EuroDW2017

Toward Privacy-Preserving IoT Data Publishing

by

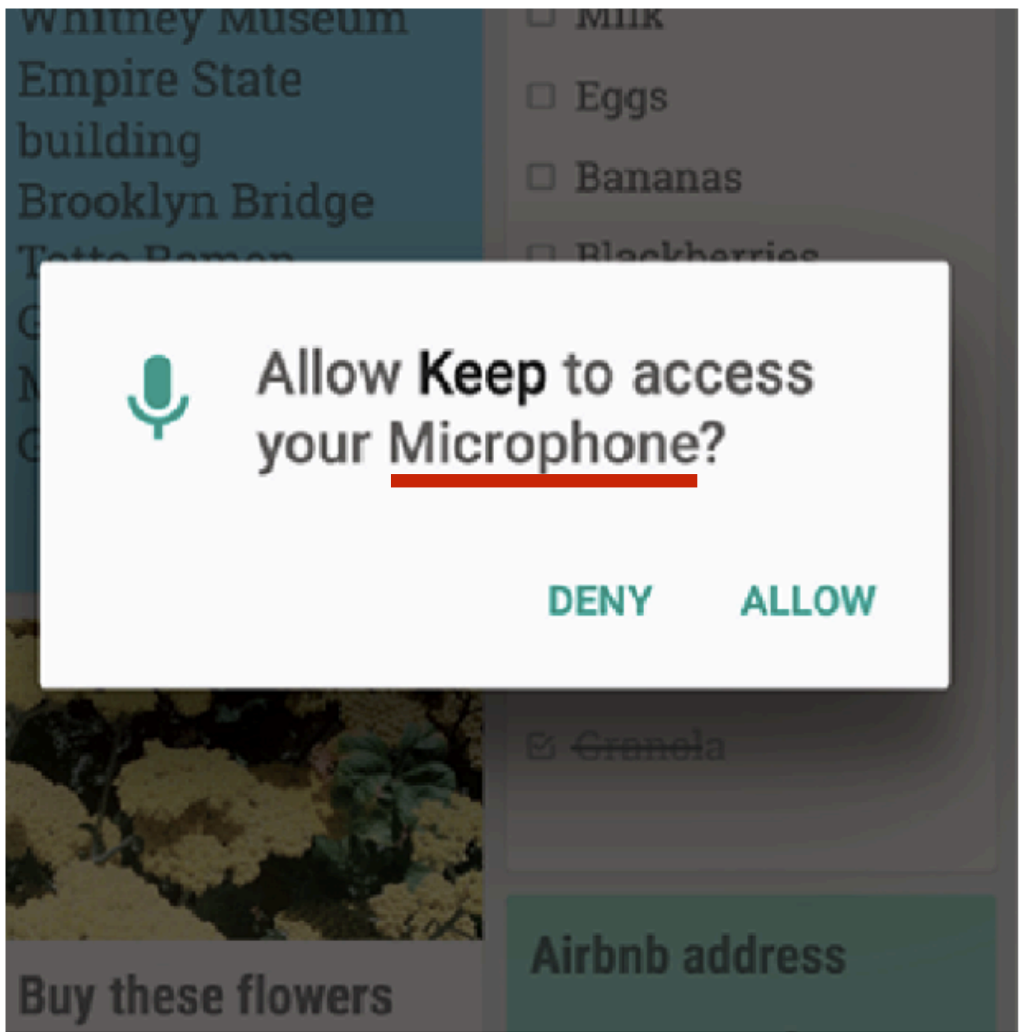
Mohammad Malekzadeh

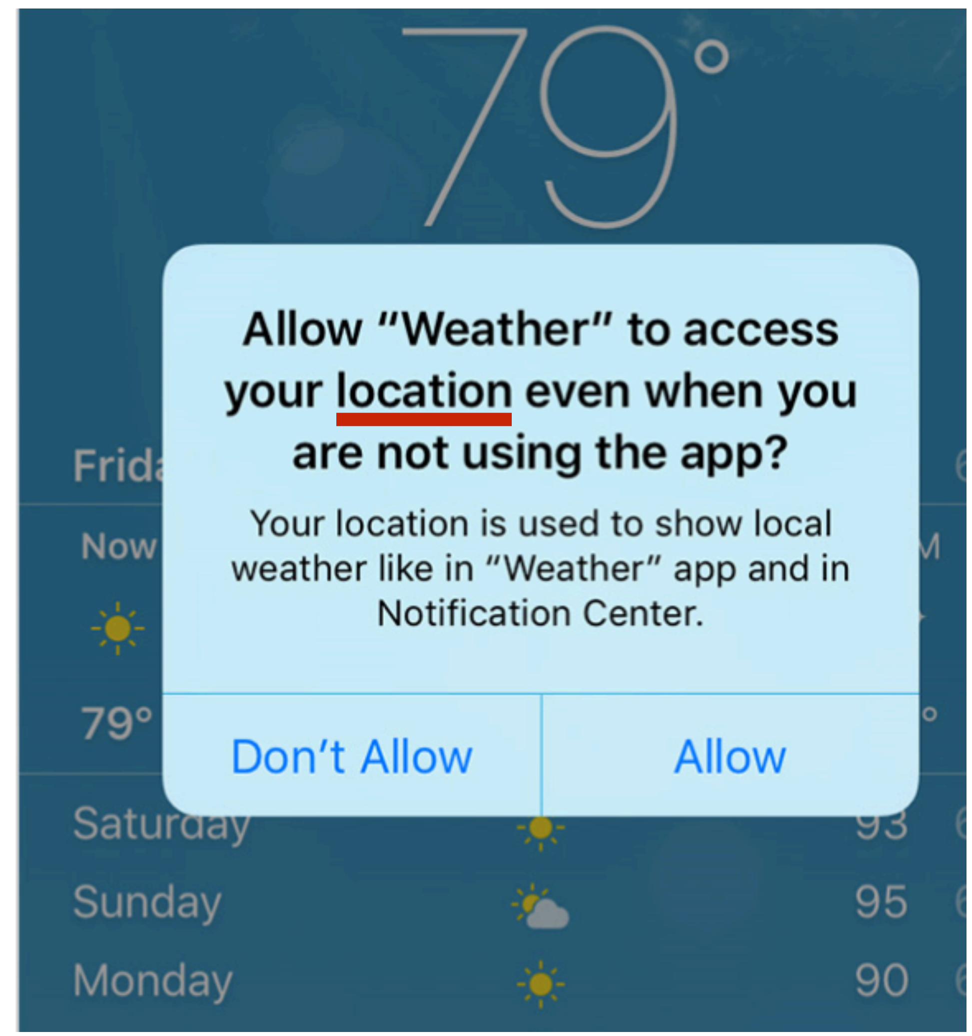
m.malekzadeh@qmul.ac.uk

April 24th 2017 - Belgrade

To begin with

User permission dialogs for Android and iOS[1]





Two valuable sources of information

Audio Signal and **Location**

AGENDA

A BACKGROUND

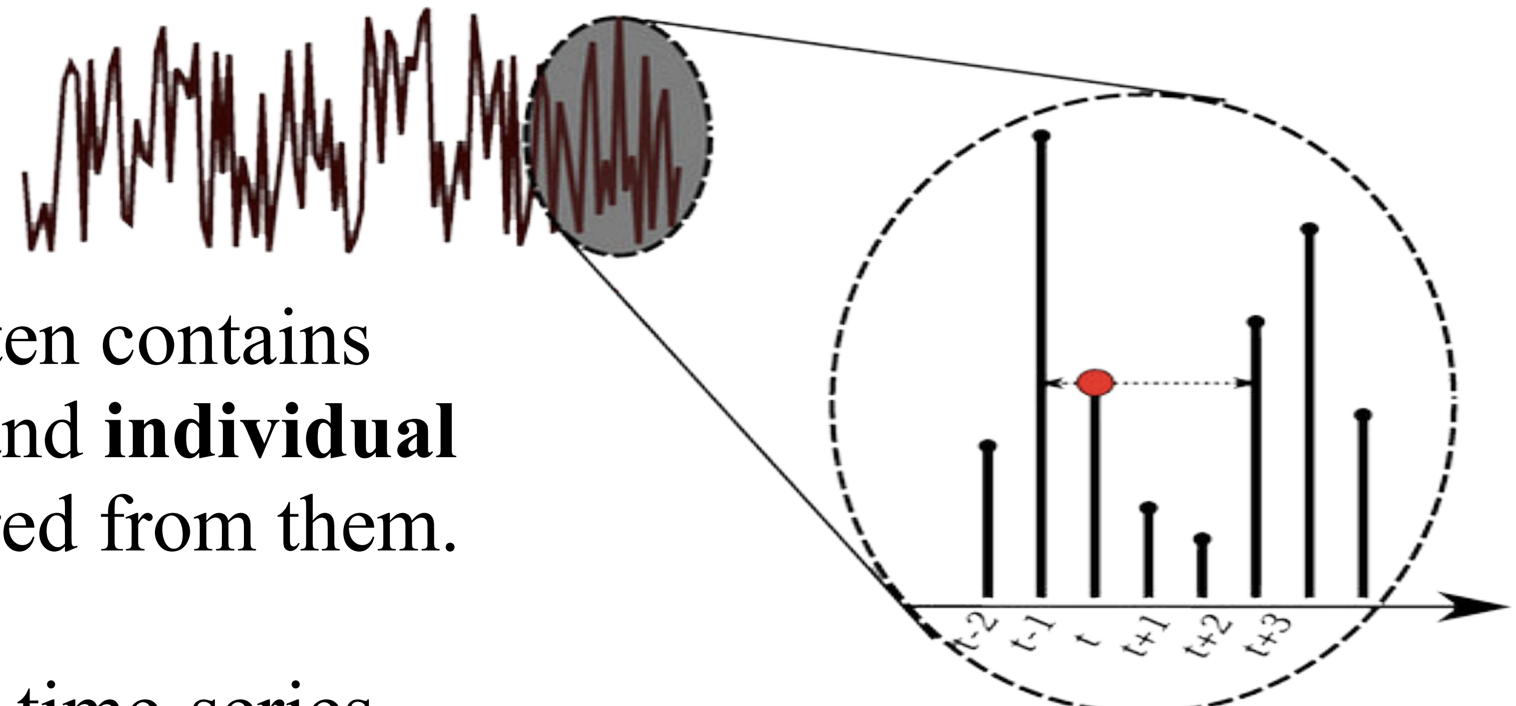
B TECHNIQUES

C CHALLENGES

D OUR IDEA AND FOCUS

A

Dichotomy



- Personal **time-series** often contains **precious information** and **individual behaviour** can be inferred from them.
- Publishing/Mining such time-series immediately can violates **individual privacy**.
- But, users **benefit** from sharing and mining their data by external entities.
  - healthcare – home security – traffic controlling – lifestyle improvement …

A

Disclosure

- There are two types of **disclosure risk**:

1) Identity disclosure:

- The association of a user's identity with a disseminated data containing confidential information

2) Attribute disclosure:

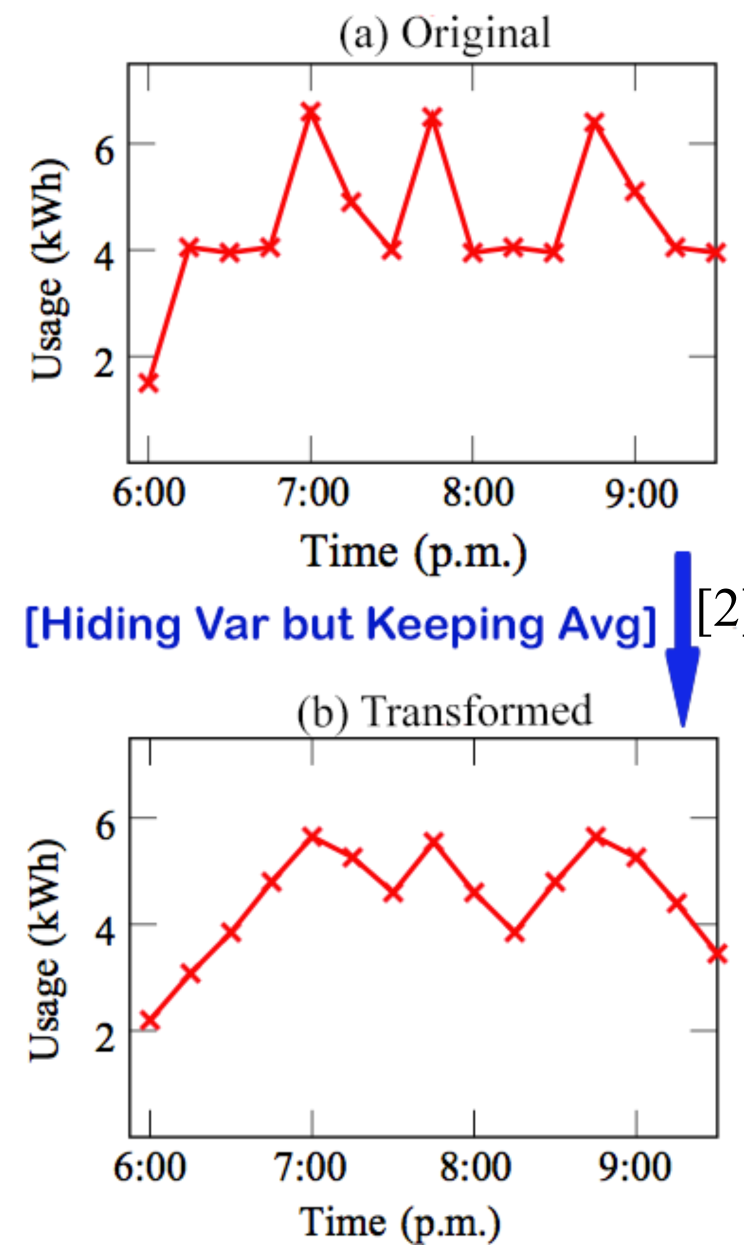
- The association of an attribute value based on the disseminated data with the user

Problem Definition

- By sharing personal IoT data:

1. **original time-series** mustn't be accurately **reconstructed**; also, **sensitive information** shouldn't be **inferred**;

2. yet, some **agreed statistics** should be accurately **estimated** despite the transformation

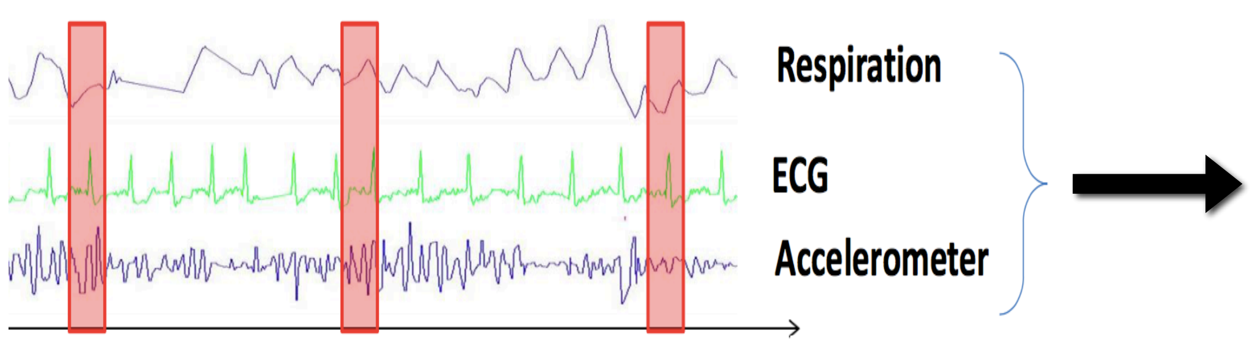


Privacy vs. Utility

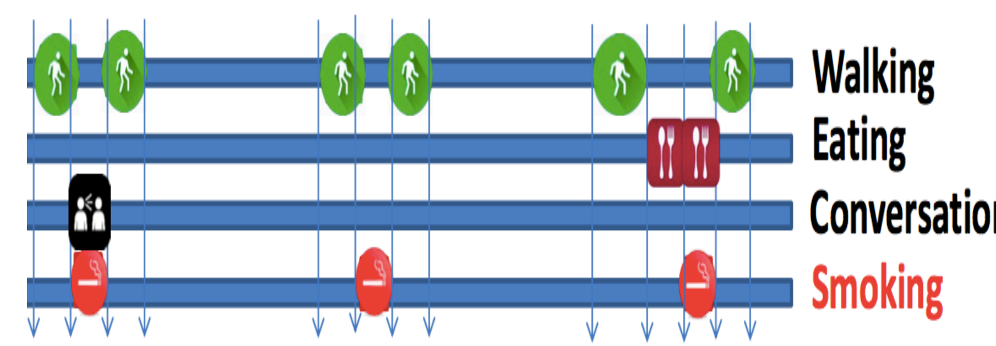
A

i. A **privacy breach** will occur when **user's sensitive information** can be inferred from published data.

ii. A **utility loss** will happen when **user-aware insensitive information** cannot be inferred by third parties' application.



Original Time-Series[3]



Several Inferences

Techniques and Threats

B

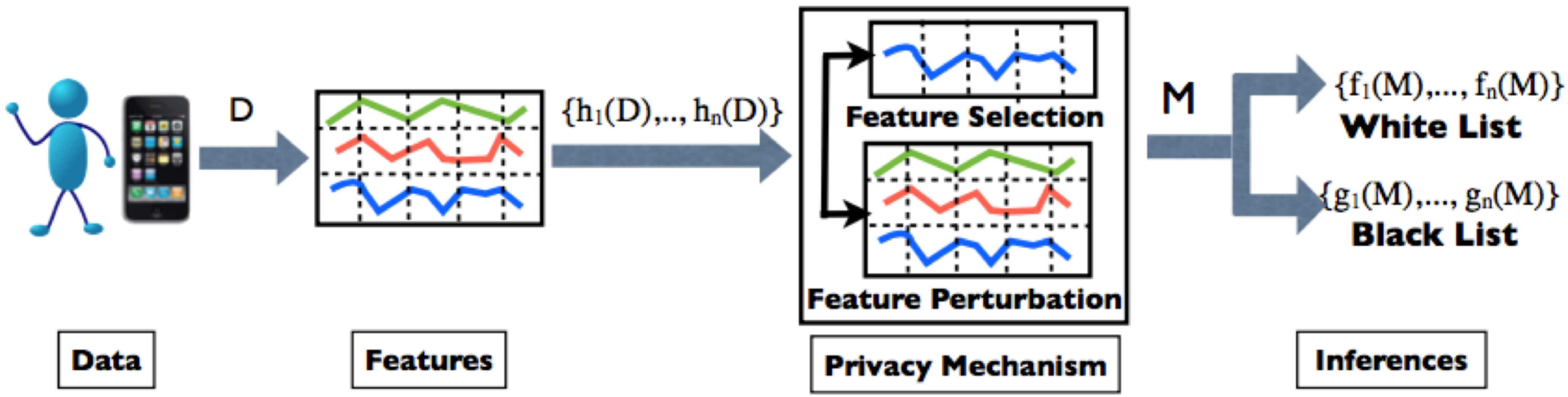
- **Anonymisation: removing personally identifiable information**
  - Threat is linkage attack using other auxiliary data sources to infer the sensitive attributes of individuals within the same dataset.

- **Randomisation: protecting the sensitive information contained in real data by adding some noise**
  - Threat is removing the noise in the data in such a way that it fits the aggregate structure of the data.

- **Data Synthesis: protect privacy and confidentiality of authentic data**
  - Threat is trying to separate real data from synthetic ones.

B

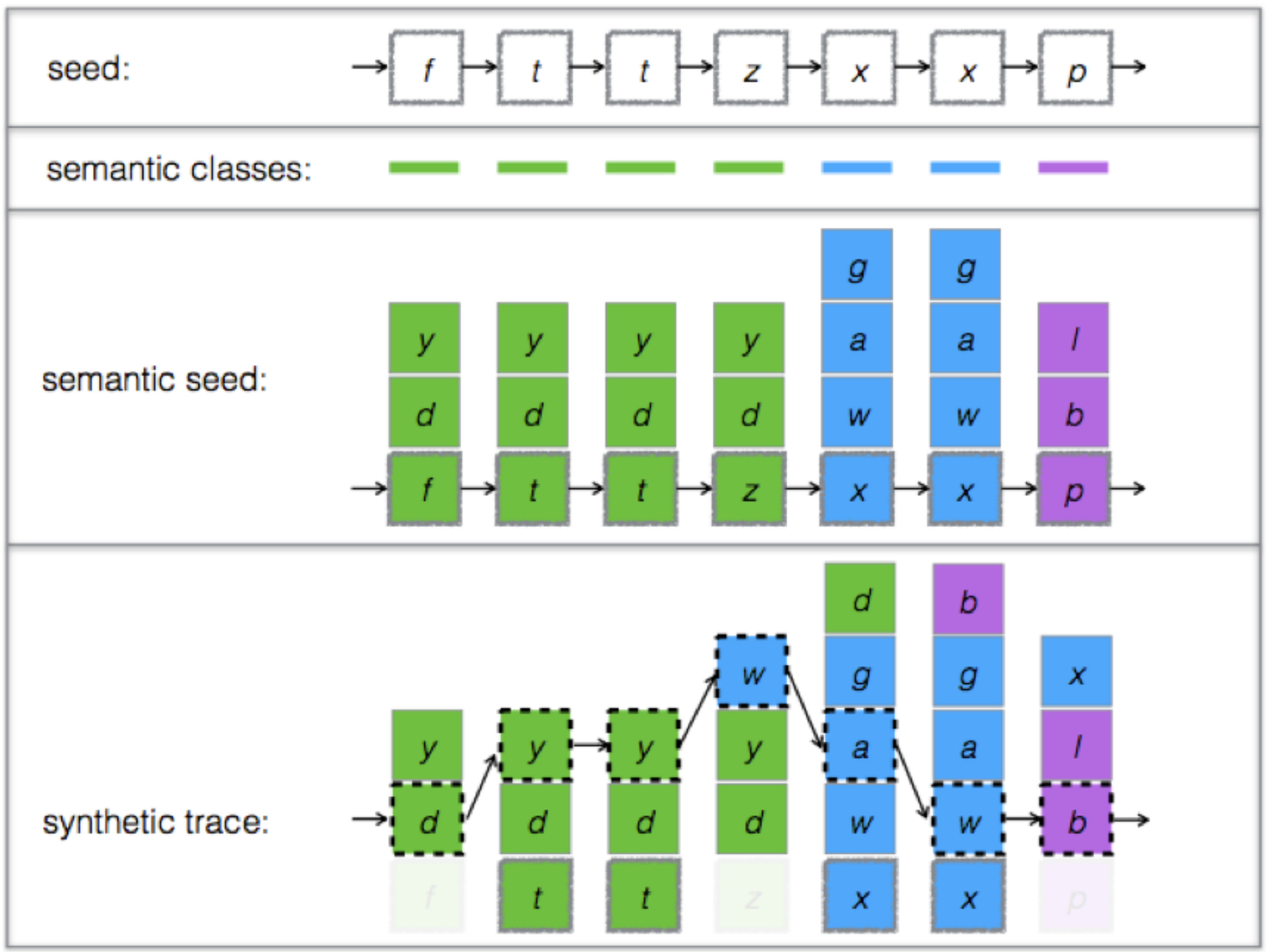
White & Black List



Feature Selection and Perturbation in Time-Series[4]

B

Location Obfuscation



Generating Fake Time-Series from Original One[5]

C

Degree of Perturbation

- current approaches devised more for the static datasets; for **dynamic cases of time-series** data, **scalability challenges** are very crucial
  - need to **further perturbation** for coping with the correlation across time-series
  - privacy-utility trade-off over time is often obtained through **optimisations**

C

Privacy Measure

- Definition of proper **measures** that can truly assess the **privacy of the protection**
  - some measures focus on protecting the identity of a participant
  - other measures focus on the amount of sensitive information that is contained in the noisy data.
- An efficient privacy measure is more **application-specific**

C

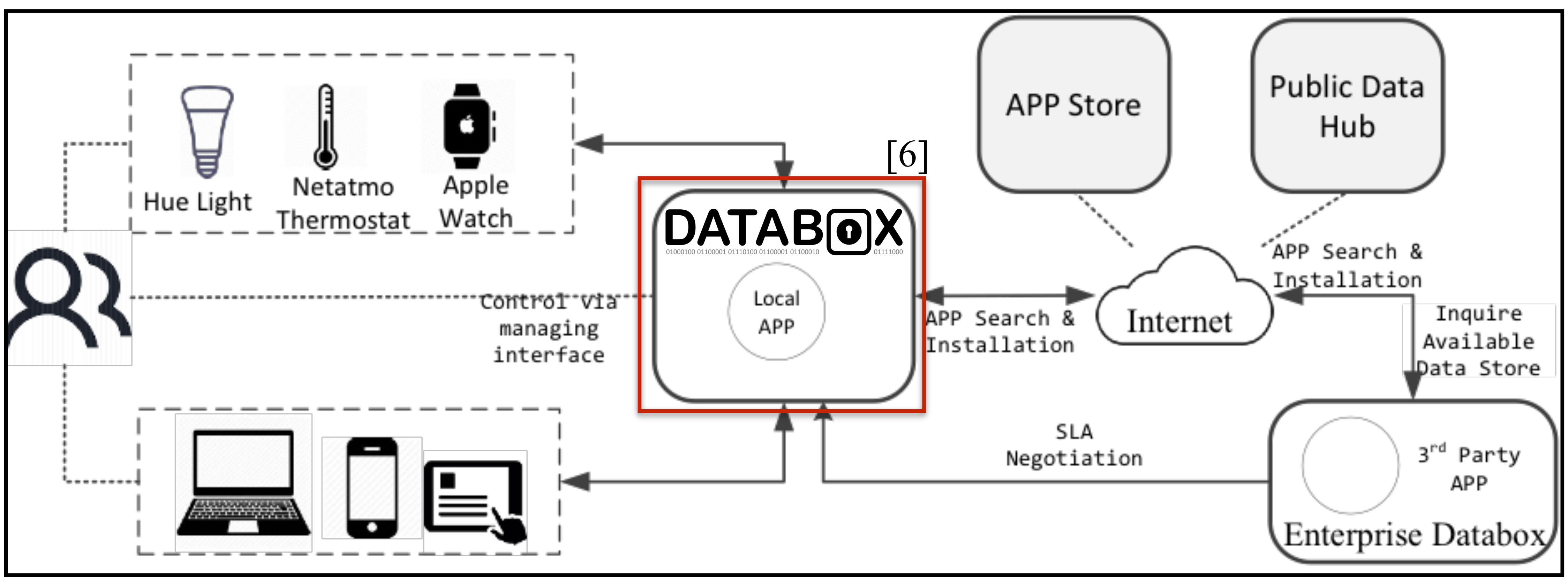
Variability in Sensitivity

- Some applications must **protect aggregate secrets** while others must protect **secrets about individuals**
- In some cases, only **certain attributes** need to be protected, and **certain individuals** may require more privacy protection than others
- Some time-series data carries **disparate information** about **various user behaviours**

D

Databox

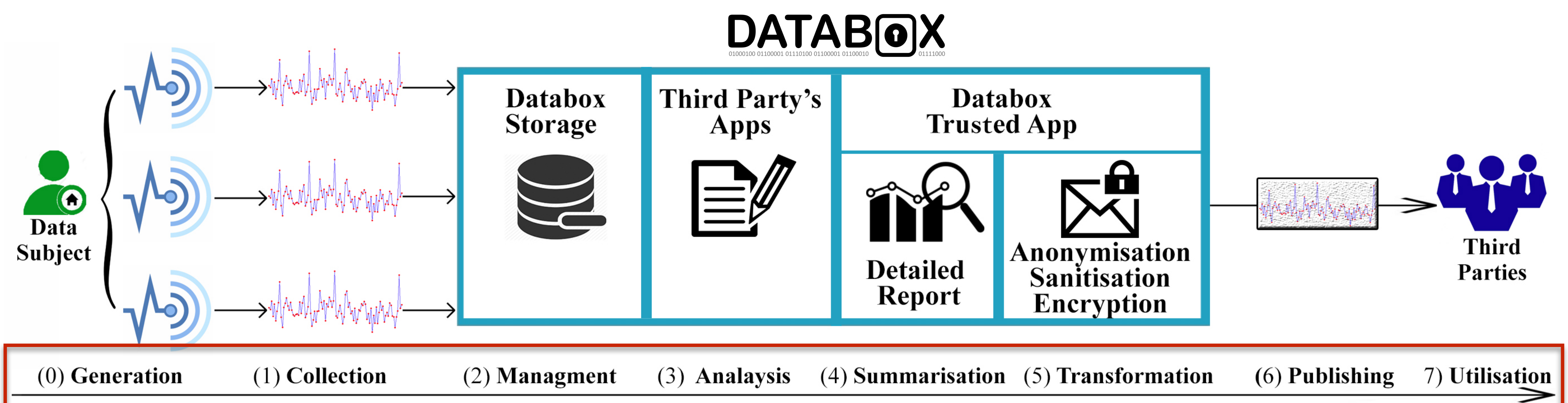
- ☑ Enforcing accountability and control by design at the users' end.
- ☑ Processing of personal data can be done locally
- ☑ Publishing IoT time-series data as a part of Databox.



D

IoT Data Flow

* Third parties' apps installed on Databox are able to request for several data from different sources and perform desired analytics on the large quantity of data.

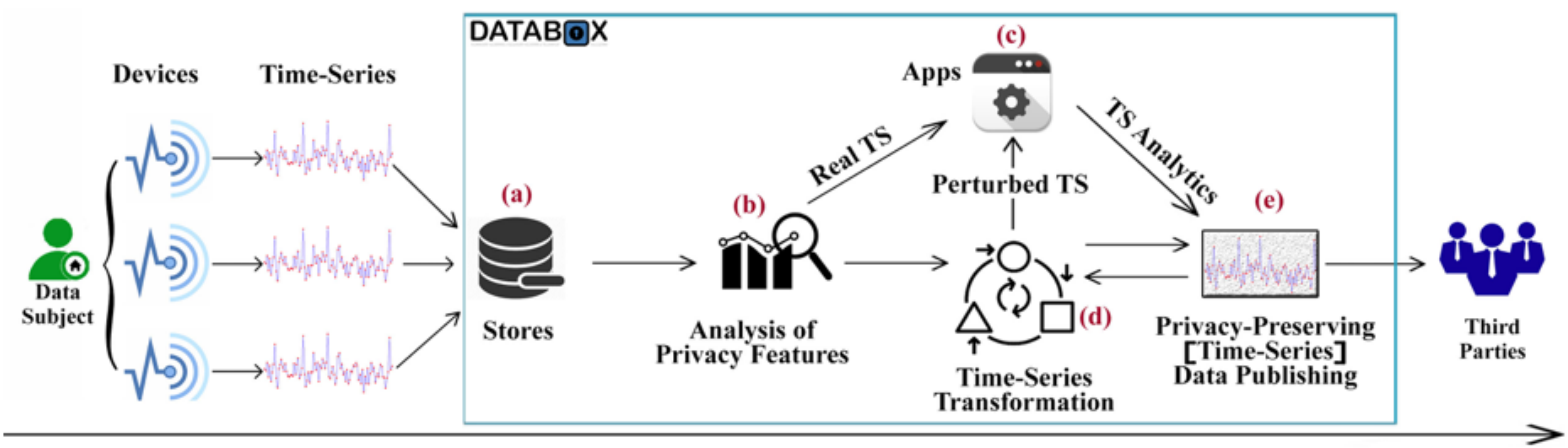


D

Approach

* Awareness of threats on collected time-series that can jeopardise user privacy, will help users to choose whether to provide apps their raw data or applying some transformation before granting access to them.

* When the processing of users' data is completed, users can allow their apps to send the results back to their providers via a privacy-preserving data publishing method.

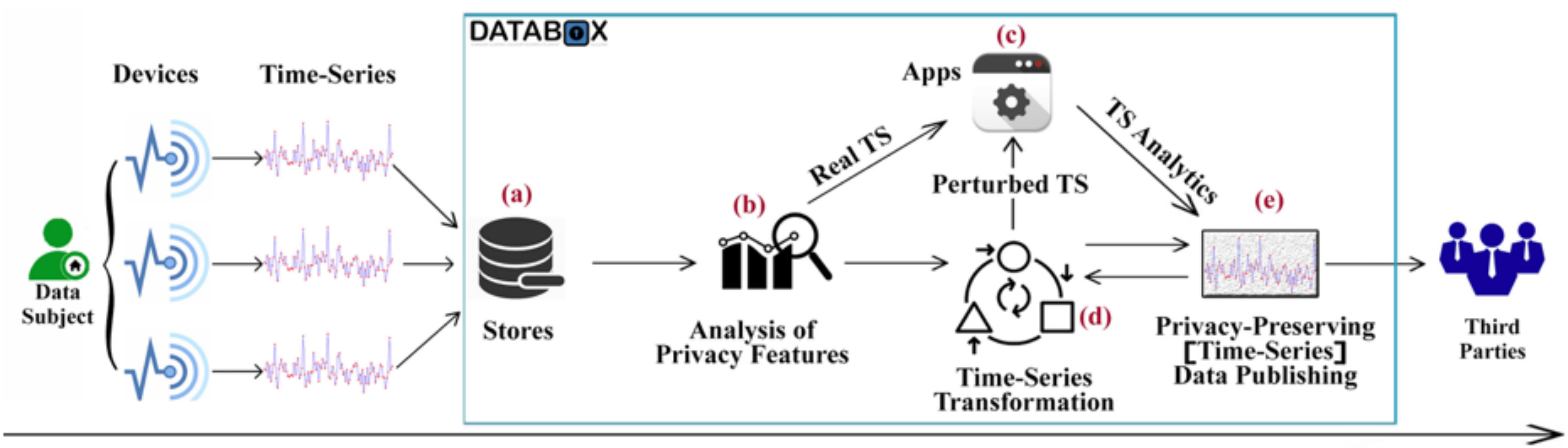


D

Goals

A. Extraction of the meaningful **statistics and characteristics** of time series which are **pertinent to user privacy**.

B. Transformation of the time-series to prevent third parties from **inferring user-specified sensitive information** while being able to accurately **achieve the agreed information**.



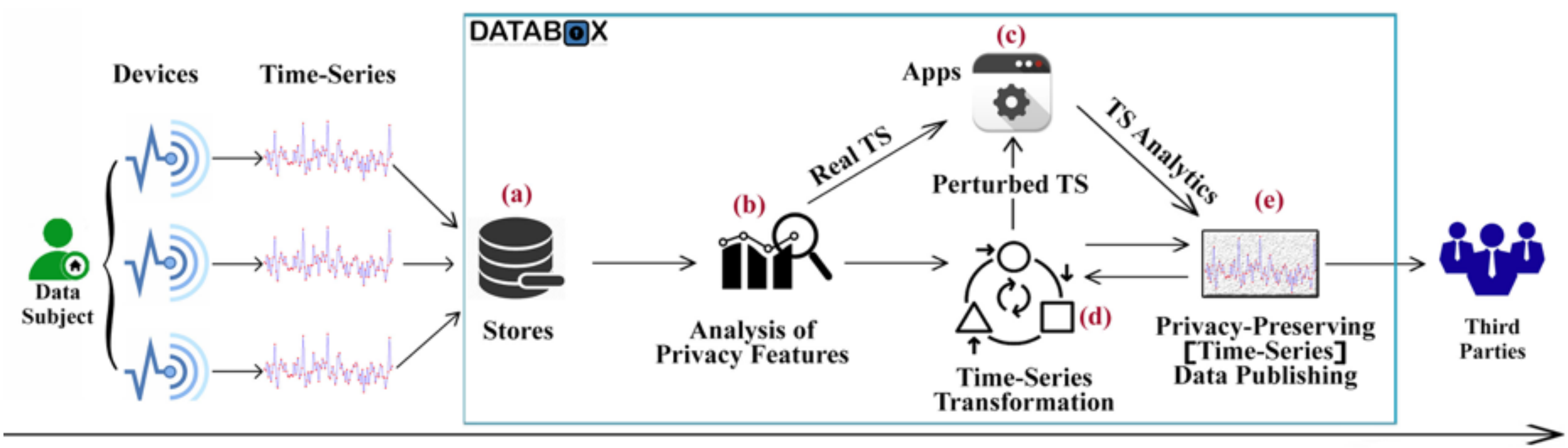
D

Future Direction

1) Suitable **measures** for evaluation of privacy-preserving IoT data publishing should be defined.

2) Real-world time-series **data** from IoT sensors (for a few specific sensor types) should be collected through development of specific apps for Databox.

3) We will consider **privacy leakage** in time-series data produced by multiple sensors for specific applications.



References

[1] Spensky, Chad, et al. "SoK: Privacy on Mobile Devices–It's Complicated." *Proceedings on Privacy Enhancing Technologies* 2016.3 (2016): 96-116.

[2] Laforet, Fabian, et al. "Individual privacy constraints on time-series data." Information Systems 54 (2015): 74-91.

[3] Saleheen, Nazir, et al. "mSieve: differential behavioral privacy in time series of mobile sensor data." *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2016.

[4] Chakraborty, Supriyo, et al. "Balancing behavioral privacy and information utility in sensory data flows." Pervasive and Mobile Computing 8.3 (2012): 331-345.

[5] Bindschaedler, Vincent, et al. "Synthesizing plausible privacy-preserving location traces." *Security and Privacy (SP), 2016 IEEE Symposium on*. IEEE, 2016.

[6] Haddadi, Hamed, et al. "Personal data: Thinking inside the box." arXiv preprint arXiv:1501.04737 (2015). Proceedings of The Fifth Decennial Aarhus Conference on Critical Alternatives. Aarhus University Press, 2015.

Thanks for your attention