Privacy-Friendly Mobility Analytics using Aggregate Location Data
ACM SIGSPATIAL 2016

Apostolos Pyrgelis, Emiliano De Cristofaro, Gordon Ross
University College London

November 3, 2016
Motivation

Mobility analytics are useful in modern cities - journey planning, congestion prevention, improving transportation service levels.

But, large scale collection of individual users' location data raises privacy concerns (life-style, political / religious inclinations).

Anonymization of location traces is ineffective.
Mobility analytics are useful in modern cities - journey planning, congestion prevention, improving transportation service levels.

But, large scale collection of individual users’ location data raises privacy concerns (life-style, political / religious inclinations).

Anonymization of location traces is ineffective.

A. Pyrgelis, E. De Cristofaro, G. Ross, UCL
Privacy-Friendly Mobility Analytics using Aggregate Locations
Motivation

- Mobility analytics are useful in modern cities - journey planning, congestion prevention, improving transportation service levels.
- But, large scale collection of individual users’ location data raises privacy concerns (life-style, political / religious inclinations).

A. Pyrgelis, E. De Cristofaro, G. Ross, UCL
Privacy-Friendly Mobility Analytics using Aggregate Locations
Mobility analytics are useful in modern cities - journey planning, congestion prevention, improving transportation service levels

But, large scale collection of individual users’ location data raises privacy concerns (life-style, political / religious inclinations)

Anonymization of location traces is ineffective
Our Proposal

Our approach:
data aggregation
for gathering location
statistics

Our goals:
1. usefulness of aggregate locations for mobility analytics
2. real-world deployability of a system for privacy-friendly location
data collection via crowd-sourcing

A. Pyrgelis, E. De Cristofaro, G. Ross, UCL
Privacy-Friendly Mobility Analytics using Aggregate Locations
Our Proposal

Our approach: *data aggregation* for gathering location statistics.
Our Proposal

Our approach: *data aggregation* for gathering location statistics

Our goals:

1. Usefulness of aggregate locations for mobility analytics
2. Real-world deployability of a system for privacy-friendly location data collection via crowd-sourcing
Our Proposal

Our approach: *data aggregation* for gathering location statistics

Our goals:

1. usefulness of aggregate locations for mobility analytics
Our Proposal

- **Our approach:** *data aggregation* for gathering location statistics
- **Our goals:**
  1. usefulness of aggregate locations for mobility analytics
  2. real-world deployability of a system for privacy-friendly location data collection via crowd-sourcing
Roadmap

- Experiment with real-world mobility datasets (TFL, SFC)
- Methodology for performing mobility analytics over aggregate locations
- 1. Forecasting traffic volumes in regions of interest (ROIs)
- 2. Detecting mobility anomalies
- 3. Improving traffic volume predictions in the presence of anomalies
- Design a privacy-respecting system for crowd-sourcing location data
- Empirical evaluation of computation / communication / energy complexities
Roadmap

- Experiment with real-world mobility datasets (TFL, SFC)
Roadmap

- Experiment with real-world mobility datasets (TFL, SFC)
- Methodology for performing mobility analytics over aggregate locations
  1. forecasting traffic volumes in regions of interest (ROIs)
  2. detecting mobility anomalies
  3. improving traffic volume predictions in the presence of anomalies

A. Pyrgelis, E. De Cristofaro, G. Ross, UCL
Privacy-Friendly Mobility Analytics using Aggregate Locations
Roadmap

- Experiment with real-world mobility datasets (TFL, SFC)
- Methodology for performing mobility analytics over aggregate locations
  1. forecasting traffic volumes in regions of interest (ROIs)
  2. detecting mobility anomalies
  3. improving traffic volume predictions in the presence of anomalies
- Design a privacy-respecting system for crowd-sourcing location data
- Empirical evaluation of computation / communication / energy complexities
Transport for London (TFL)

Logs of anonymized oyster card trips including Underground (LUL), National Rail (NR), Overground (LRC), Docklands Light Railway (DLR)

Monday, March 1 to Sunday, March 28, 2010 (4 weeks)

60 million trips as performed by 4 million unique users, over 582 stations

We build hourly time series (TS) of stations ($Y_t$), counting # of users tapping-in/out at each station

A. Pyrgelis, E. De Cristofaro, G. Ross, UCL

Privacy-Friendly Mobility Analytics using Aggregate Locations
Transport for London (TFL)

- Logs of anonymized oyster card trips including Underground (LUL), National Rail (NR), Overground (LRC), Docklands Light Railway (DLR)

A. Pyrgelis, E. De Cristofaro, G. Ross, UCL
Logs of anonymized oyster card trips including Underground (LUL), National Rail (NR), Overground (LRC), Docklands Light Railway (DLR)

Monday, March 1 to Sunday, March 28, 2010 (4 weeks)
Transport for London (TFL)

- Logs of anonymized oyster card trips including Underground (LUL), National Rail (NR), Overground (LRC), Docklands Light Railway (DLR)
- Monday, March 1 to Sunday, March 28, 2010 (4 weeks)
- 60 million trips as performed by 4 million unique users, over 582 stations
Transport for London (TFL)

- Logs of anonymized oyster card trips including Underground (LUL), National Rail (NR), Overground (LRC), Docklands Light Railway (DLR)
- Monday, March 1 to Sunday, March 28, 2010 (4 weeks)
- 60 million trips as performed by 4 million unique users, over 582 stations
- We build hourly time series (TS) of stations ($Y_t$), counting # of users tapping-in/out at each station
<table>
<thead>
<tr>
<th>Date</th>
<th>Canary Wharf</th>
<th>Clapham Common</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mar-03</td>
<td>0</td>
<td>500</td>
</tr>
<tr>
<td>Mar-06</td>
<td>2000</td>
<td>2500</td>
</tr>
<tr>
<td>Mar-09</td>
<td>6000</td>
<td>3000</td>
</tr>
<tr>
<td>Mar-12</td>
<td>8000</td>
<td></td>
</tr>
<tr>
<td>Mar-15</td>
<td>10000</td>
<td></td>
</tr>
<tr>
<td>Mar-18</td>
<td>12000</td>
<td></td>
</tr>
<tr>
<td>Mar-21</td>
<td>14000</td>
<td></td>
</tr>
<tr>
<td>Mar-24</td>
<td>16000</td>
<td></td>
</tr>
<tr>
<td>Mar-27</td>
<td>18000</td>
<td></td>
</tr>
</tbody>
</table>

A. Pyrgelis, E. De Cristofaro, G. Ross, UCL

Privacy-Friendly Mobility Analytics using Aggregate Locations
Canary Wharf TS.

A. Pyrgelis, E. De Cristofaro, G. Ross, UCL
We observe daily / weekly seasonality and stationarity.

Canary Wharf TS.

Clapham Common TS.
We observe daily / weekly **seasonality** and stationarity
San Francisco Cab Network (SFC)

Mobility traces of 536 cabs in San Francisco between May 19 to June 8, 2008 (3 weeks)

11 million GPS coordinates

San Francisco grid of 100 x 100 regions, each of 0.19 × 0.14 sq mi

We build hourly time series (TS) for ROIs ($Y_t$), counting # of taxis that have reported presence

A. Pyrgelis, E. De Cristofaro, G. Ross, UCL
Privacy-Friendly Mobility Analytics using Aggregate Locations
Mobility traces of 536 cabs in San Francisco between May 19 to June 8, 2008 (3 weeks)
Mobility traces of 536 cabs in San Francisco between May 19 to June 8, 2008 (3 weeks)
11 million GPS coordinates
San Francisco Cab Network (SFC)

- Mobility traces of 536 cabs in San Francisco between May 19 to June 8, 2008 (3 weeks)
- 11 million GPS coordinates
- San Francisco grid of 100 × 100 regions, each of 0.19 × 0.14 sq mi
Mobility traces of 536 cabs in San Francisco between May 19 to June 8, 2008 (3 weeks)

11 million GPS coordinates

San Francisco grid of $100 \times 100$ regions, each of $0.19 \times 0.14$ sq mi

We build hourly time series (TS) for ROIs ($Y_t$), counting # of taxis that have reported presence
- Mobility traces of 536 cabs in San Francisco between May 19 to June 8, 2008 (3 weeks)
- 11 million GPS coordinates
- San Francisco grid of 100 x 100 regions, each of 0.19 x 0.14 sq mi
- We build hourly time series (TS) for ROIs ($Y_t$), counting # of taxis that have reported presence

SFC 100 x 100 grid.
We observe daily, weekly patterns and stationarity.
Region 7160 TS.
We observe daily, weekly patterns and stationarity

A. Pyrgelis, E. De Cristofaro, G. Ross, UCL
Privacy-Friendly Mobility Analytics using Aggregate Locations
We observe daily, weekly patterns and stationarity
Removing Seasonality

Additive decomposition of TS:

$$D_t = Y_t - Y_{t-1}$$


Date

0
1000
2000
3000
4000
5000
6000
7000
8000

# of Oysters

Green Park


Date

4000
3000
2000
1000
0
1000
2000

# of Oysters

Green Park Aggregate TS.

De-seasonalized time series ($D_t$) show strong auto-regressive structure.
Removing Seasonality

Additive decomposition of TS: $D_t = Y_t - \overline{Y}_t$
Removing Seasonality

Additive decomposition of TS: \( D_t = Y_t - \overline{Y_t} \)

Green Park Aggregate TS.
Removing Seasonality

Additive decomposition of TS: \( D_t = Y_t - \overline{Y_t} \)

Green Park Aggregate TS.  
Green Park De-seasonalized TS.
Removing Seasonality

Additive decomposition of TS: \( D_t = Y_t - \overline{Y_t} \)

De-seasonalized time series \( (D_t) \) show strong **auto-regressive** structure
Forecasting Traffic Volumes in ROIs

\[ \hat{Y}_t = \hat{D}_t + Y_t, \] using a sliding window

Evaluate accuracy via absolute forecast error ($e_t = |Y_t - \hat{Y}_t|$)
ARMA_{SEAS} modeling
**Forecasting Traffic Volumes in ROIs**

- ARMA\textsubscript{SEAS} modeling
- \( \hat{Y}_t = \hat{D}_t + \bar{Y}_t \), using a sliding window

Evaluate accuracy via absolute forecast error (\( e_t = |Y_t - \hat{Y}_t| \))
Forecasting Traffic Volumes in ROIs

- ARMA_{SEAS} modeling

\[ \hat{Y}_t = \hat{D}_t + \overline{Y}_t \], using a sliding window

- Evaluate accuracy via absolute forecast error (\( e_t = |Y_t - \overline{Y}_t| \))
Experiments
Forecasting Traffic Volumes in ROIs

Comparison to a baseline black-box ARMA model on $Y_t$.
Improved predictions when considering seasonal effects (e.g. TFL average error: 19% vs 600%).

A. Pyrgelis, E. De Cristofaro, G. Ross, UCL
Privacy-Friendly Mobility Analytics using Aggregate Locations
Experiments
Forecasting Traffic Volumes in ROIs

- Experiment with *top* 100 TFL stations and SFC ROIs
Experiments
Forecasting Traffic Volumes in ROIs

- Experiment with top 100 TFL stations and SFC ROIs
- 5 days of data for training ($D_t$) - 1 day of testing ($\widehat{Y_t}$ vs $Y_t$)
Experiments
Forecasting Traffic Volumes in ROIs

- Experiment with top 100 TFL stations and SFC ROIs
- 5 days of data for training ($D_t$) - 1 day of testing ($\hat{Y}_t$ vs $Y_t$)

![Graph](Green Park Predictions, March 25.)

Comparison to a baseline black-box ARMA model on $Y_t$. Improved predictions when considering seasonal effects (e.g. TFL average error : 19% vs 600%).
Experiments
Forecasting Traffic Volumes in ROIs

- Experiment with *top* 100 TFL stations and SFC ROIs
- 5 days of data for training ($D_t$) - 1 day of testing ($\hat{Y}_t$ vs $Y_t$)

Green Park Predictions, March 25.
Region 8755 Predictions, June 5.
Experiments
Forecasting Traffic Volumes in ROIs

- Experiment with top 100 TFL stations and SFC ROIs
- 5 days of data for training ($D_t$) - 1 day of testing ($\hat{Y}_t$ vs $Y_t$)

Green Park Predictions, March 25.
- Comparison to a baseline black-box ARMA model on $Y_t$

Region 8755 Predictions, June 5.

A. Pyrgelis, E. De Cristofaro, G. Ross, UCL
Privacy-Friendly Mobility Analytics using Aggregate Locations
Experiments
Forecasting Traffic Volumes in ROIs

- Experiment with top 100 TFL stations and SFC ROIs
- 5 days of data for training ($D_t$) - 1 day of testing ($\hat{Y}_t$ vs $Y_t$)

Green Park Predictions, March 25.
- Comparison to a baseline black-box ARMA model on $Y_t$
- Improved predictions when considering seasonal effects (e.g. TFL average error: 19% vs 600%)

Region 8755 Predictions, June 5.
Detecting Traffic Anomalies in ROIs

ARMA SEAS modeling - rely on absolute forecast error ($e_t$)

Apply the $3\sigma$ rule, with confidence interval:

$$\lambda = \mu + 3\sigma$$

Detect an anomaly at time $t$ if:

$$e_t > \lambda$$

A. Pyrgelis, E. De Cristofaro, G. Ross, UCL

Privacy-Friendly Mobility Analytics using Aggregate Locations
ARMA_{SEAS} modeling - rely on absolute forecast error ($e_t$)
ARMA\textsubscript{SEAS} modeling - rely on absolute forecast error \( (e_t) \)

Apply the 3\( \sigma \) rule, with confidence interval: \( \lambda = \mu + 3\sigma \)
ARMA\textsubscript{SEAS} modeling - rely on absolute forecast error ($e_t$)

- Apply the $3\sigma$ rule, with confidence interval: $\lambda = \mu + 3\sigma$

- Detect an anomaly at time $t$ if: $e_t > \lambda$
Experiments
Detecting Traffic Anomalies in ROIs

Train the ARMA SEAS model with 1 week data, test it against the rest of weeks.

Top 100 TFL stations: 896 anomalies
Top 100 SFC blocks: 366 anomalies

North Greenwich, March 10.

Region 8261, May 31.

Note: no ground truth for anomalies

A. Pyrgelis, E. De Cristofaro, G. Ross, UCL
Privacy-Friendly Mobility Analytics using Aggregate Locations
Experiments
Detecting Traffic Anomalies in ROIs

- Train the ARMA_{SEAS} model with 1 week data, test it against the rest of weeks
Experiments
Detecting Traffic Anomalies in ROIs

- Train the ARMA_{SEAS} model with 1 week data, test it against the rest of weeks
- Top 100 TFL stations: 896 anomalies
Experiments
Detecting Traffic Anomalies in ROIs

- Train the ARMA$_{SEAS}$ model with 1 week data, test it against the rest of weeks
- Top 100 TFL stations : 896 anomalies
- Top 100 SFC blocks: 366 anomalies
Experiments
Detecting Traffic Anomalies in ROIs

- Train the ARMA\textsubscript{SEAS} model with 1 week data, test it against the rest of weeks
- Top 100 TFL stations: 896 anomalies
- Top 100 SFC blocks: 366 anomalies

North Greenwich, March 10.
Experiments
Detecting Traffic Anomalies in ROIs

- Train the ARMA$_{SEAS}$ model with 1 week data, test it against the rest of weeks
- Top 100 TFL stations: 896 anomalies
- Top 100 SFC blocks: 366 anomalies

North Greenwich, March 10.

Region 8261, May 31.

A. Pyrgelis, E. De Cristofaro, G. Ross, UCL
Experiments
Detecting Traffic Anomalies in ROIs

- Train the ARMA_{SEAS} model with 1 week data, test it against the rest of weeks
- Top 100 TFL stations: 896 anomalies
- Top 100 SFC blocks: 366 anomalies

North Greenwich, March 10.

Region 8261, May 31.

Note: no ground truth for anomalies
Predicting Traffic Volumes during Anomalies

Can we improve our predictions in the presence of an anomaly?

Discover correlated ROIs by sliding their time series - (Spearman correlation)

Use a VAR model to capture linear inter-dependencies

A. Pyrgelis, E. De Cristofaro, G. Ross, UCL

Privacy-Friendly Mobility Analytics using Aggregate Locations
Can we improve our predictions in the presence of an anomaly?
Can we improve our predictions in the presence of an anomaly?

Discover *correlated* ROIs by sliding their time series - (Spearman correlation)
Can we improve our predictions in the presence of an anomaly?

Discover correlated ROIs by sliding their time series - (Spearman correlation)

Use a VAR model to capture linear inter-dependencies between time series
Experiments
Predicting Traffic Volumes during Anomalies

Experiment with 10% of anomalies of TFL (90 anoms) and SFC (30 anoms)
Train a VAR model including information from 10 correlated ROIs
Compare against a baseline: ARMA SEAS model trained on local data

Arsenal Exit Traffic Predictions during an Anomaly, March 20.
ARMA SEAS Error 93% vs VAR Error 59%

A. Pyrgelis, E. De Cristofaro, G. Ross, UCL
Privacy-Friendly Mobility Analytics using Aggregate Locations
Experiments
Predicting Traffic Volumes during Anomalies

- Experiment with 10% of anomalies of TFL (90 anoms) and SFC (30 anoms)
Experiments
Predicting Traffic Volumes during Anomalies

- Experiment with 10% of anomalies of TFL (90 anoms) and SFC (30 anoms)
- Train a VAR model including information from 10 correlated ROIs
Experiments
Predicting Traffic Volumes during Anomalies

- Experiment with 10% of anomalies of TFL (90 anoms) and SFC (30 anoms)
- Train a VAR model including information from 10 correlated ROIs
- Compare against a baseline: ARMA_{SEAS} model trained on local data
Experiments
Predicting Traffic Volumes during Anomalies

- Experiment with 10% of anomalies of TFL (90 anoms) and SFC (30 anoms)
- Train a VAR model including information from 10 correlated ROIs
- Compare against a baseline: ARMA_{SEAS} model trained on local data

Arsenal Exit Traffic Predictions during an Anomaly, March 20.
- ARMA_{SEAS} Error 93% vs VAR Error 59%
Experiments
Predicting Traffic Volumes during Anomalies

Overall, significant improvement in predictions when considering information from correlated ROIs

TFL: 29% improvement in predictions
SFC: 18% improvement in predictions

A. Pyrgelis, E. De Cristofaro, G. Ross, UCL
Privacy-Friendly Mobility Analytics using Aggregate Locations
Overall, significant improvement in predictions when considering information from correlated ROIs
Overall, significant improvement in predictions when considering information from correlated ROIs

TFL: 29% improvement in predictions
Overall, significant improvement in predictions when considering information from correlated ROIs

- TFL: 29% improvement in predictions
- SFC: 18% improvement in predictions
What next?

Analytics on aggregate locations offer interesting insights.

Can we collect aggregate locations directly from users, with privacy?

Challenges: Efficiency, scalability, fault-tolerance.

Good news: promising results by Melis et al. (NDSS 2016).

A. Pyrgelis, E. De Cristofaro, G. Ross, UCL

Privacy-Friendly Mobility Analytics using Aggregate Locations
What next?

- Analytics on aggregate locations offer interesting insights
What next?

Analytics on aggregate locations offer interesting insights.
Can we collect aggregate locations directly from users, with privacy?
What next?

- Analytics on aggregate locations offer interesting insights
- Can we collect aggregate locations directly from users, with privacy?
- Challenges: Efficiency, scalability, fault-tolerance
What next?

- Analytics on aggregate locations offer interesting insights
- Can we collect aggregate locations directly from users, with **privacy**?
- Challenges: Efficiency, scalability, fault-tolerance
- Good news: promising results by Melis et al. (NDSS 2016)\(^1\)

---

\(^1\)Luca Melis, George Danezis, and Emiliano De Cristofaro: Efficient private statistics with succinct sketches, NDSS (2016).
Mobility Data Donors (MDD) Framework

Design a collaborative framework for aggregate location data collection (users vs aggregator)

**Client Side:**
- Users install MDD app
- MDD runs on the background, collecting GPS coordinates
- Aggregator periodically triggers privacy-preserving aggregation, assigning users to groups
- MDD encrypts entries in the matrix that represents user locations

**Server Side:**
- Aggregator collects the encrypted matrices and decrypts ONLY aggregate location counts - combines aggregates if collected from multiple groups

A. Pyrgelis, E. De Cristofaro, G. Ross, UCL

Privacy-Friendly Mobility Analytics using Aggregate Locations
Mobility Data Donors (MDD) Framework

- Design a collaborative framework for aggregate location data collection (users vs aggregator)
Design a collaborative framework for aggregate location data collection (users vs aggregator)

Client Side:
Mobility Data Donors (MDD) Framework

- Design a collaborative framework for aggregate location data collection (users vs aggregator)

  **Client Side:**
  - Users install MDD app
Design a collaborative framework for aggregate location data collection (users vs aggregator)

**Client Side:**
- Users install MDD app
- MDD runs on the background, collecting GPS coordinates
Mobility Data Donors (MDD) Framework

- Design a collaborative framework for aggregate location data collection (users vs aggregator)

- **Client Side:**
  - Users install MDD app
  - MDD runs on the background, collecting GPS coordinates
  - Aggregator periodically triggers privacy-preserving aggregation, assigning users to groups

---

A. Pyrgelis, E. De Cristofaro, G. Ross, UCL
Design a collaborative framework for aggregate location data collection (users vs aggregator)

**Client Side:**
- Users install MDD app
- MDD runs on the background, collecting GPS coordinates
- Aggregator periodically triggers privacy-preserving aggregation, assigning users to groups
- MDD encrypts entries in the matrix that represents user locations
Design a collaborative framework for aggregate location data collection (users vs aggregator)

**Client Side:**
- Users install MDD app
- MDD runs on the background, collecting GPS coordinates
- Aggregator periodically triggers privacy-preserving aggregation, assigning users to groups
- MDD encrypts entries in the matrix that represents user locations

**Server Side:**
Design a collaborative framework for aggregate location data collection (users vs aggregator)

**Client Side:**
- Users install MDD app
- MDD runs on the background, collecting GPS coordinates
- Aggregator periodically triggers privacy-preserving aggregation, assigning users to groups
- MDD encrypts entries in the matrix that represents user locations

**Server side:**
- Aggregator collects the encrypted matrices and decrypts **ONLY** aggregate location counts - combines aggregates if collected from multiple groups
MDD Experimental Evaluation

Javascript/Node.js implementation of the secure aggregation protocol by Melis et al.

Port of client side to run on Android, via Apache Cordova

Cryptographic operations: Edc25519 elliptic curve - 128 bit security

Android device: Samsung Galaxy A3, 1.2 GHz quad-core Snapdragon 410, 1.5 GB RAM, Lollipop v5.0.2

PowerTutor app for power monitoring

A. Pyrgelis, E. De Cristofaro, G. Ross, UCL

Privacy-Friendly Mobility Analytics using Aggregate Locations
• Javascript/Node.js implementation of the secure aggregation protocol by Melis et al.
• Port of client side to run on Android, via Apache Cordova
Javascript/Node.js implementation of the secure aggregation protocol by Melis et al.

Port of client side to run on Android, via Apache Cordova

Cryptographic operations: Edc25519 elliptic curve - 128 bit security
- Javascript/Node.js implementation of the secure aggregation protocol by Melis et al.
- Port of client side to run on Android, via Apache Cordova
- Cryptographic operations: Edc25519 elliptic curve - 128 bit security
- Android device: Samsung Galaxy A3, 1.2 GHz quad-core Snapdragon 410, 1.5 GB RAM, Lollipop v5.0.2
Javascript/Node.js implementation of the secure aggregation protocol by Melis et al.

Port of client side to run on Android, via Apache Cordova

Cryptographic operations: Edc25519 elliptic curve - 128 bit security

Android device: Samsung Galaxy A3, 1.2 GHz quad-core Snapdragon 410, 1.5 GB RAM, Lollipop v5.0.2

PowerTutor app for power monitoring
<table>
<thead>
<tr>
<th># of Mobile Users</th>
<th>Execution Time (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>Encryption</td>
</tr>
<tr>
<td>200</td>
<td>10.7KB public keys, 4.54KB encrypted ROI matrix</td>
</tr>
<tr>
<td>300</td>
<td>826mJ encryption, 609mJ / 322mJ download / upload (via Wi-Fi)</td>
</tr>
</tbody>
</table>

Note: Succinct data representation can be used, if more fine-grained data need to be collected (e.g. O-D matrices)

A. Pyrgelis, E. De Cristofaro, G. Ross, UCL

Privacy-Friendly Mobility Analytics using Aggregate Locations
- ROI matrix of size $(582, 2)$
- $\sim 7s$ encryption for groups of 200 mobile users

**TFL Execution Time - Encryption Phase.**
Results - TFL

- ROI matrix of size (582, 2)
- ~ 7s encryption for groups of 200 mobile users

TFL Execution Time - Encryption Phase.

- 10.7KB public keys, 4.54KB encrypted ROI matrix
• ROI matrix of size (582, 2)
• \( \sim 7 \text{s} \) encryption for groups of 200 mobile users

10.7KB public keys, 4.54KB encrypted ROI matrix
826mJ encryption, 609mJ / 322mJ download / upload (via Wi-Fi)

TFL Execution Time - Encryption Phase.
- ROI matrix of size (582, 2)
- ~ 7s encryption for groups of 200 mobile users

TFL Execution Time - Encryption Phase.

- 10.7KB public keys, 4.54KB encrypted ROI matrix
- 826mJ encryption, 609mJ / 322mJ download / upload (via Wi-Fi)

**Note**: Succinct data representation can be used, if more fine grained data need to be collected (e.g. O-D matrices)
Conclusions

Mobility analytics over crowd-sourced aggregate location data

Time series modeling with seasonality for:
1. forecasting traffic volumes in ROIs
2. detecting anomalies
3. improving traffic volume predictions in the presence of anomalies

Experiments on real-world mobility datasets (TFL, SFC)

Privacy-respecting system for data collection

Mobile application framework (MDD) and empirical evaluation in terms of computation / communication / energy overhead

A. Pyrgelis, E. De Cristofaro, G. Ross, UCL

Privacy-Friendly Mobility Analytics using Aggregate Locations
Conclusions

- Mobility analytics over crowd-sourced aggregate location data
Conclusions

- Mobility analytics over *crowd-sourced* aggregate location data
- Time series modeling with seasonality for:
  1. forecasting traffic volumes in ROIs
  2. detecting anomalies
  3. improving traffic volume predictions in the presence of anomalies

Experiments on real-world mobility datasets (TFL, SFC)

Privacy-respecting system for data collection

Mobile application framework (MDD) and empirical evaluation in terms of computation / communication / energy overhead

A. Pyrgelis, E. De Cristofaro, G. Ross, UCL

Privacy-Friendly Mobility Analytics using Aggregate Locations
Conclusions

- Mobility analytics over *crowd-sourced* aggregate location data
- Time series modeling with seasonality for:
  1. forecasting traffic volumes in ROIs
  2. detecting anomalies
  3. improving traffic volume predictions in the presence of anomalies
- Experiments on real-world mobility datasets (TFL, SFC)
Conclusions

- Mobility analytics over *crowd-sourced* aggregate location data
- Time series modeling with seasonality for:
  1. forecasting traffic volumes in ROIs
  2. detecting anomalies
  3. improving traffic volume predictions in the presence of anomalies
- Experiments on real-world mobility datasets (TFL, SFC)
- Privacy-respecting system for data collection
- Mobile application framework (MDD) and empirical evaluation in terms of computation / communication / energy overhead
Future Work

Evaluate our methodology on different mobility datasets
Privacy quantification and analysis of aggregate location data
group sizes
characteristics of ROIs (density, size, time)
semantics of ROIs
Future Work

- Evaluate our methodology on different mobility datasets

A. Pyrgelis, E. De Cristofaro, G. Ross, UCL
Future Work

- Evaluate our methodology on different mobility datasets
- Privacy quantification and analysis of aggregate location data
Future Work

- Evaluate our methodology on different mobility datasets
- Privacy quantification and analysis of aggregate location data
  - group sizes
Future Work

- Evaluate our methodology on different mobility datasets
- Privacy quantification and analysis of aggregate location data
  - group sizes
  - characteristics of ROIs (density, size, time)
Future Work

- Evaluate our methodology on different mobility datasets
- Privacy quantification and analysis of aggregate location data
  - group sizes
  - characteristics of ROIs (density, size, time)
  - semantics of ROIs
Thanks for your attention! Any questions?

Contact Details:
apostolos.pyrgelis.14@ucl.ac.uk
Thanks for your attention! Any questions?
Thanks for your attention! Any questions?

Contact Details: apostolos.pyrgelis.14@ucl.ac.uk