Privacy-preserving Smart Cities
¿utopia or reality?

Carmela Troncoso
(IMDEA Software Institute)
Smart City Expo World Congress
18\textsuperscript{th} November 2015
(Big) data

Data processing & analytics
(Big) data

Data processing & analytics
Data processing & analytics

(Big) Personal data

✔ Whereabouts
✔ Shopping
✔ Religion
✔ Restaurants
✔ Relationships
✔ Friends
✔ Professional
✔ Habits at home
✔ ...

...
(Big) Personal data

Data processing & analytics
1) Ethical issues, public opinion
2) Legal framework - Data Protection:
   - consent
   - proportionality
   - purpose limitation

(Big) Personal data

Data processing & analytics
1) Ethical issues, public opinion
2) Legal framework - Data Protection:
   consent
   proportionality
   purpose limitation

(Big) Personal data

Data processing & analytics

Value or privacy?
Two technological paths to reconciliation

- Data anonymization
- Advanced cryptography (processing in the encrypted domain)
Anonymization

EU legislation evolves to harder constraints [Art. 29 WP’s opinion on anonymization techniques]

3 criteria for anonymization

1- No singling out of individuals
Metadata are unique!

- **Location:**
  - “the median size of the individual’s anonymity set in the U.S. working population is 1, 21 and 34,980, for granularity of a census block, census track and county”
  - “if the location of an individual is specified hourly, and with a spatial resolution equal to that given by the carrier’s antennas, four spatio-temporal points are enough to uniquely identify 95% of the individuals.” [15 months, 1.5M people]”

- **Browser:** “83,6 % of browsers have unique fingerprints”

- **Demographic:** “It was found that 87 % (216 million of 248 million) of the population in the United States had reported characteristics that likely made them unique based only on {5-digit ZIP, gender, date of birth}”

- **Credit card transactions:** “need four purchases to identify an individual on the anonymized credit card records, or three purchases if the prices are known” [3 months 1.1 million people]
EU legislation evolves to harder constraints Art. 29 WP’s opinion on anonymization techniques

3 criteria for anonymization

1- No singling out of individuals
   Metadata are unique!

   - Location:
     ● “the median size of the individual’s anonymity set in the U.S. working population is 1, 21 and 34,980, for granularity of a census block, census track and county.”
     ● “if the location of an individual is specified hourly, and with a spatial resolution equal to that given by the carrier’s antennas, four spatio-temporal points are enough to uniquely identify 95% of the individuals.” [15 months, 1.5M people]

   - Browser:
     “83.6% of browsers have unique fingerprints”

   - Demographic:
     “It was found that 87% (216 million of 248 million) of the population in the United States had reported characteristics that likely made them unique based only on {5-digit ZIP, gender, date of birth}”

   - Credit card transactions:
     “need four purchases to identify an individual on the anonymized credit card records, or three purchases if the prices are known” [3 months 1.1 million people]

2- No linking data from one individual

   - Social network data: take two graphs representing social networks and map the nodes to each other based on the graph structure alone—no usernames, no nothing (Netflix Prize, Kaggle contest)
     • Techniques to do this automatically
EU legislation evolves to harder constraints [Art. 29 WP’s opinion on anonymization techniques]

3 criteria for anonymization

1- No singling out of individuals
   Metadata are unique!
   - Location:
     "the median size of the individual’s anonymity set in the U.S. working population is 1, 21 and 34,980, for granularity of a census block, census track and county"
     "if the location of an individual is specified hourly, and with a spatial resolution equal to that given by the carrier's antennas, four spatio-temporal points are enough to uniquely identify 95% of the individuals. [15 months, 1.5M people]"
   - Browser: "83.6% of browsers have unique fingerprints"
   - Demographic: "It was found that 87% (216 million of 248 million) of the population in the United States had reported characteristics that likely made them unique based only on {5-digit ZIP, gender, date of birth}"
   - Credit card transactions: "need four purchases to identify an individual on the anonymized credit card records, or three purchases if the prices are known [3 months, 1.1 million people]"

2- No linking data from one individual
   - Social network data:
     take two graphs representing social networks and map the nodes to each other based on the graph structure alone—no usernames, no nothing (Netflix Prize, Kaggle contest)
     Techniques to do this automatically

3- No inference about individuals
   - Location: infer workplace, home, religion, ...
   - Energy: infer concrete appliances, home habits ...

What is data analytics about?
EU legislation evolves to harder constraints Art. 29 WP’s opinion on anonymization techniques

3 criteria for anonymization

1- No singling out of individuals
   Metadata are unique!
   - Location:
     "the median size of the individual’s anonymity set in the U.S. working population is 1, 21 and 34,980, for granularity of a census block, census track and county"
     "if the location of an individual is specified hourly, and with a spatial resolution equal to that given by the carrier’s antennas, four spatio-temporal points are enough to uniquely identify 95% of the individuals." [15 months, 1.5M people] 
   - Browser: "83.6% of browsers have unique fingerprints" 
   - Demographic: "It was found that 87% (216 million of 248 million) of the population in the United States had reported characteristics that likely made them unique based only on {5-digit ZIP, gender, date of birth}"
   - Credit: "need four purchases to identify an individual on the anonymized credit card records, or three purchases if the prices are known" [3 months 1.1 million people]

2- No linking data
   - Social network data: "take two graphs representing social networks and map the nodes to each other based on the graph structure alone — no usernames, no nothing (Netflix Prize, Kaggle contest)"
   - Techniques to do this automatically

3- No inference about individuals
   - Location: infer workplace, home, religion, ... 
   - Energy: infer concrete appliances, home habits ...

What is data analytics about?
EU legislation evolves to harder constraints Art. 29 WP’s opinion on anonymization techniques

3 criteria for anonymization

1- No singling out of individuals

- Location:
  - “the median size of the individual’s anonymity set in the U.S. working population is 1, 21 and 34,980”
  - “if the location of an individual is specified hourly, and with a spatial resolution equal to that given by the carrier’s antennas, four spatio-temporal points are enough to uniquely identify 95% of the individuals.” [15 months, 1.5M people]

- Browser: “83.6% of browsers have unique fingerprints”

- Demographic: “It was found that 87% (216 million of 248 million) of the population in the United States had reported characteristics that likely made them unique based only on {5-digit ZIP, gender, date of birth}”

- Credit card transactions: “need four purchases to identify an individual on the anonymized credit card records, or three purchases if the prices are known” [3 months 1.1 million people]

2- No linking data from one individual

- Social network data: take two graphs representing social networks and map the nodes to each other based on the graph structure alone — no usernames, no nothing (Netflix Prize, Kaggle contest)

3- No inference about individuals

- Location: infer workplace, home, religion,...

- Energy: infer concrete appliances, home habits...

Art 29 - Risk of de-anonymization

- Traditional identification suppression methods will not do the trick (hash, encryption, random noise...)

- But...
  - We can evaluate anonymity degree and remaining information
  - General anonymization ← little utility
  - Targeted (application dependent) anonymization ← better utility

IMPOSSIBLE IS NOTHING
Advanced cryptography
Processing in the encrypted domain
Advanced cryptography
Processing in the encrypted domain

41.373925, 2.149896
Fira de Barcelona
Advanced cryptography
Processing in the encrypted domain

41.373925, 2.149896
Advanced cryptography
Processing in the encrypted domain
Advanced cryptography
Processing in the encrypted domain

41.373925, 2.149896
Advanced cryptography
Processing in the encrypted domain

Fira de Barcelona

41.373925, 2.149896

\&_D#$^a\)

\&$%_XY
Advanced cryptography
Processing in the encrypted domain

Data encrypted at the user side (local key management)

Provider cannot read
But can process!!

Fira de Barcelona

41.373925, 2.149896
Advanced cryptography
Processing in the encrypted domain

Data encrypted at the user side
(local key management)

Provider cannot read
But can process!!

Best of both worlds: service AND privacy!
What “magic” is possible?
- Private searches
- Private billing
- Private comparison
- Private sharing
- Private statistics computation
Privacy-preserving Smart Cities
¿utopia or reality?

No personal data involved: is a reality!

Personal data: not yet guaranteed, but there is a path!
- Anonymization and privacy evaluation
- Advanced cryptography

Carmela Troncoso
Carmela.troncoso@imdea.org
www.software.imdea.org
Privacy-preserving Smart Cities
¿utopia or reality?

No personal data involved: is a reality!

Personal data: not yet guaranteed, but there is a path!
- Anonymization and privacy evaluation
- Advanced cryptography

We need to work together to walk this path!

Carmela Troncoso
Carmela.troncoso@imdea.org
www.software.imdea.org