on Sustainable Energy*, *12*(1), 533–542. https://doi.org/10.1109/TSTE.2020.3009615
2. Mieth, R., et. al. (2023). *Data Valuation from Data-Driven Optimization*. http://arxiv.org/abs/2305.01775
3. Han, L, et. al.. (2021). *Trading Data for Wind Power Forecasting: A Regression Market with Lasso Regularization*. http://arxiv.org/abs/2110.07432
4. Chen, L., et. al. (2021). Toward Future Information Market: An Information Valuation Paradigm. *2021 IEEE Power & Energy Society General Meeting (PESGM)*, 1–5. https://doi.org/10.1109/PESGM46819.2021.9638205
5. Zhao, Y, et. Al.. (2018). *Federated Learning with Non-IID Data*. https://doi.org/10.48550/arXiv.1806.00582

Statistical Distances- Wasserstein Metric Valuation

- Which distance to use?
  - The Wasserstein distance between two distributions is:

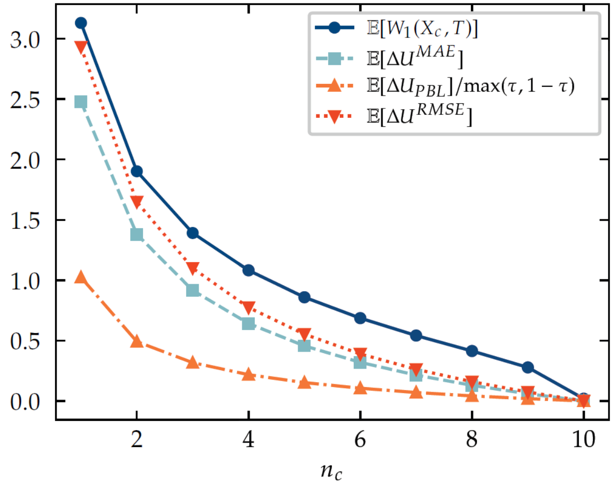
$$W_p(X,Y) = \inf_{X\sim\mu, Y\sim\nu} (E\|X-Y\|^p)^{1/p}$$

  - Wasserstein metric/distance has a number of advantages:
    - Continuous, well defined even when distributions do not overlap.
    - Has metric properties.
    - Calculated efficiently and privately in the case p = 1 [1,2].
    - Effect of differential privacy can be bounded[1]:

$$W_1(X_1 + X_L, X_2) \leq W_1(X_1, X_2) + \frac{\Delta}{\epsilon}, where\ X_L \sim Lap\left(0, \frac{\Delta}{\epsilon}\right)$$

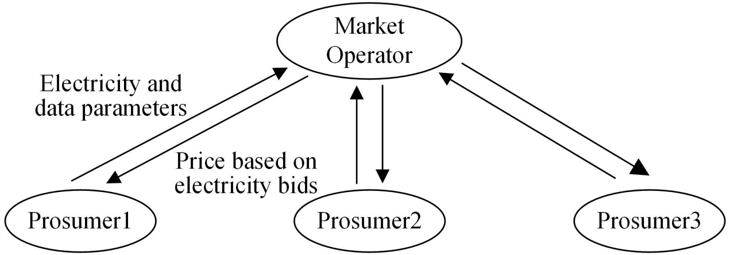
- How does it relate to model performance/error?
  - Lipschitz bound: *Given a K-Lipschitz loss function $l(x_i)$ and its expected value $U(X_i) = \mathbb{E}[l(x_i)]$, the difference between the loss obtained with X_1 or X_2 is bounded by the 1-Wasserstein distance between them*[3]*:*

$$|U(X) - U(Y)| \leq L \cdot W_1(X,Y)$$

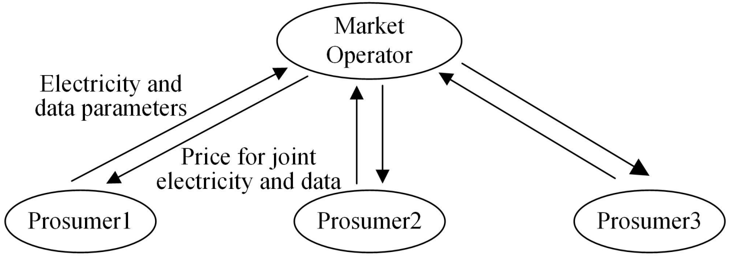


1. Chhachhi, S., & **Teng, F**. (2023). *On the 1-Wasserstein Distance between Location-Scale Distributions and the Effect of Differential Privacy*. http://arxiv.org/abs/2304.14869
2. Blanco-Justicia, A., & Domingo-Ferrer, J. (2020). Privacy-Preserving Computation of the Earth Mover's Distance. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 12472 LNCS*, 409–423.
3. Ghorbani, A., Kim, M. P., & Zou, J. (2020). *A Distributional Framework for Data Valuation*. http://arxiv.org/abs/2002.12334

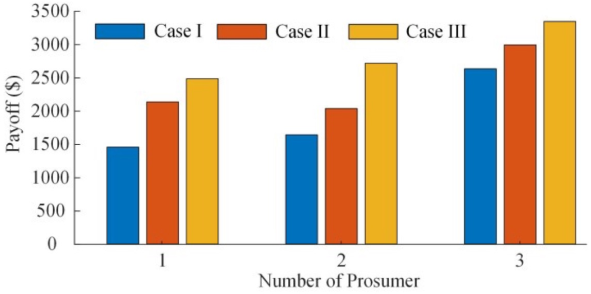
A Joint Energy and Data Market - Case study

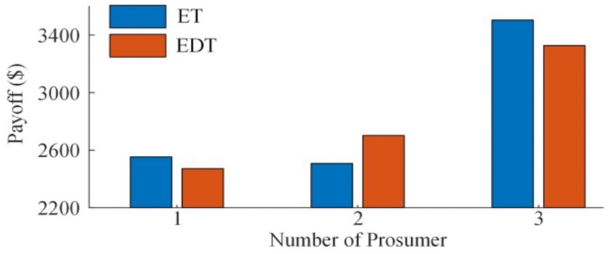


(a)



(b)





M. Yan and **F .Teng*** “Towards Joint Electricity and Data Trading: A Scalable Cooperative Game Theoretic Approach”, *Arxiv*

Summary

- Some of the data has embedded within significant amounts of personal or commercially sensitive information, particularly if combined with other data sources
- The majority of consumers are unaware but when informed have significant privacy concerns. Privacy concerns are heterogonous among consumers.
- Privacy-preserving techniques can ensure protection while providing access
  - DP is appliable for energy system data but needs to understand the trade-offs
  - A distributed framework for control, optimization and learning plays a critical role in energy digitalization
  - A hybrid centralized/decentralised data processing framework may be eventually needed
- A joint energy and data trading mechanism is needed for the future market