Privacy Preserving Aggregation of Distributed Mobility Data Streams

Thomas Liebig
TU Dortmund
Mobility Data Aggregation Problem

- How many people have been at which location?
Re-Identification problems

- removing personal identifier unsufficient
- sending position information in bulks (trajectories or histograms) unsufficient
- Easy re-identification in mobility data
- Intrusted server

→ streaming solution required which prevents server from accessing individual mobility traces

Problem described in [Andrienko et. al. 12] (with many examples and references to laws)
Problem

☐ secure this aggregation:
Related approach [Kopp et. al. 12]
Re-Identification problems

- removing personal identifier unsufficient
- sending position information in bulks (trajectories or histograms) unsufficient
- Easy re-identification in mobility data
- Intrusted server

→ streaming solution required which prevents server from accessing individual mobility traces

Problem described in [Andrienko et. al. 12] (with many examples and references to laws)
Related approach [Monreale et. al. 13]
Re-Identification problems

- removing personal identifier unsufficient
- sending position information in bulks (trajectories or histograms) unsufficient
- Easy re-identification in mobility data
- Intrusted server

→ streaming solution required which prevents server from accessing individual mobility traces

Problem described in [Andrienko et. al. 12] (with many examples and references to laws)
Proposed Solution (3 steps)

- utilize homomorphic encryption
Proposed Solution I (RSA)

- based on two cryptographic keys (like in ssh)

\[ M := \text{Encryption} (H, \text{public key}) \rightarrow \text{Decryption} (M, \text{private key}) = H \]
Proposed Solution II (Damgaard Algorithm)

- sharing the keys (homomorphic cryptography)

\[
\begin{align*}
M1 &= \text{Encryption (H1, public key1)} \\
M2 &= \text{Encryption (H2, public key2)} \\
M3 &= \text{Encryption (H3, public key3)} \\
\text{Decryption (M1*M2*M3, private key)} &= H1 + H2 + H3
\end{align*}
\]
Proposed Solution III (Hash Chain)

- prevent timing attacks

Proposed Solution:

\[ M1 := \text{Encryption} (H1^*T, \text{public key}1) \]
\[ M2 := \text{Encryption} (H2^*T, \text{public key}2) \]
\[ M3 := \text{Encryption} (H3^*T, \text{public key}3) \]

\[ \text{Decryption} (M1^*M2^*M3, \text{private key}) = (H1+H2+H3)^*T \]
Summary

- privacy-by-design concept for mobility stream aggregation
- full utility
- no randomization, no approximation

Next steps:

- further algorithms (clustering, classification, subgroup discovery) without revealing the data
Privacy Preserving Aggregation of Distributed Mobility Data Streams

Thomas Liebig

TU Dortmund