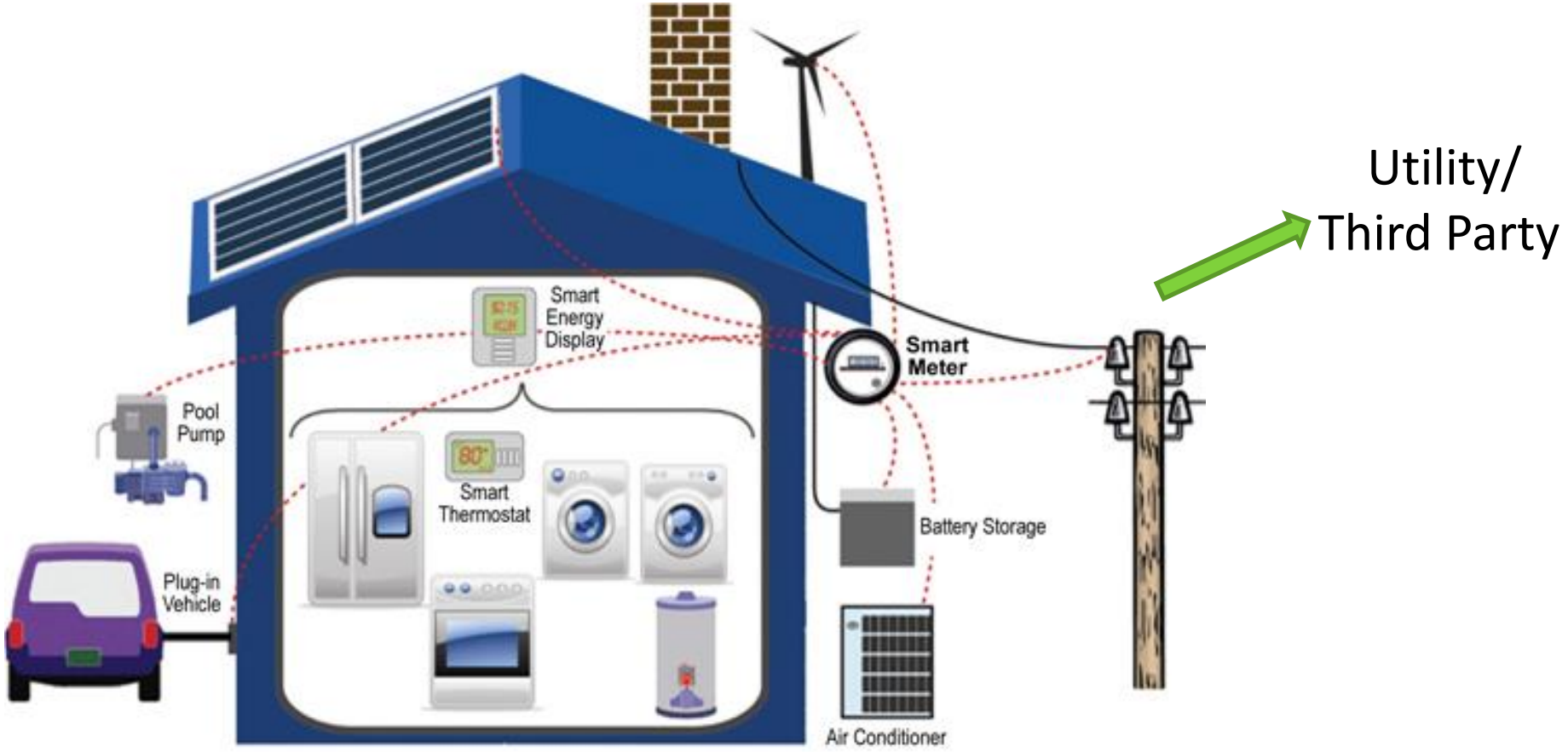
Privacy-preserving Data-aggregation for Internet-of-things in Smart Grid

Aakanksha Chowdhery
Postdoctoral Researcher, Microsoft Research
ac@microsoft.com

Collaborators: Victor Bahl, Ratul Mahajan, Frank Mcsherry, Abhradeep Thakurta

Smart meters/devices in home



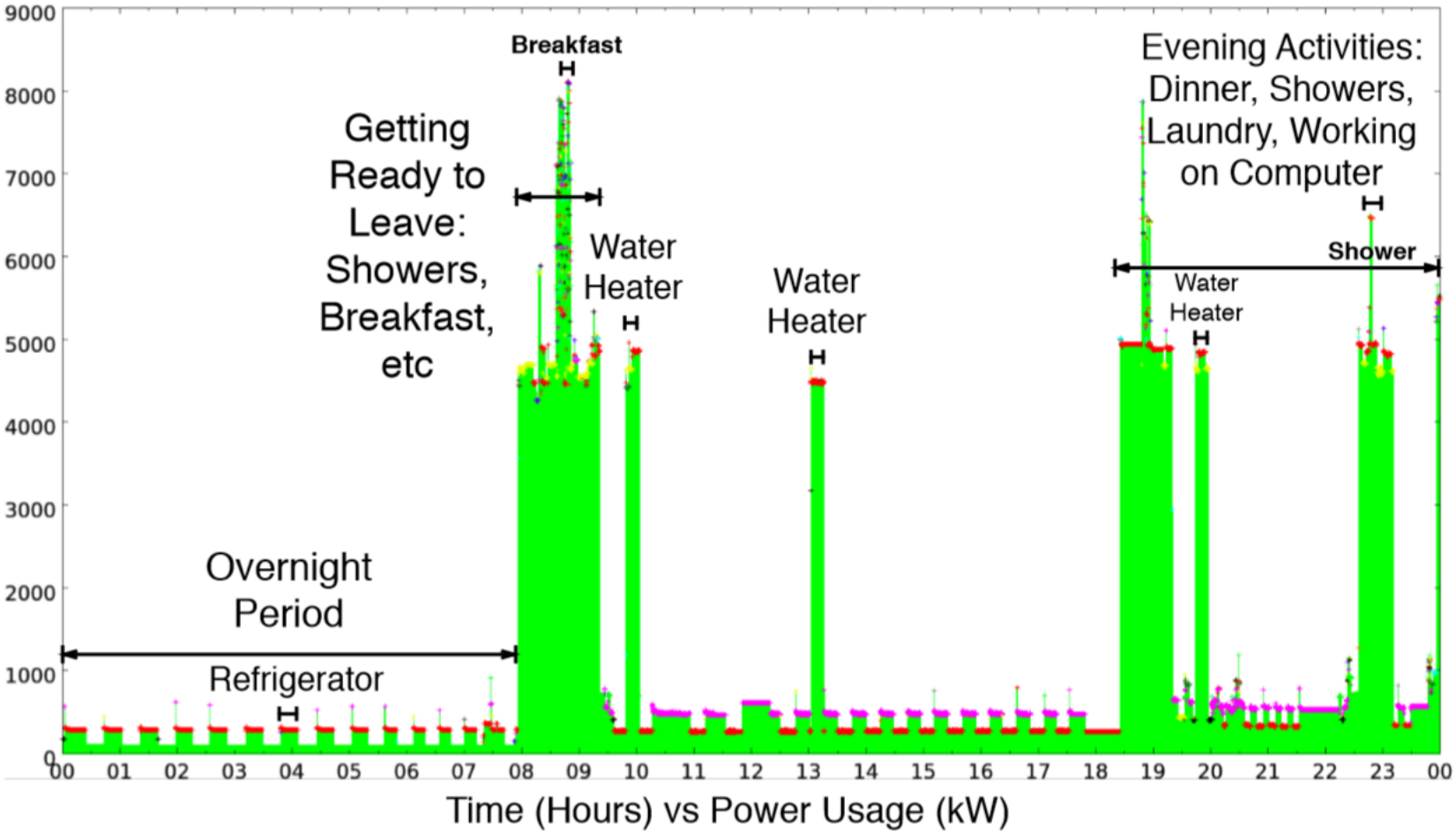
- **Measure** fine-grained energy use
- **Collected data** transmitted by smart meter & aggregated
  - at Energy data center
  - data consumer: utility/third party

Smart meter data enables…

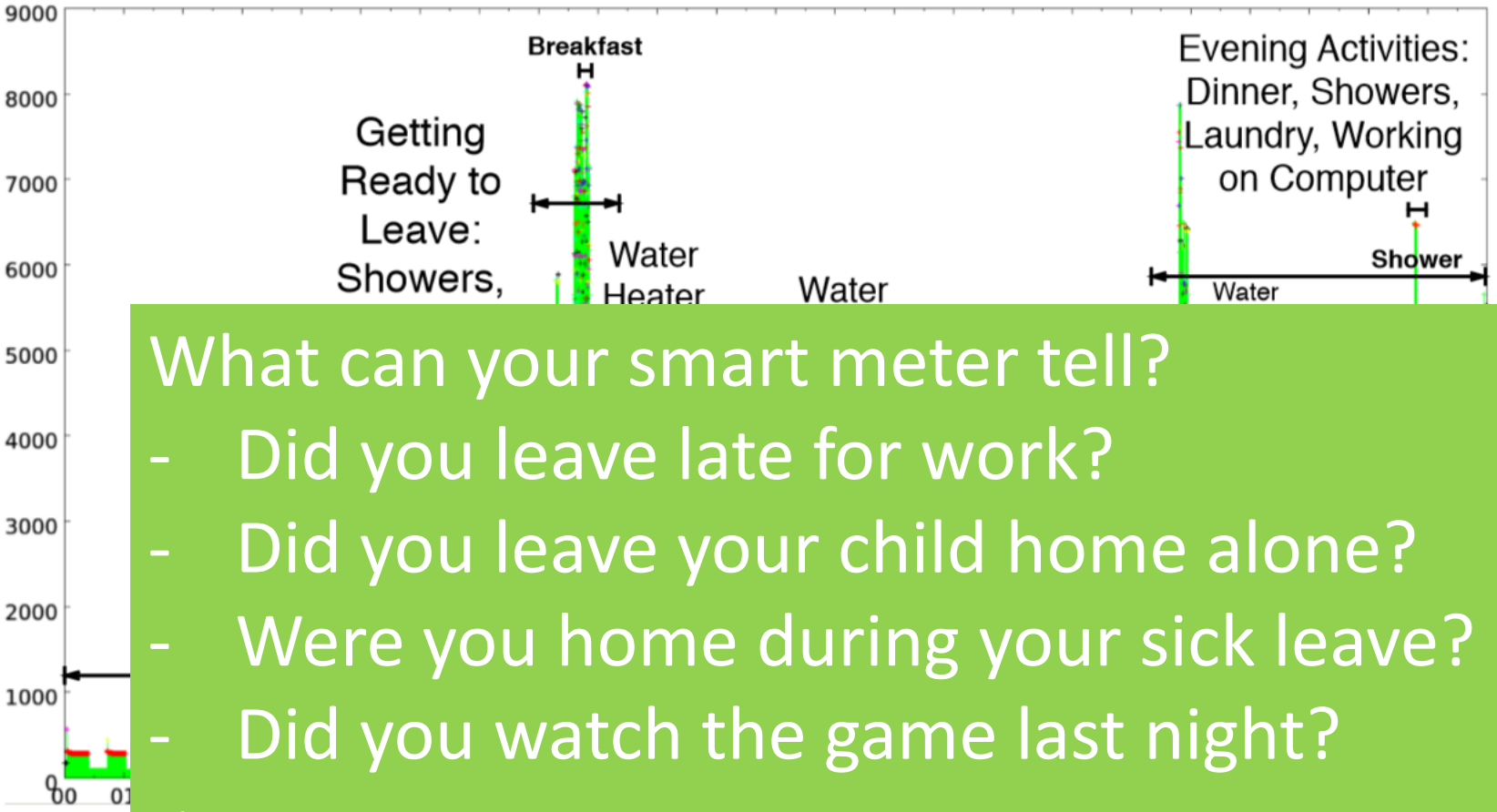
- Billing - with time-of-use pricing
- Fraud detection
- Demand response
- Load monitoring and forecasting
- Power outage notifications
- Energy Efficiency analysis & optimization

etc…

Privacy Concerns



Privacy Concerns



What can your smart meter tell?

- Did you leave late for work?
- Did you leave your child home alone?
- Were you home during your sick leave?
- Did you watch the game last night?

(Molina-Markham et al, Private Memoirs of Smart Meters, BuildSys'10)

Privacy Concerns

- Energy Industry – maximize revenues
- Third-party companies - target marketing material
  - e.g. building & insulation
- Hackers – real-time mass surveillance, burglary

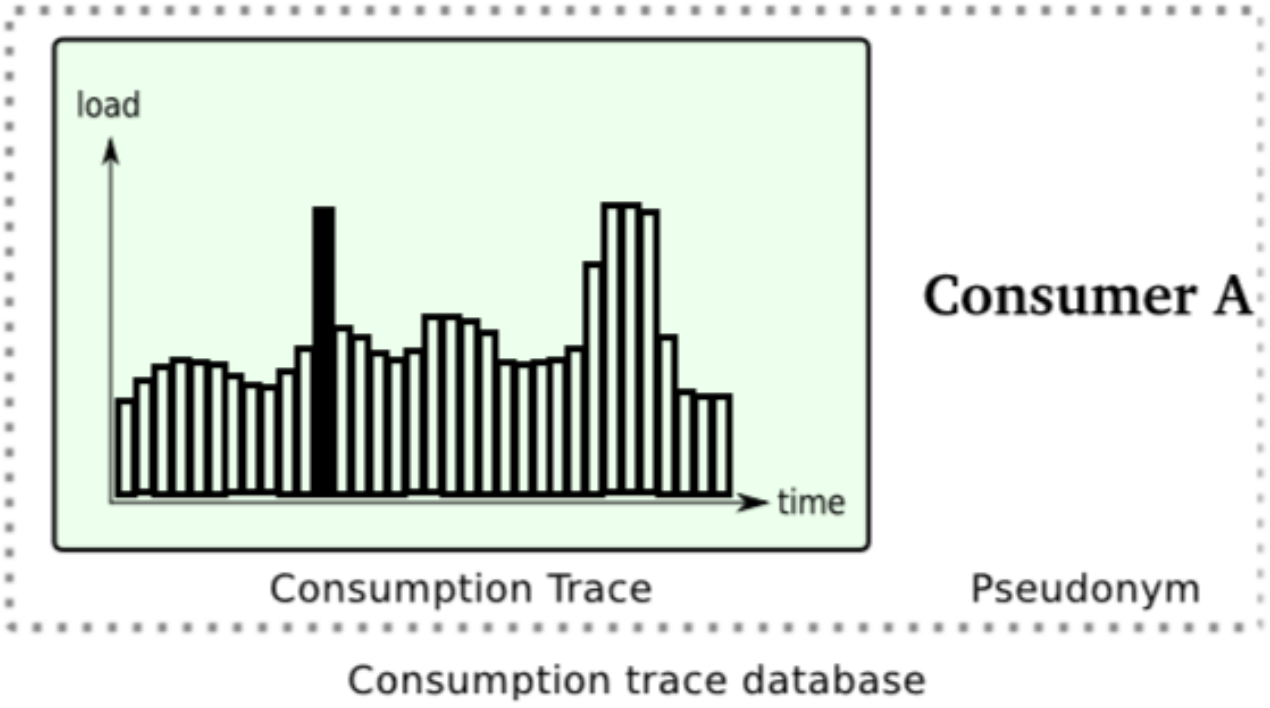
Data Privacy compromised if leak personally identifiable information/attributes

Current Privacy Policies

- Under "*Fair Information Practice Principles*" at Federal
  - Detailed readings - sensitive
  - Requires consumer awareness & consent

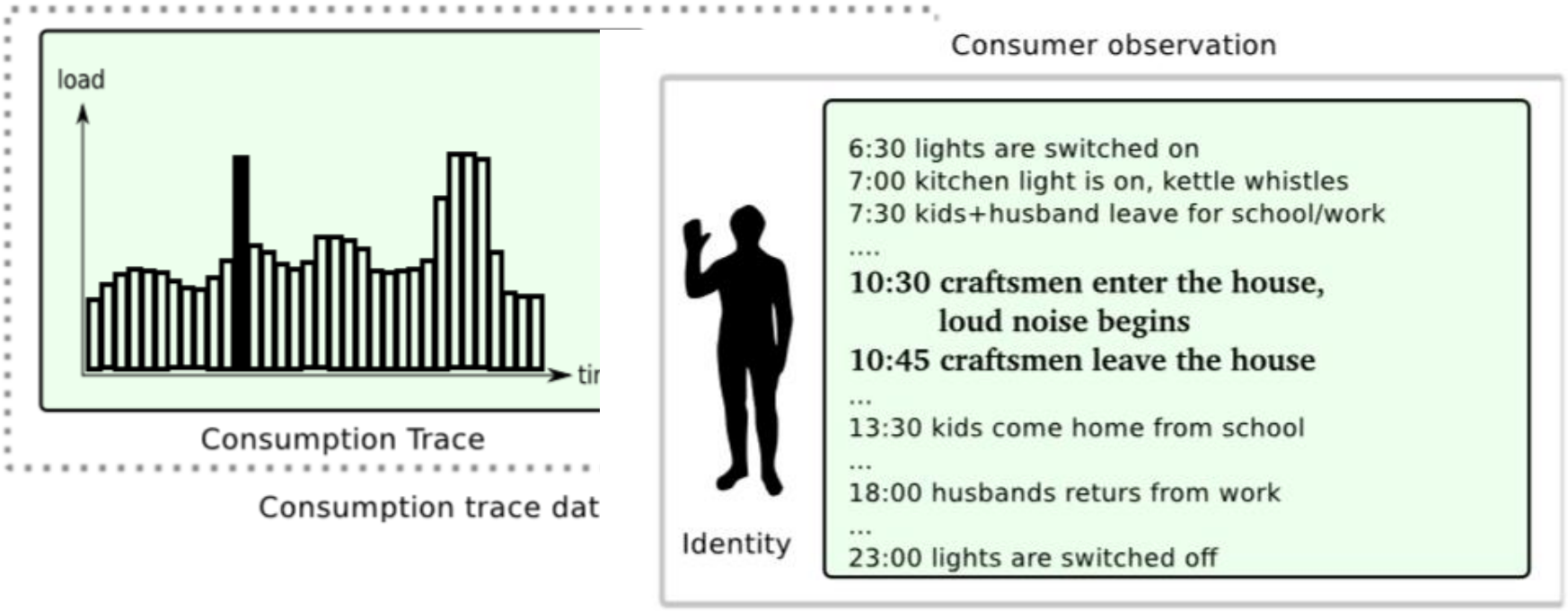
- California Public Utilities protect smart meter data (rulings in 2011 & April 2014)
  - Utilities can't sell customer's personal/consumption data
  - Third parties can't use it for secondary commercial use

Pseudo-nymizing smart meter data…



- Separate consumption trace & household identity

Naïve Pseudo-nymizing is fragile



- Correlate two data sources overlapping in time
- Attack: Linking by anomaly

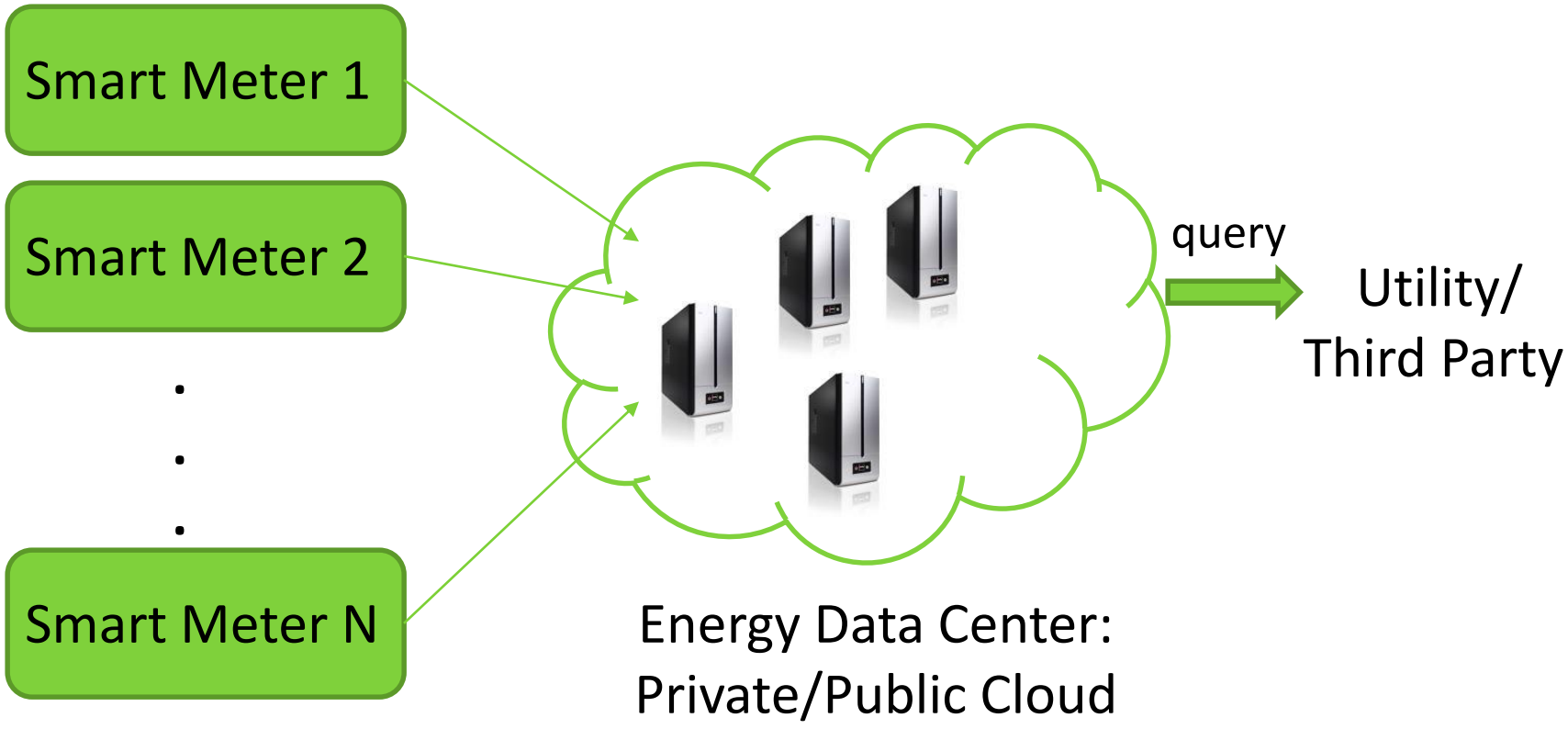
Jawurek et al, *Smart Metering de-pseudonymization*, ACSAC 2011

Privacy-enhancing Technologies

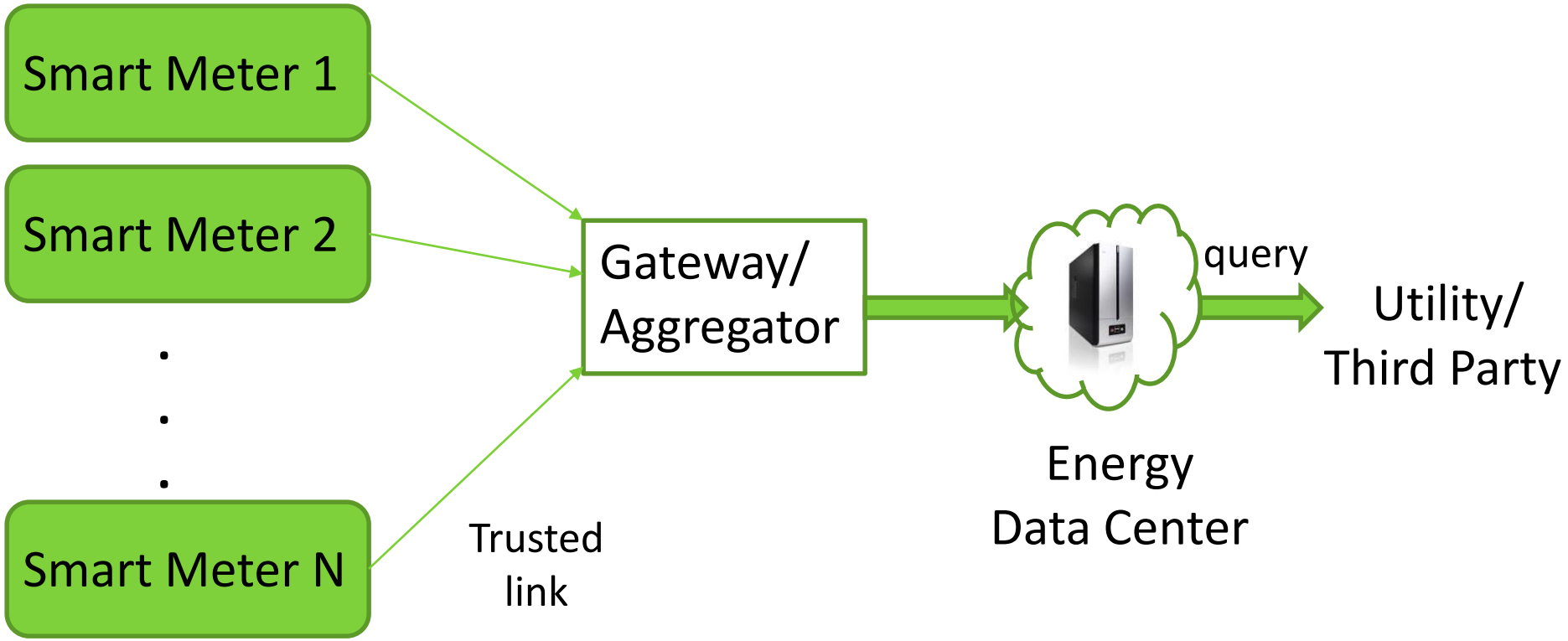
Prevent privacy violations before they occur

- Pseudo-nymizing
- Trusted third party
  - Aggregates
  - Adds noise (differential privacy)
- Cryptographic computation

System Model

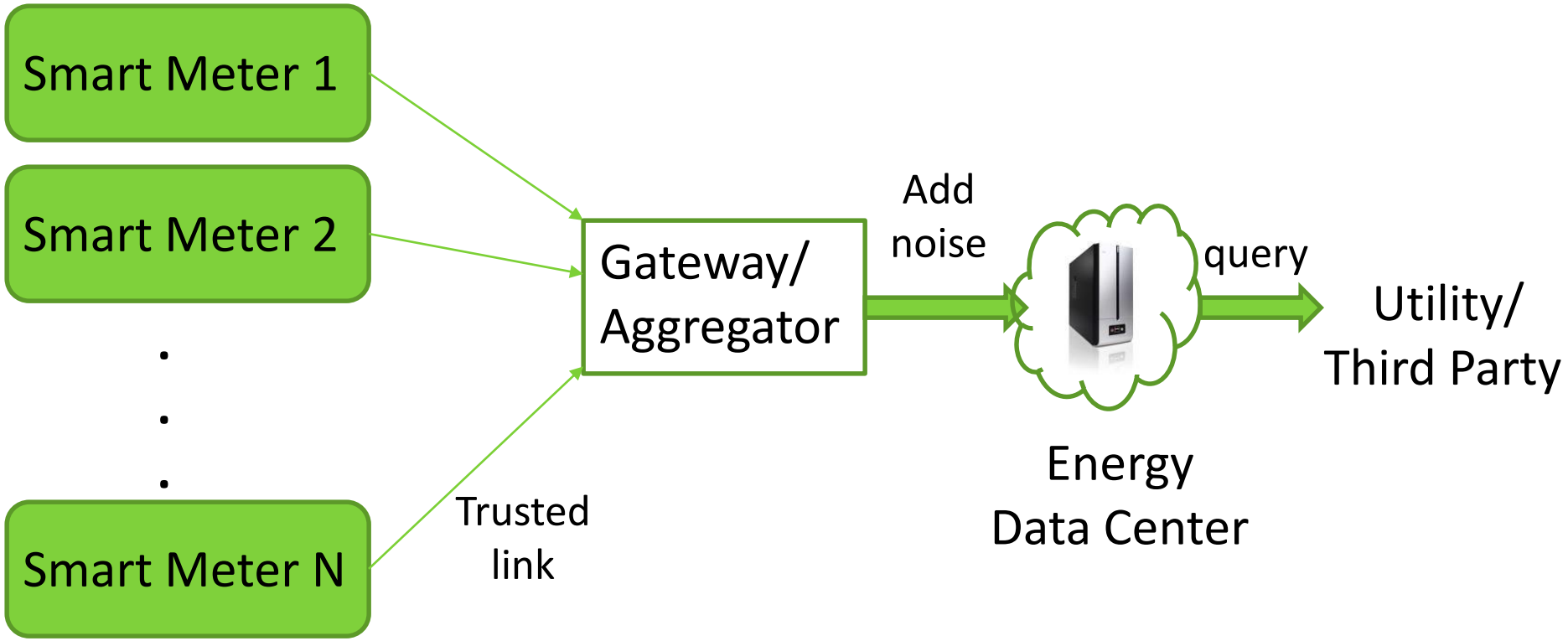


Trusted third party aggregates…



- Gateway aggregates the high-frequency readings
- No private data items sent, yet some individual identifiable

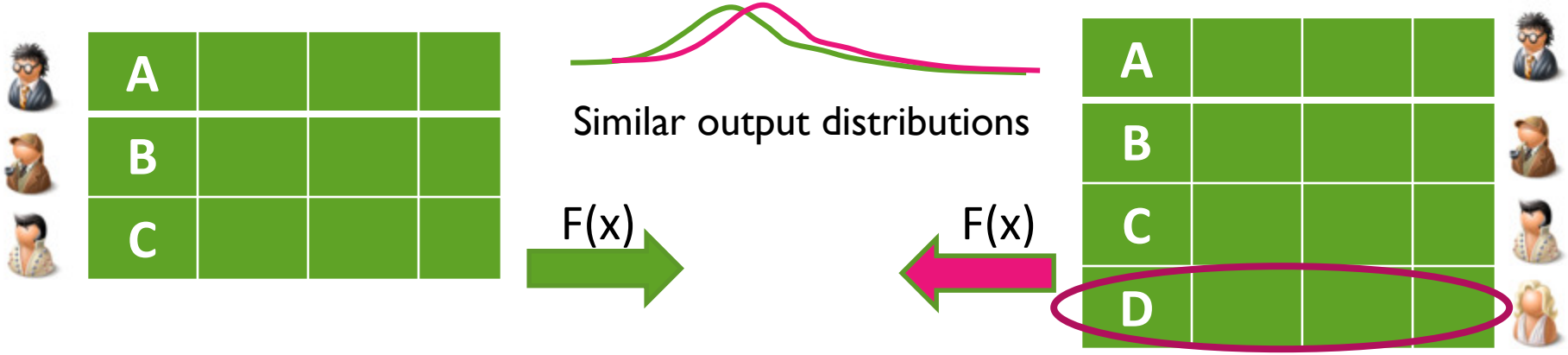
Trusted third party adds noise…



- Differential privacy - add random noise to aggregate

Differential privacy (intuition)

A mechanism is differentially private if every output is produced with similar probability whether any given input is included or not

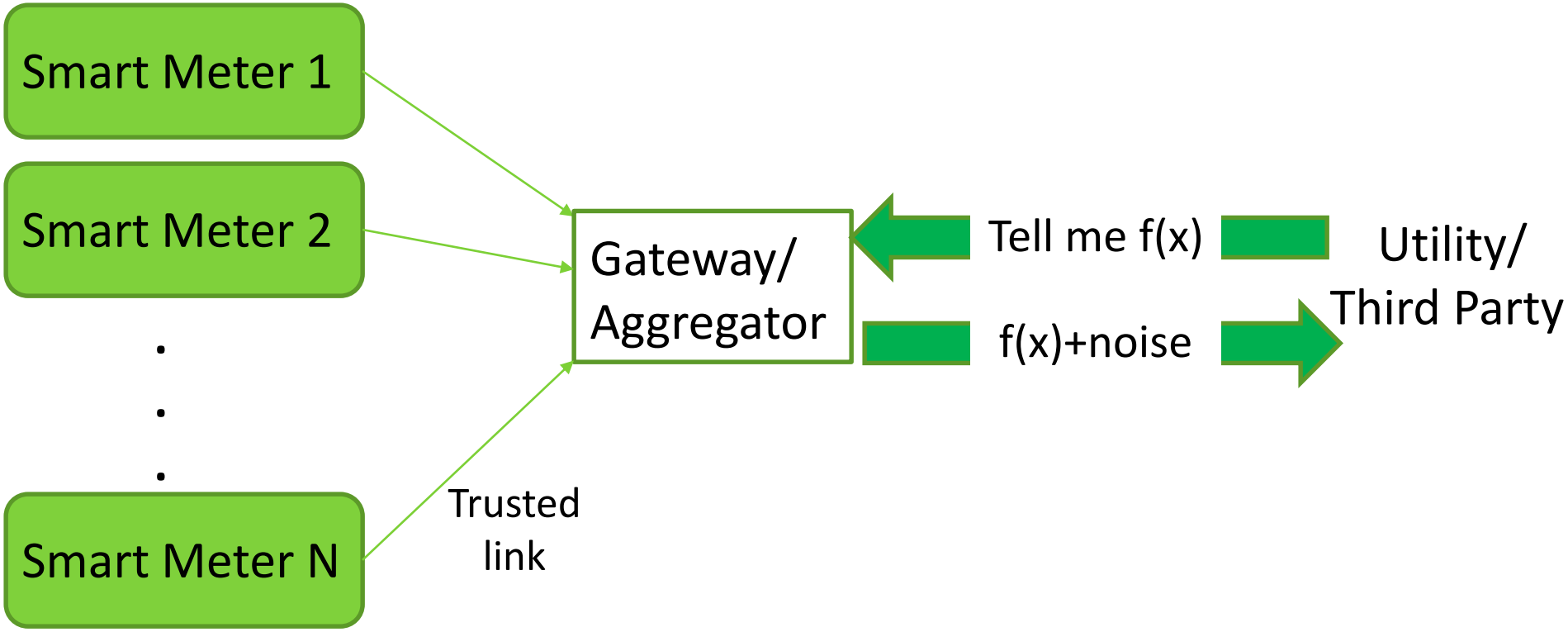


Bounded risk for D if she includes her data!

Cynthia Dwork. *Differential Privacy*. ICALP 2006

Achieving differential privacy

- A simple differentially private mechanism



- How much noise should one add?

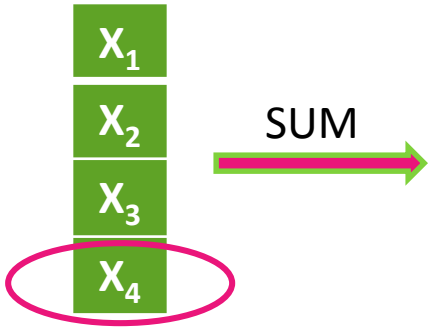
Achieving differential privacy

- Function sensitivity (intuition): Maximum effect of any single input on the output
  - Aim: Need to conceal this effect to preserve privacy

- Example: Computing the aggregate mean of the readings has low sensitivity
  - Any single user's reading does not affect the final mean by too much
  - Calculating the maximum reading has high sensitivity

Achieving differential privacy

- Function sensitivity (intuition): Maximum effect of any single input on the output
  - Aim: Need to conceal this effect to preserve privacy
- Example: SUM over input elements drawn from [0, M]

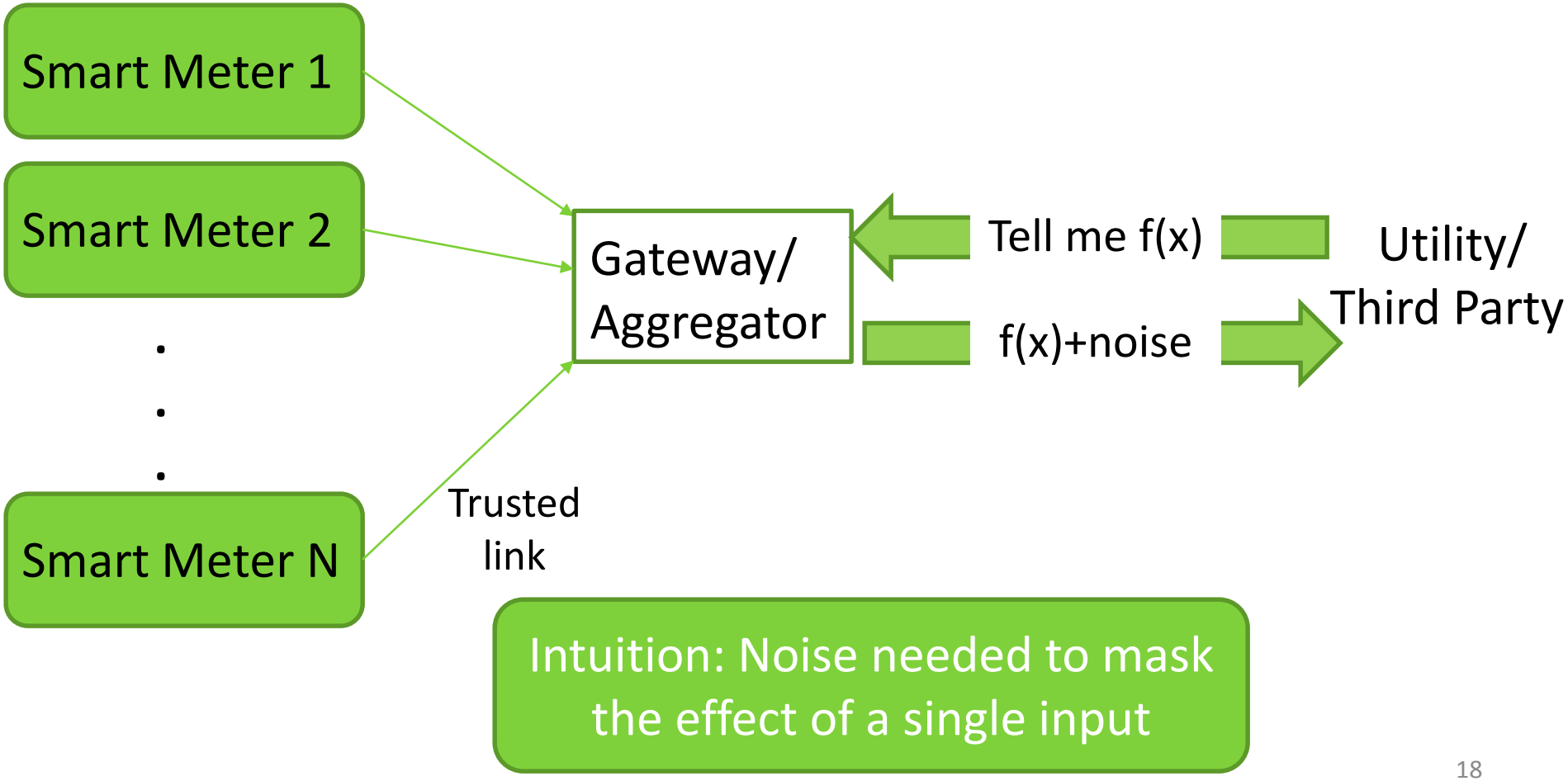


Sensitivity = M

Max. effect of any input element is M

Achieving differential privacy

- A simple differentially private mechanism



Privacy-enhancing Technologies

Prevent privacy violations before they occur

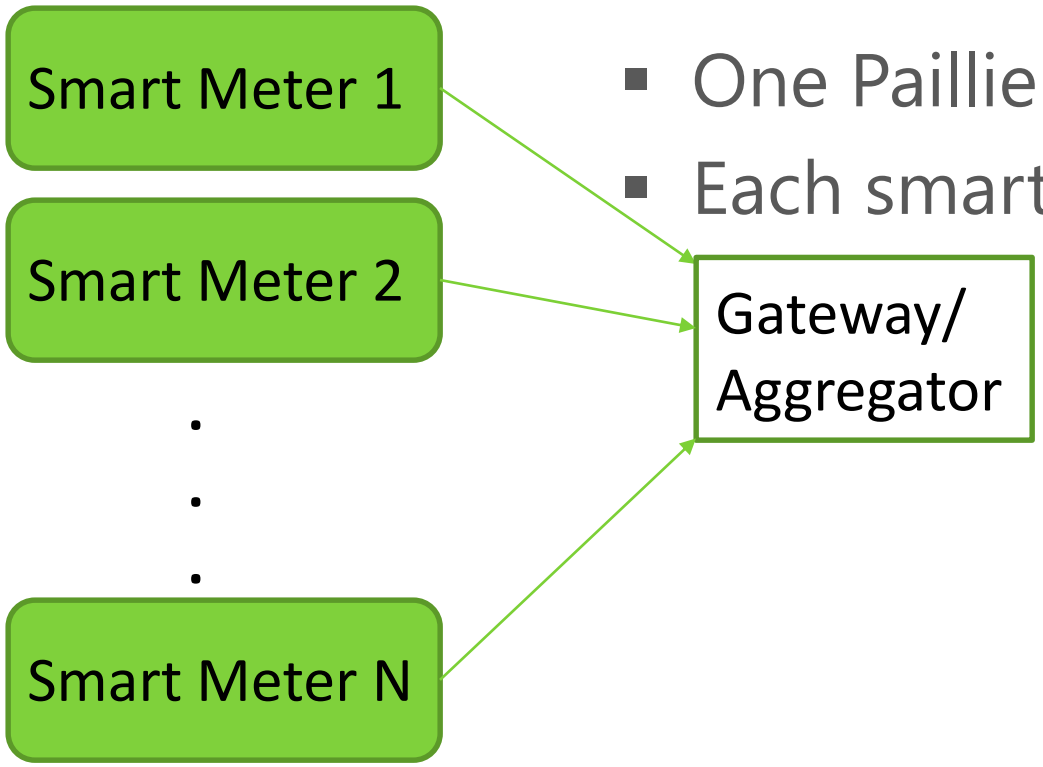
- Pseudo-nymizing
- Trusted third party
  - Aggregates
  - Adds noise (differential privacy)
- Cryptographic computation

Cryptographic Computation

- Strongest privacy/security guarantee
- Aggregate via homomorphic encryption
  - The product of encryptions of two messages is *an* encryption of the sum of the two messages.
  - Paillier cryptosystem - additively homomorphic
- Enables spatial/temporal aggregation

Erkin et al. *Private computation of spatial and temporal power consumption with smart meters,* ACNS 2012

Cryptographic Computation



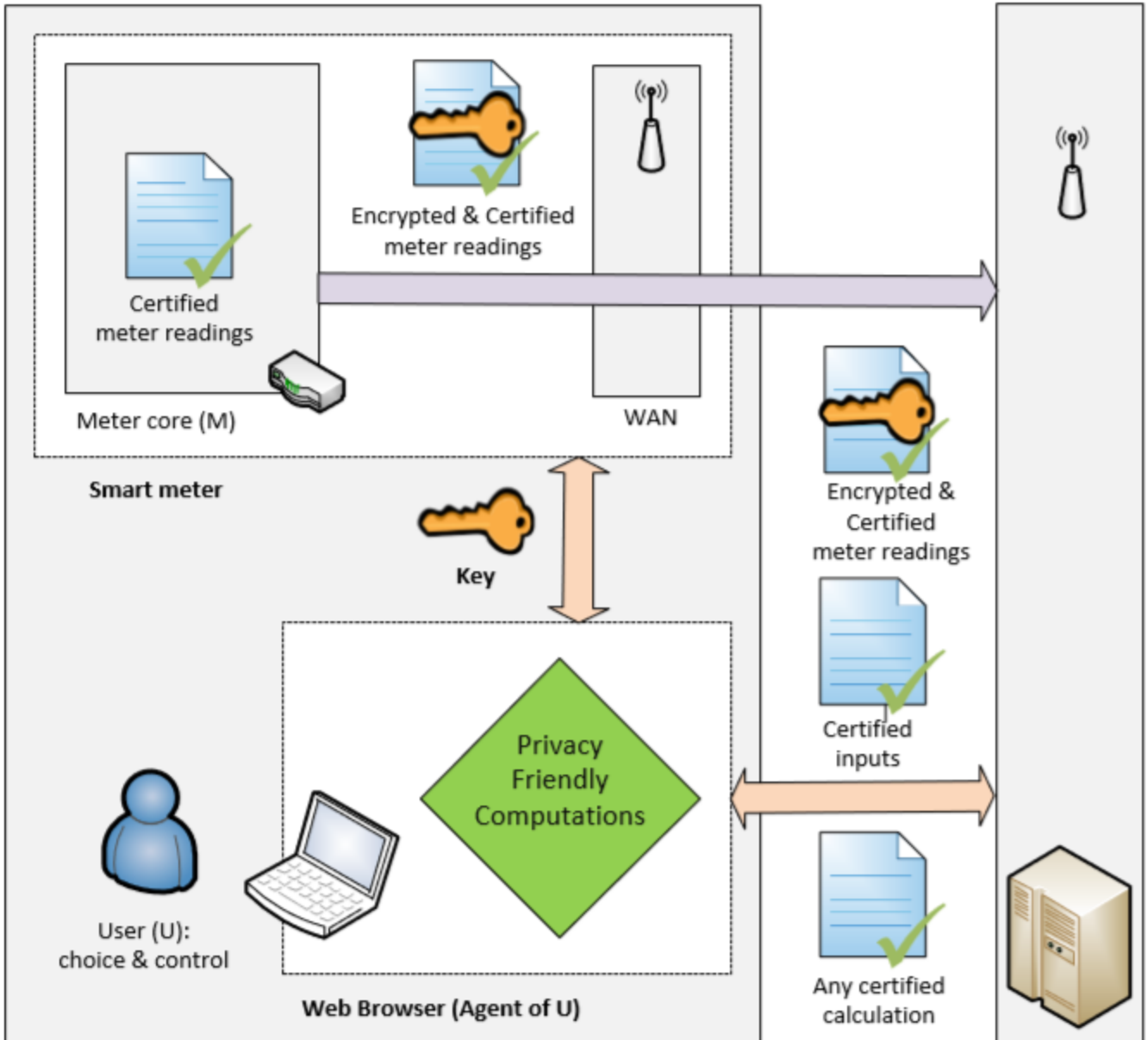
- One Paillier public key
- Each smart meter encrypts
- Aggregator combines the encrypted readings
  - Can decrypt the sum of readings
  - Can't decrypt the individual (modified Paillier scheme)

Erkin et al. *Private computation of spatial and temporal power consumption with smart meters,* ACNS 2012

Cryptographic Computation

- Time-of-use pricing & billing
  - require individual meter readings?
- Integrity
  - certify meter readings and bill calculations?

Cryptographic Computation



Rial et al. Privacy-Preserving Smart Metering; WPES 2011

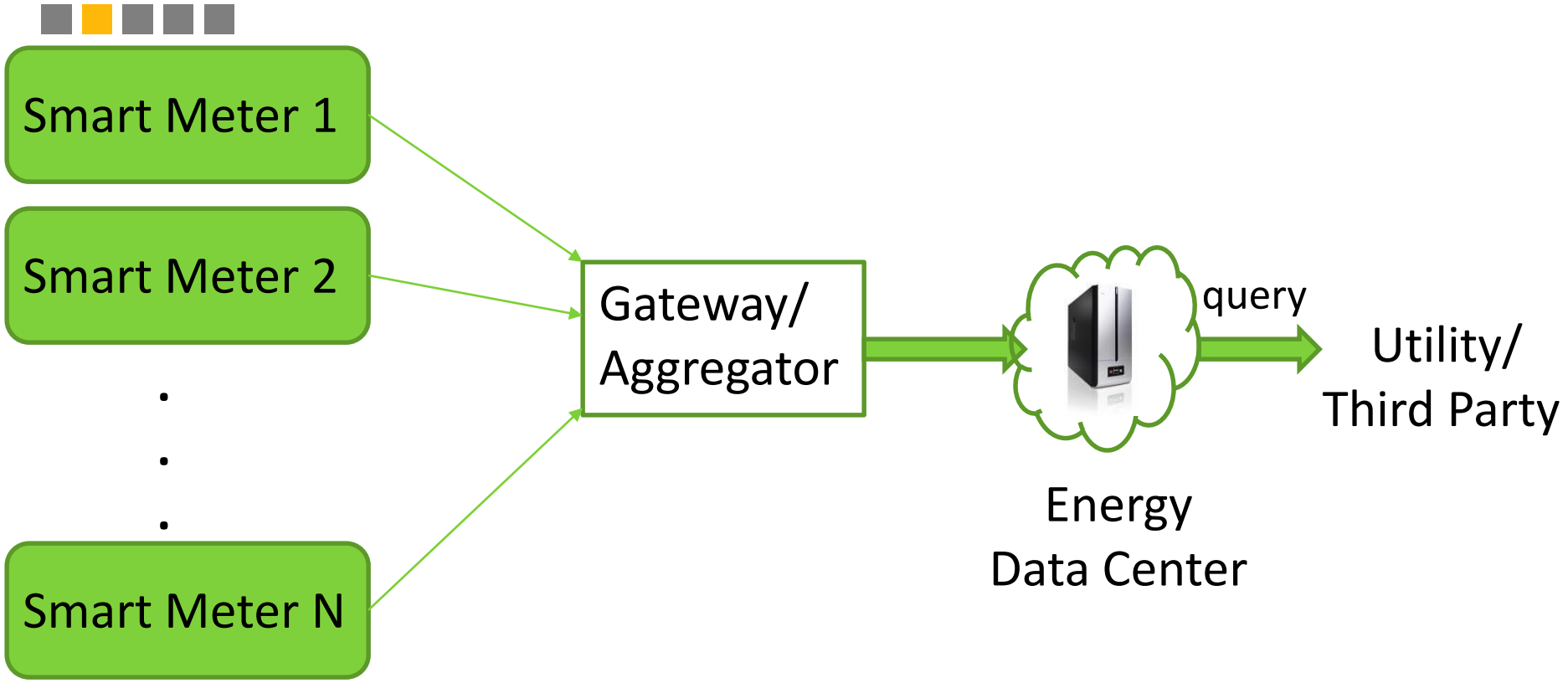
Cryptographic Computation

- Time-of-use pricing & billing
  - require individual meter readings?
  - No – use homomorphic encryption

- Integrity
  - certify meter readings and billing calculations
  - Use **zero-knowledge proof**
    - Smart meter proves to the utility (the *verifier*) that the reading and calculation is true,
    - Doesn't reveal individual readings

Rial et al. Privacy-Preserving Smart Metering; WPES 2011

Recap: Privacy-enhancing Technologies



- Pseudo-nymizing
- Trusted third party aggregates & adds noise
- Cryptographic computation

Implementation Overheads

- Smart meter: low computation power & memory
  - No overhead with Pseudo-nymizing & trusted third party
  - Additional computation/hardware for cryptographic
- Communication bandwidth
  - Pseudo-nymizing < Trusted third party <= Cryptographic
- Computation at the aggregator
  - Increases with the complexity of the protocol
- Scalability

Conclusions

- Smart-meter data can be privacy intrusive
  - Personally identifiable information
  - Time granularity matters
- Anonymizing the readings is not sufficient
- Privacy-enhancing technologies can prevent privacy violations before they occur
  - Trusted third party can aggregate the data & add noise using differential privacy
  - Cryptographic computation enables verifiable spatio-temporal aggregations

THANK YOU!