Data Protection Case Study II

Carmela Troncoso

2nd March 2018

https://spring.epfl.ch/

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Problem Statement

ACME Transports would to optimize their business. Why not crowdsourcing? Millions of people looking at your problem. So they decide to publish a number of trajectories followed by their tracks over time

Q1: How much information is in this raw data? How difficult is it to get it?

Q2: Can the data be obfuscated?

Q3: And what about publishing aggregates?

Q1: How much information is in this raw data? How difficult is it to get it?

Q1: How much information is in this raw data? How difficult is it to get it?

- Time of trips
- Routes = customers
- Address of drivers?

In raw data this data can easily be inferred even with naïve algorithms

Q2: Can the data be obfuscated?

Q2: Can the data be obfuscated?

- Obfuscation: modification, suppression, generalization
- From the information available to the adversary
  - Define property to be protected: "not possible to identify customer"
  - Probabilistic analysis
- Hide drivers addresses ~ Anonymization?
  - Three properties of Art 29.
    - Differential Inference Testing A Practical Approach to Evaluate Anonymized Data Ali Kassem, Gergely Acs, Claude Castelluccia

Q2: Can the data be obfuscated?

- Obfuscation: modification, suppression, generalization
- From the information available to the adversary
  - Define property to be protected: "not possible to identify customer"
  - Probabilistic analysis
- Hide drivers addresses (~ Anonymization)?
  - Three properties of Art 29.
    - Differential Inference Testing A Practical Approach to Evaluate Anonymized Data Ali Kassem, Gergely Acs, Claude Castelluccia

Q3: And what about publishing aggregates? Where should ACME store these aggregates?

Q3: And what about publishing aggregates?

- Location Aggregates enable:
  - Inference of further data (3rd property of Art 29)
  - Membership attacks (1st property – linkability)

- Differential privacy is of little help (correlation and sensitivity)

No hope for ACME?

- Offer an API
  - Studying leakage of query
- Ad-hoc algorithms for particular statistics
  - Beware of subsequent releases
  - No crowdsourcing anymore, predefined possibility
- Synthetic data
  - Infancy

No hope for ACME?

But everything depends on the adversary!!

- Offer an API
  - Studying leakage of query
- Ad-hoc algorithms for particular statistics
  - Beware of subsequent releases
  - No crowdsourcing anymore, predefined possibility
- Synthetic data
  - Infancy

No hope for ACME?

But everything depends on the adversary!!
(see work by Hubaux/Ford for data sharing)

- Offer an API
  - Studying leakage of query
- Ad-hoc algorithms for particular statistics
  - Beware of subsequent releases
  - No crowdsourcing anymore, predefined possibility
- Synthetic data
  - Infancy