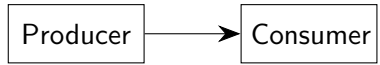
An Information-Theoretic Approach to Time-Series Data Privacy

W-P2DS 2018

Yousef Amar
Hamed Haddadi, Richard Mortier

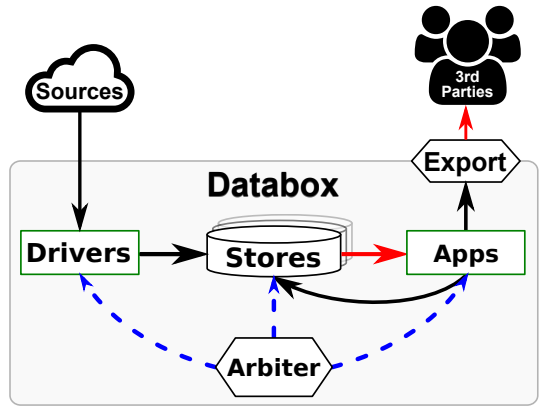
Problem

- Opaque privacy contexts
- Coarse access control
- Context-dependent filtering
- How can we measure privacy and risk online and adjust the flow of data based on risk?

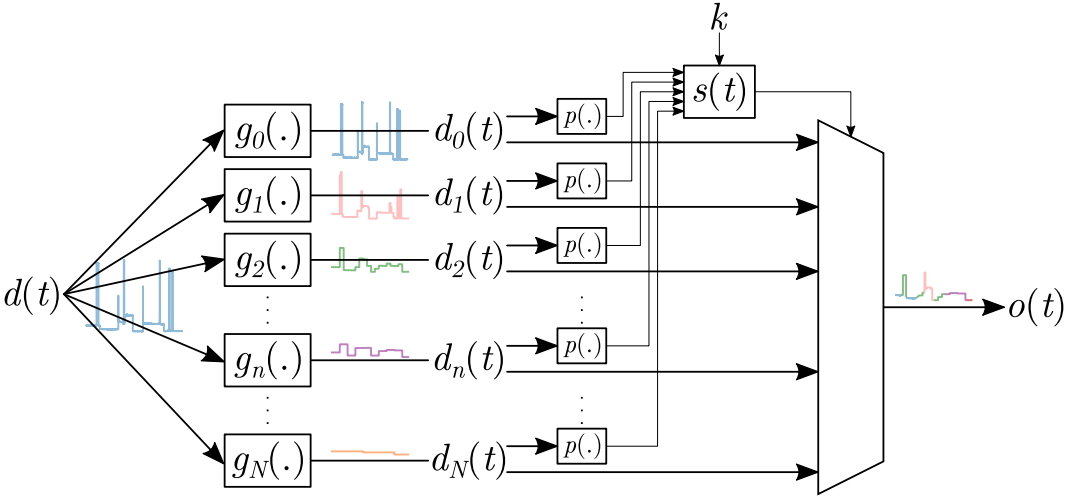


Context

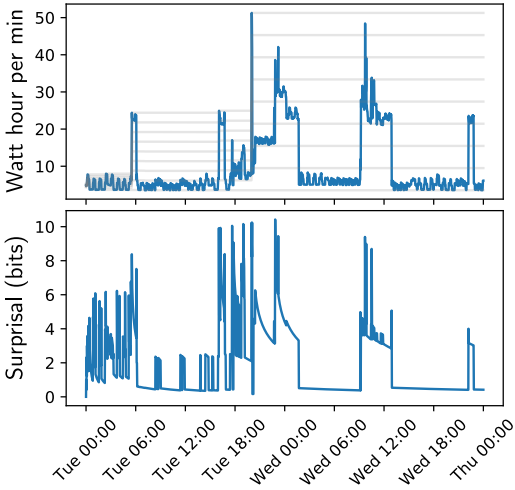
- Home IoT devices
  - Low-latency
  - Limited resources
- Streaming, high-frequency time series data
- Implemented over the Databox platform

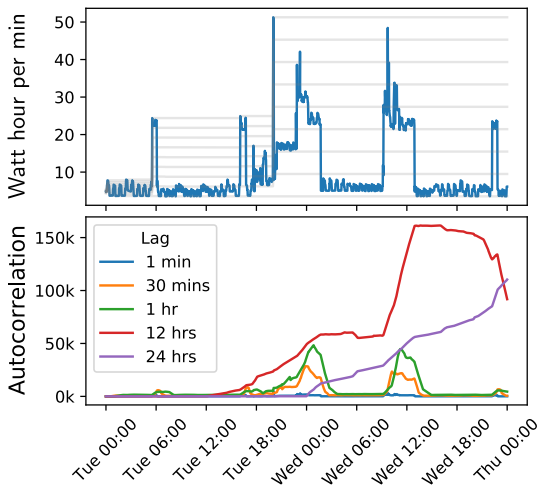


Implementation



Implementation





Evaluation

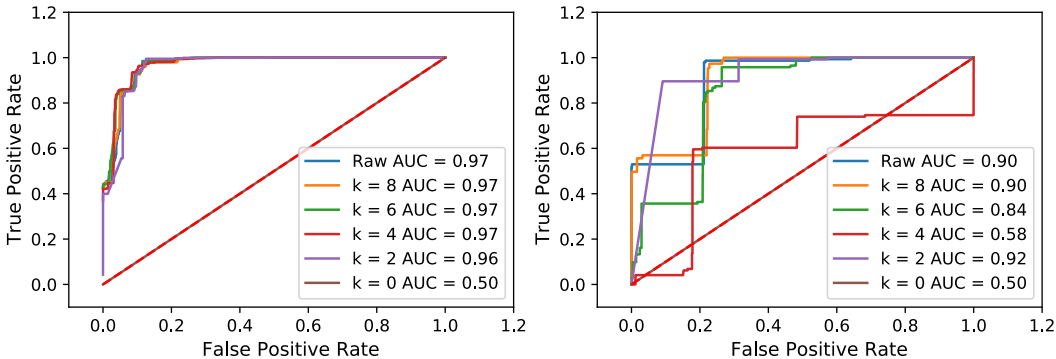


Figure: Receiver Operating Characteristic (ROC) curves for washer-dryer (utility; left) and microwave (attack; right)

Evaluation

- Gains in privacy
- Without impacting utility
- Negligible latency overhead
- Future Work
  - Mutual information
  - Smooth interpolation between levels of granularity
  - User-defined policies

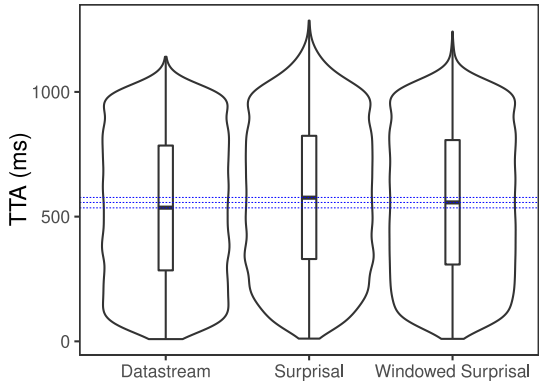


Figure: Distributions of time to availability under different conditions

Thank you for your attention!

Questions?

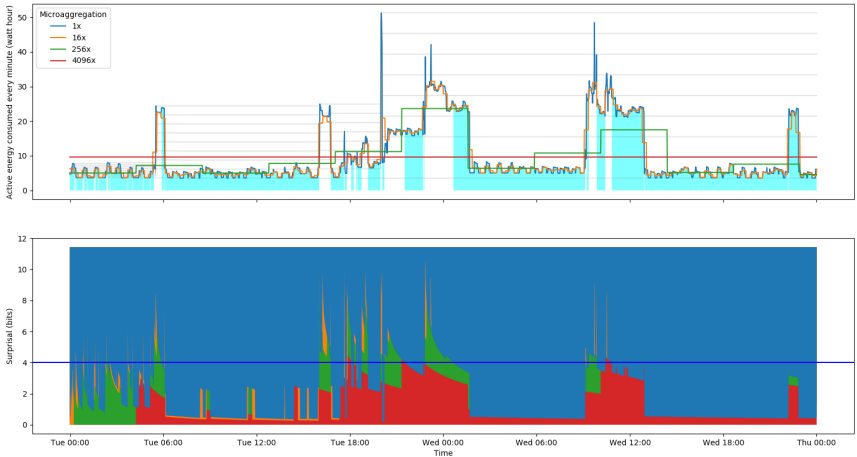
More info: `http://www.databoxproject.uk/`
Contribute: `https://github.com/me-box`

Surprisal

The self-information $\mathrm{I}(\omega_n)$ associated with outcome ω_n with probability $\mathrm{P}(\omega_n)$ is defined as:

$$\mathrm{I}(\omega_n) = -\log(\mathrm{P}(\omega_n)) = \log\left(\frac{1}{\mathrm{P}(\omega_n)}\right)$$

Thresholds


Microaggregation
1x
16x
256x
4096x
Active energy consumed every minute (watt hour)
0
10
20
30
40
50
Surprisal (bits)
0
2
4
6
8
10
12
Tue 00:00
Tue 06:00
Tue 12:00
Tue 18:00
Wed 00:00
Wed 06:00
Wed 12:00
Wed 18:00
Thu 00:00
Time