NLS
FINNISH GEOSPATIAL
RESEARCH INSTITUTE
FGI

Building an Open Personal Trajectory Repository

GEOProcessing 2022

June 26, 2022 to June 30, 2022 - Porto, Portugal

IARIA

Dr. Ville Mäkinen[a], M.Sc. Anna Brauer[a,b], Prof. Juha Oksanen[a]

a) Finnish Geospatial Research Institute FGI

b) University of Helsinki

ville.p.makinen@nls.fi

Short resume

- Ville Mäkinen
- PhD on theoretical physics, University of Jyväskylä, Finland
- 2014 → FGI, department of Geoinformatics and Cartography
  - Leading Geodata Science and Geocomputation research group
- Various geospatial analyses, algorithm development, high-performance computing, personal location data

Background

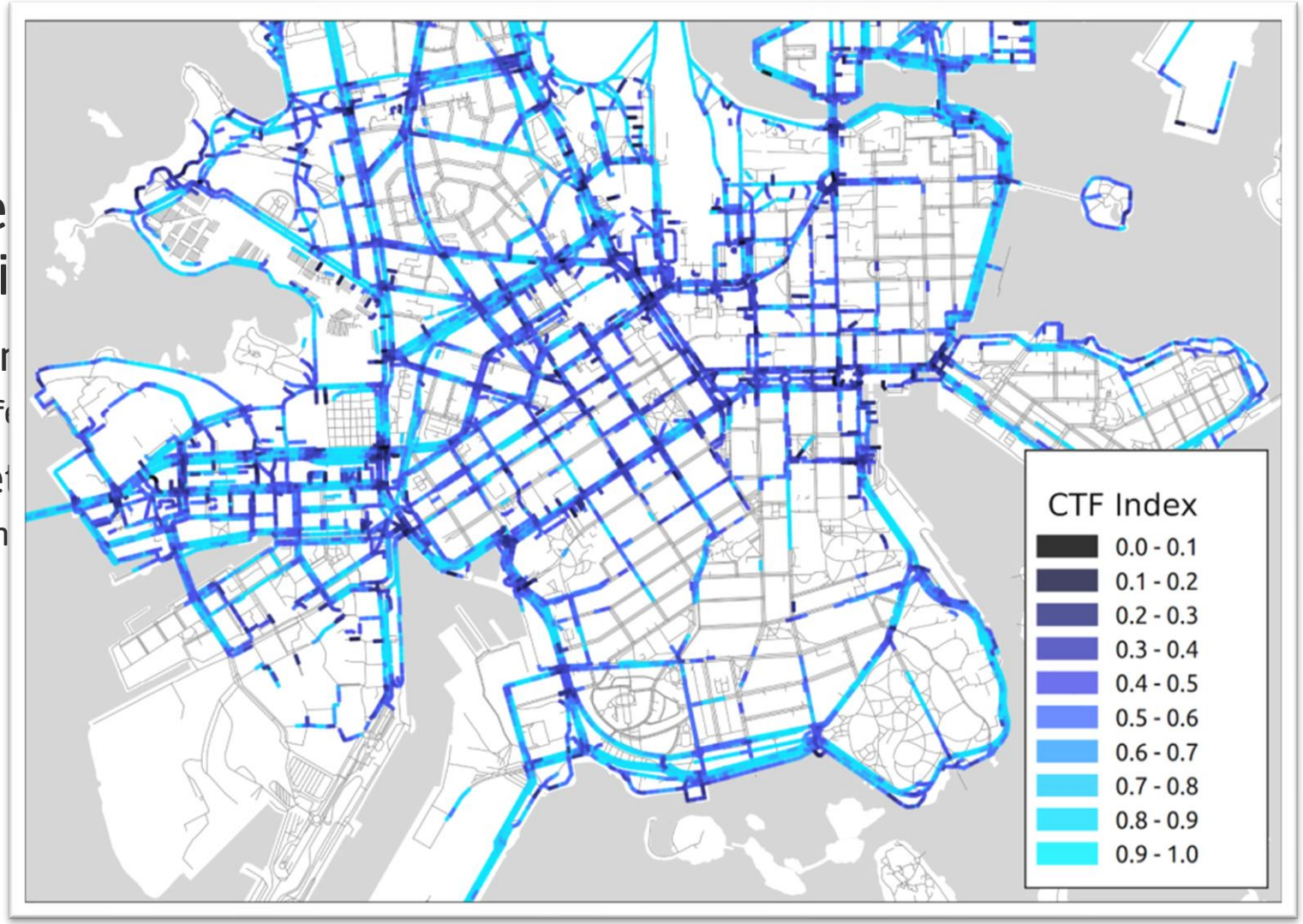
Huge amount of accurate personal movement data is gathered by GNSS-enabled smart devices

- Would be invaluable for studying the non-motorized traffic in cities
  - Missing routes, bottlenecks, unsafe routes
- Would allow for much more detailed analyses than simple aggregations
  - Time-dependence, direction, cycling fluency

Background

Huge amount of accurate
GNSS-enabled smart devi

- Would be invaluable for studyi
  - Missing routes, bottlenecks, unsaf
- Would allow for much more de
  - Time-dependence, direction, cyclin



Background

Huge amount of accurate personal movement data is gathered by GNSS-enabled smart devices

- Would be invaluable for studying the non-motorized traffic in cities
  - Missing routes, bottlenecks, unsafe routes
- Would allow for much more detailed analyses than simple aggregations
  - Time-dependence, direction, cycling fluency
- As personal data it compromises the privacy of the data subjects
  - Revelation of sensitive locations, re-identification, …

Background

Huge amount of accurate personal movement data is gathered by GNSS-enabled smart devices

- Would be invaluable for studying the non-motorized traffic in cities
  - Missing routes, bottlenecks, unsafe routes
- Would allow for much more detailed analyses than simple aggregations
  - Time-dependence, direction, cycling fluency
- As personal data it compromises the privacy of the data subjects
  - Revelation of sensitive locations, re-identification, linkage
- We are studying methods for publishing personal trajectories in such a way that
  - The privacy of the data subjects is respected
  - The utility of the data is preserved as much as possible

Motivation

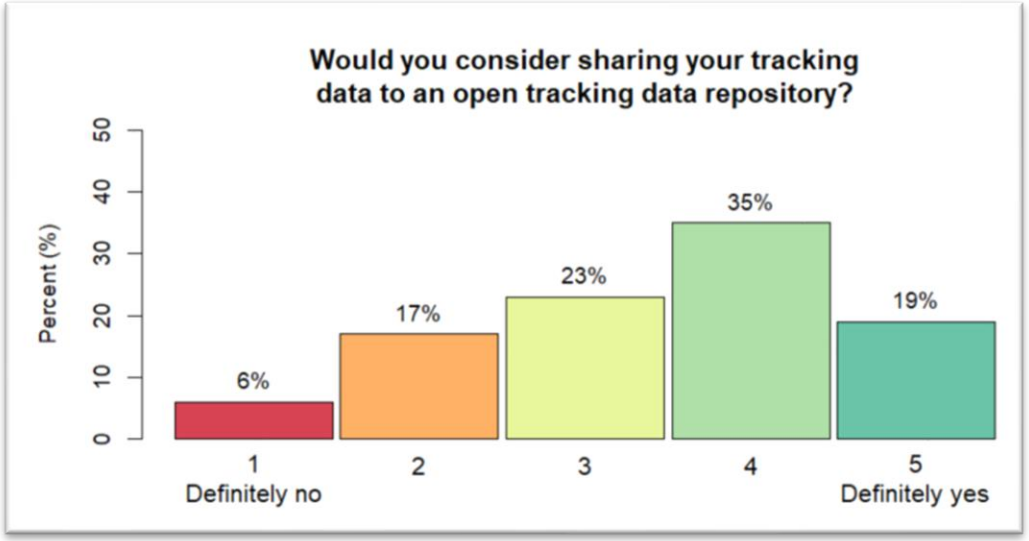
Our reasons to start to develop and build the data donation service

- There are almost no accurate human mobility data sets available – needed for research
  - Geolife (Microsoft Research Asia) – outdated (2007 – 2012)
- The questionnaire about building such a service indicated that citizens are interested in the topic and are willing to donate their data
  - Do the answers reflect the reality?

Would citizens contribute their personal location data to an open database? Preliminary results from a survey

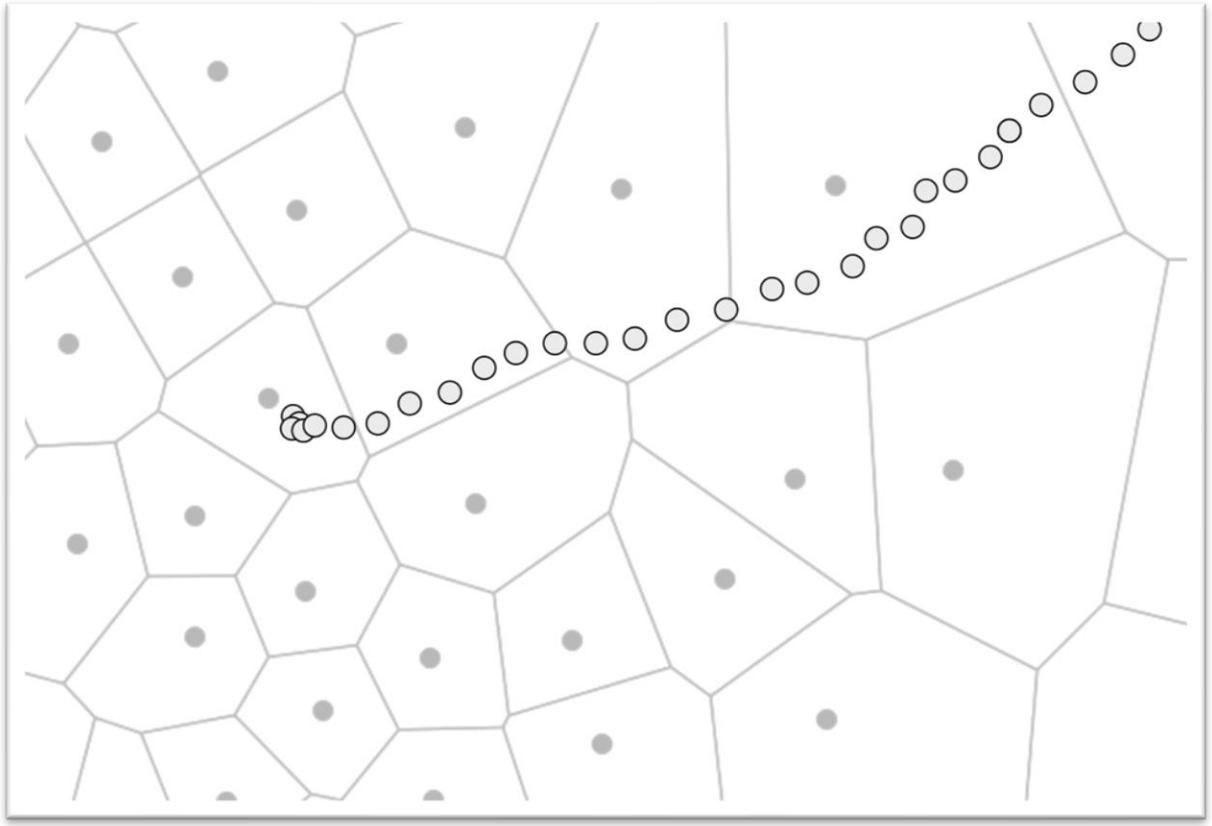
Vilma Jokinen*, Ville Mäkinen*, Anna Brauer* and Juha Oksanen*

*Finnish Geospatial Research Institute FGI, National Land Survey of Finland



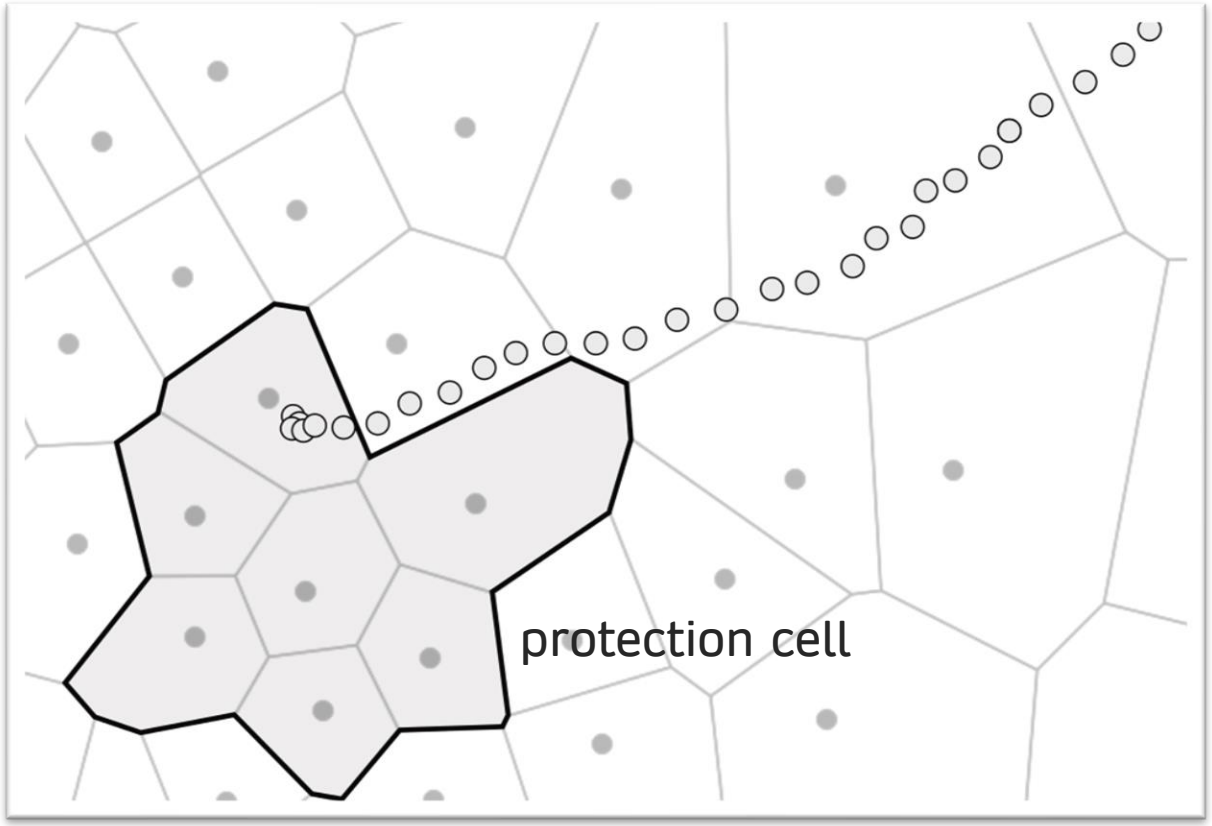
Materials and methods 1/2

The main tool: Site-dependent Trajectory Truncation (S-TT)



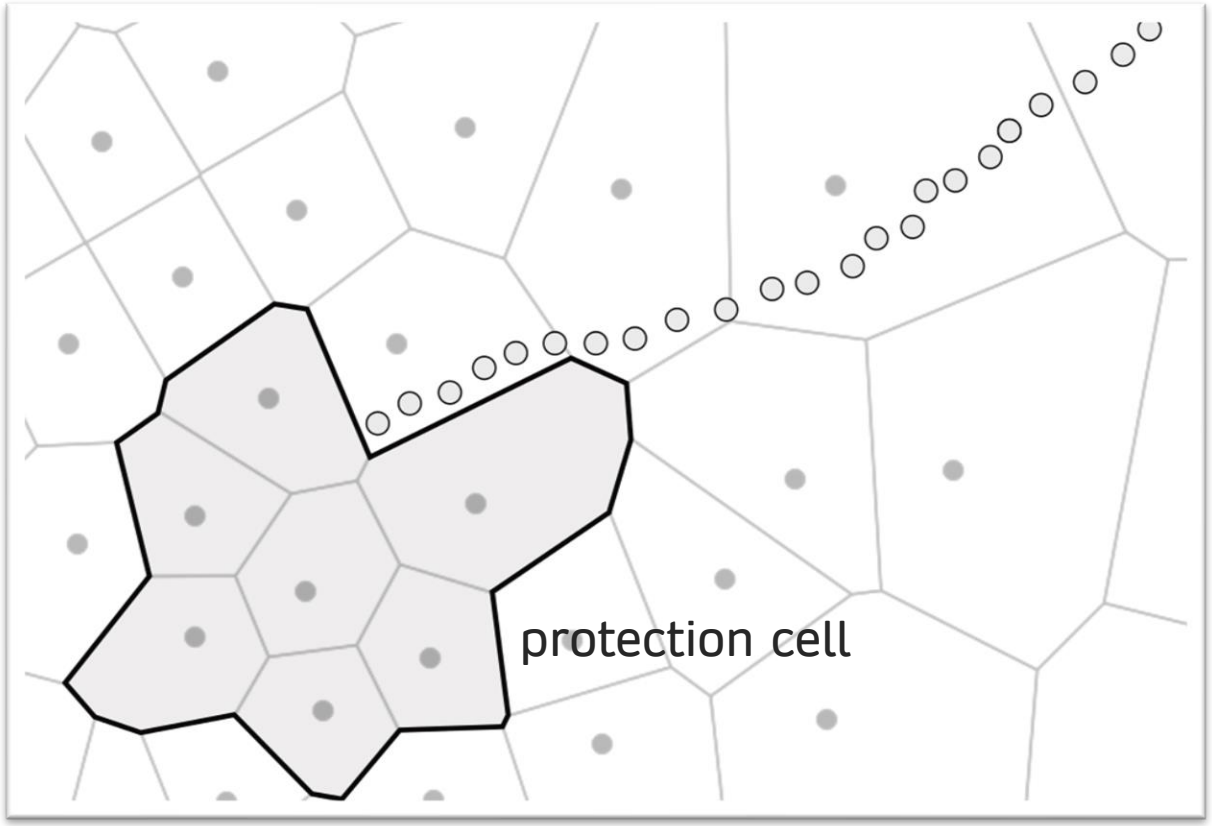
Materials and methods 1/2

The main tool: Site-dependent Trajectory Truncation (S-TT)



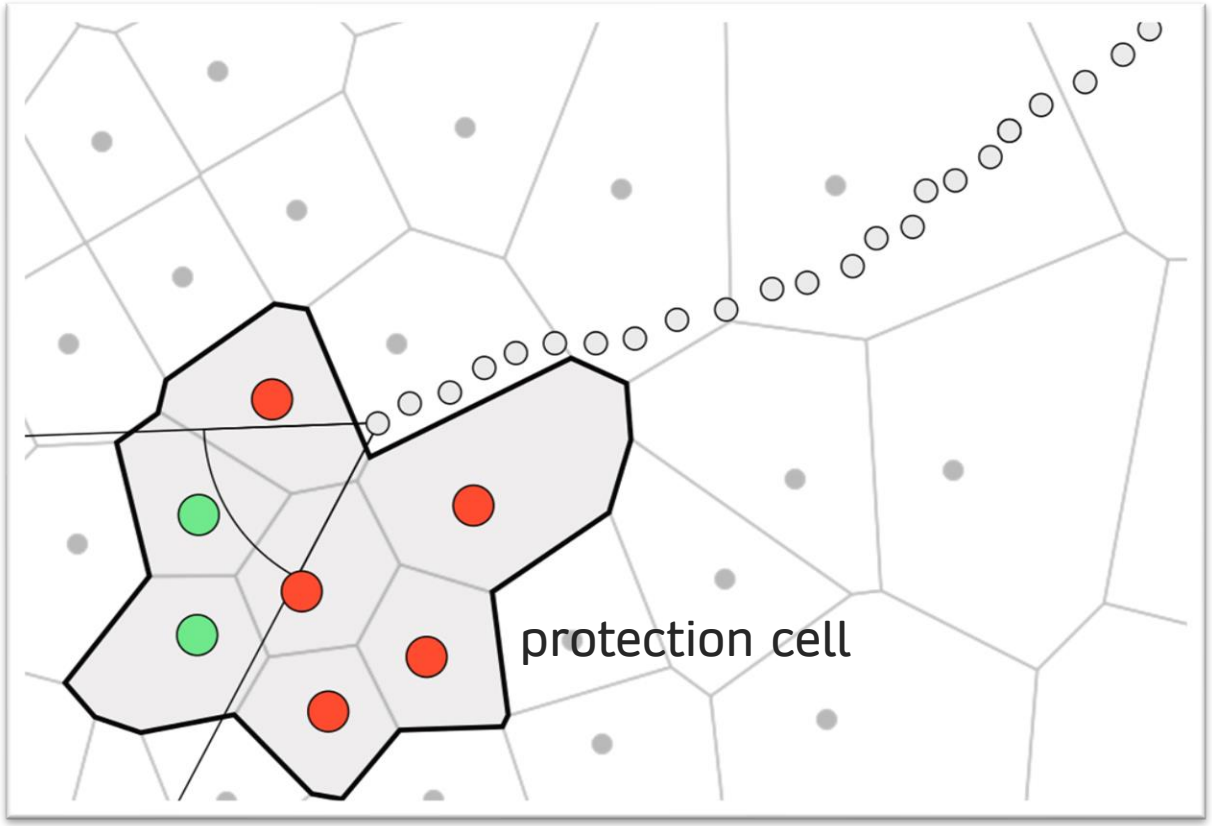
Materials and methods 1/2

The main tool: Site-dependent Trajectory Truncation (S-TT)



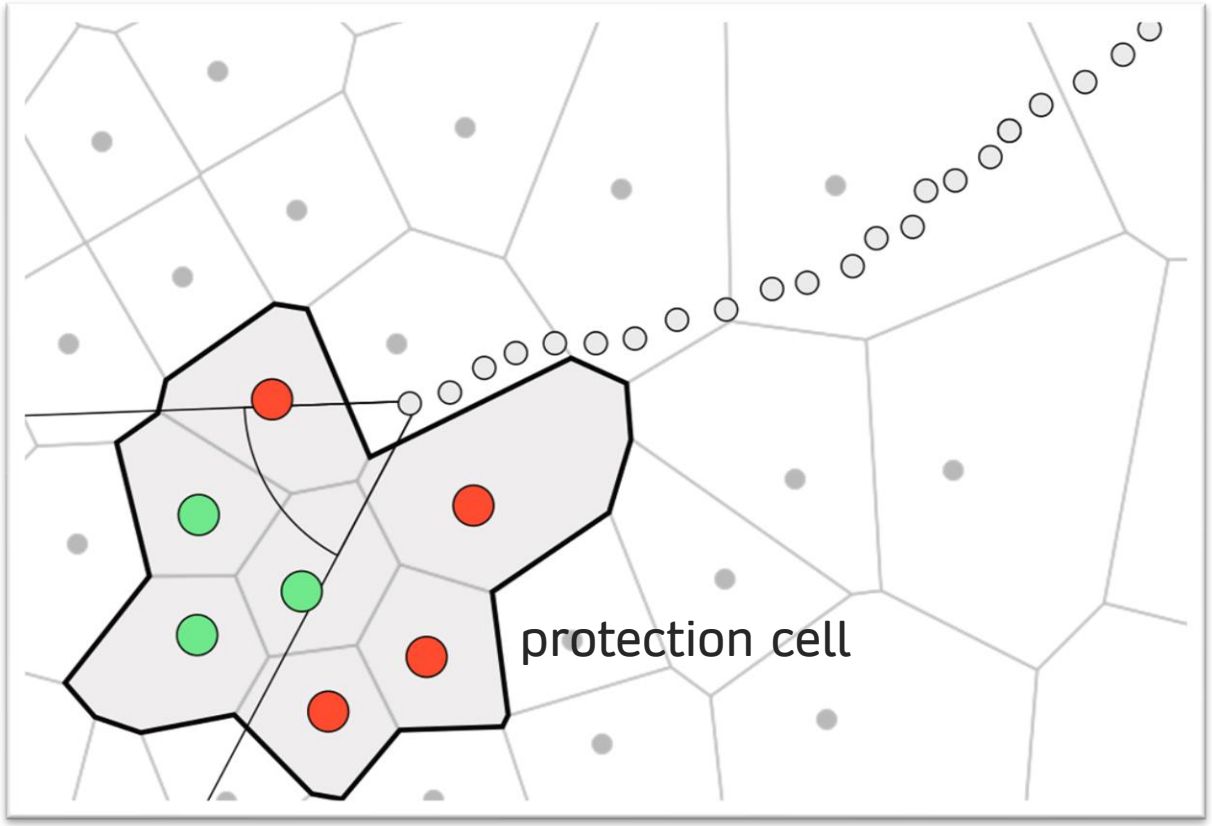
Materials and methods 1/2

The main tool: Site-dependent Trajectory Truncation (S-TT)



Materials and methods 1/2

The main tool: Site-dependent Trajectory Truncation (S-TT)

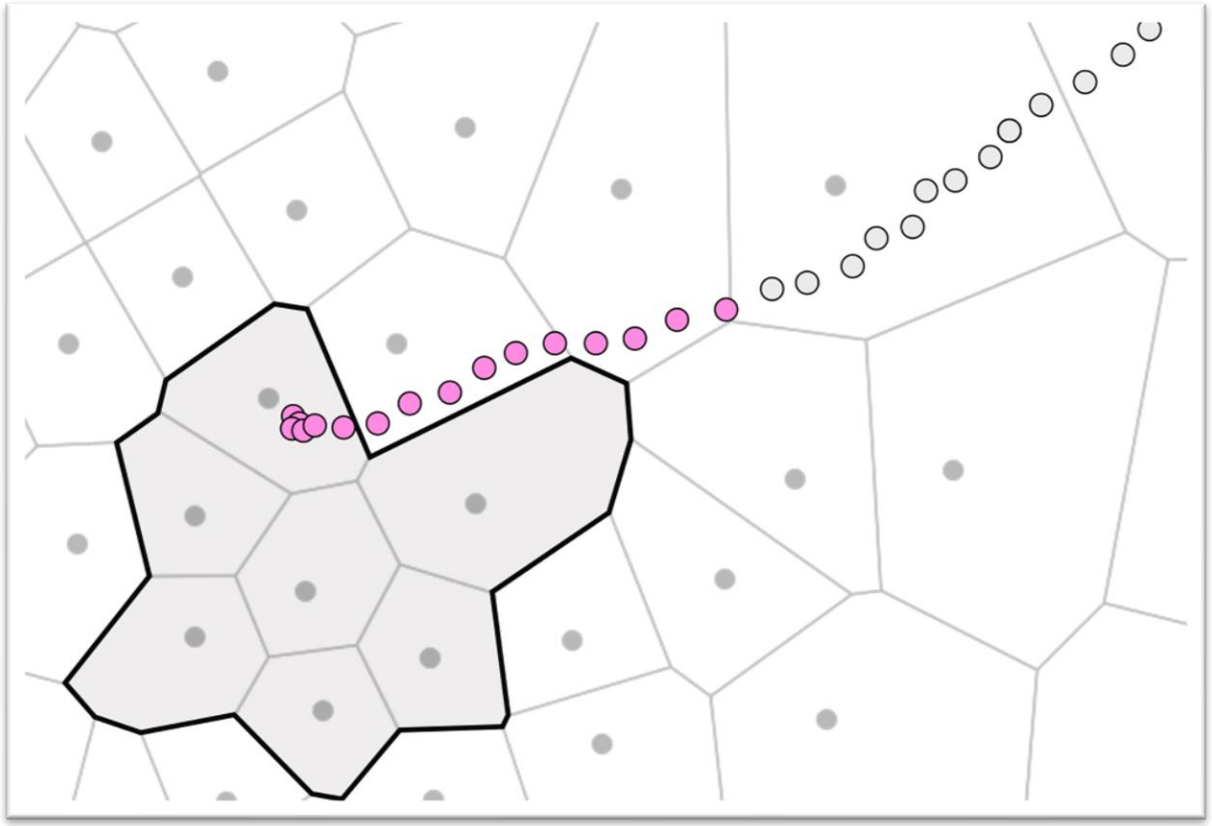


Materials and methods 1/2

The main tool: Site-dependent Trajectory Truncation (S-TT)

Materials and methods 1/2

The main tool: Site-dependent Trajectory Truncation (S-TT)



Materials and methods 1/2

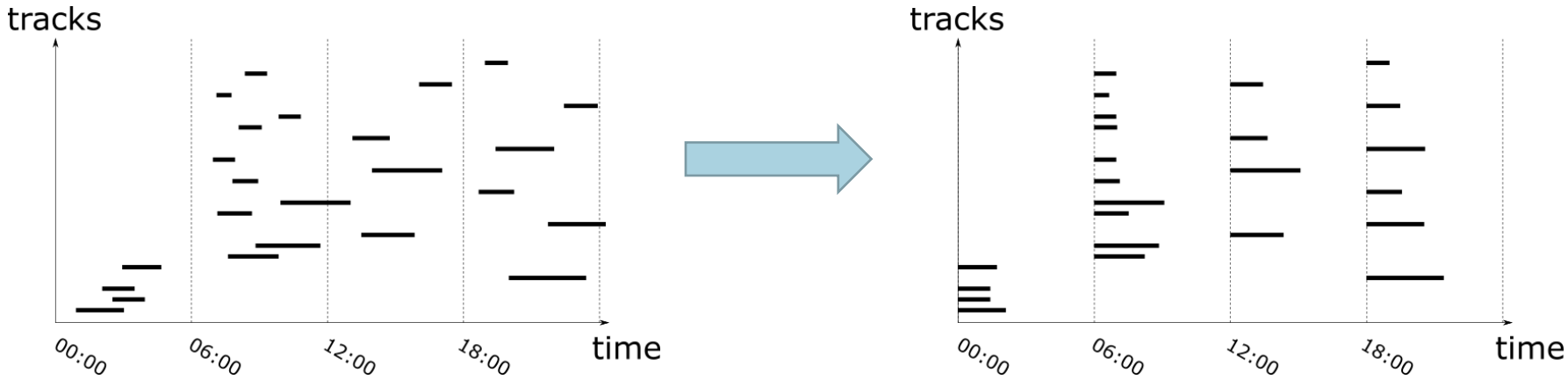
The main tool: Site-dependent Trajectory Truncation (S-TT)

- Assumption: the sensitive locations the trajectories may reveal are buildings
- Building data required from the processing area
- Buildings are clustered into “protection sets” of at least K buildings
- S-TT: the attacker can narrow down the sensitive location only to a cluster, bot not to any individual site in the cluster

Materials and methods 2/2

Processing of a trajectory

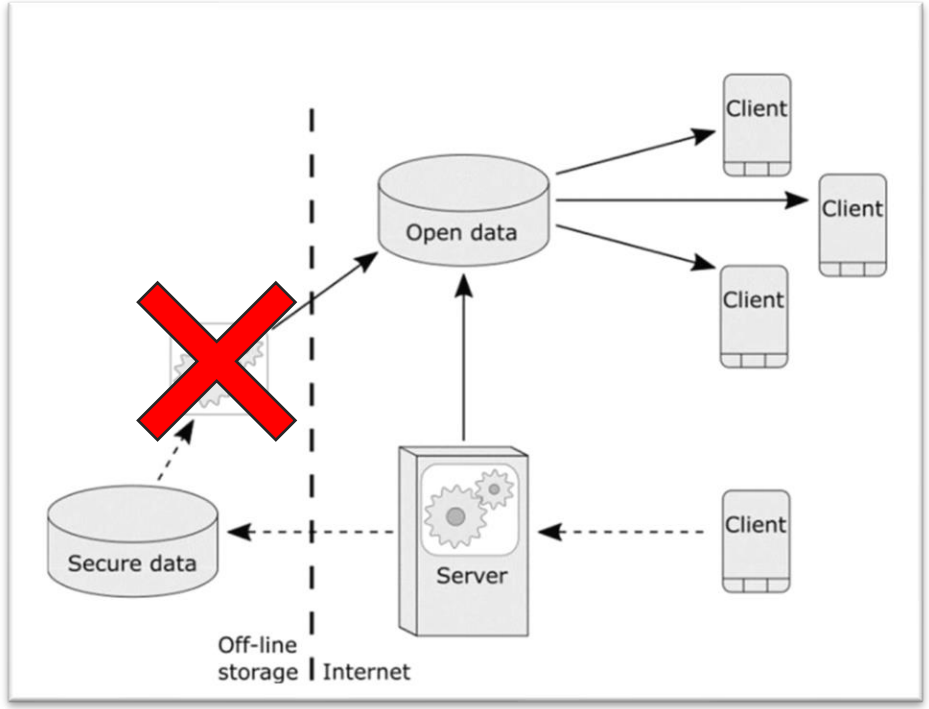
- Detect stay points – split trajectories
- Truncate the parts using the Site-dependent Trajectory Truncation
- Shift the timestamps of the trajectory points
  - Day is split into fixed time ranges
  - The trajectory is shifted to start at the beginning of the time range



Current status

A development version running online (not yet public)

- The original trajectories are encrypted, then moved to an offline location (for research only)
- The original trajectories are processed immediately after upload
- Users decide which processed trajectories are stored in the open database
- Possible open data products are created only from the anonymized trajectories



Current issues

We do not have feedback from end users yet

- Getting data that is recorded with third-party applications is not straightforward
- Laborious to upload trajectory data

⇒ Would a dedicated mobile app attract more users?

How to motivate users to participate?

- As a research pilot, we cannot provide any long-term effects (city planning)
- We cannot (and are not trying to) compete with the commercial actors

Conclusions

We are exploring new ways to use citizens' personal movement data for common good

- Difficult to develop reliable anonymization methods without proper test data
- Research pilot – in a production version the anonymization could take place on users' devices
- We are not competing with commercial actors

Acknowledgments & references

The GeoPrivacy project is funded by the Finnish Cultural Foundation.

[CTF] Brauer, A., Mäkinen, V. and Oksanen, J. “Characterizing cycling traffic fluency using big mobile activity tracking data”, Computers, Environment and Urban Systems 2021, 85, 101553 https://doi.org/10.1016/j.compenvurbsys.2020.101553

[GeoLife] https://www.microsoft.com/en-us/download/details.aspx?id=52367

[Survey] Jokinen, V., Mäkinen, V., Brauer, A. and Oksanen, J. “Would citizens contribute their personal location data to an open database? Preliminary results from a survey”, in Basiri, A., Gartner, G., & Huang, H. (Eds.). (2021). LBS 2021: Proceedings of the 16th International Conference on Location Based Services. https://doi.org/10.34726/1741

[S-TT] Brauer, A., Mäkinen, V., Forsch, A., Oksanen, J. and Haunert, J.-H. “My home is my secret: concealing sensitive locations by context-aware trajectory truncation”, Internation Journal of Geographical Information Science 2022, early access. https://doi.org/10.1080/13658816.2022.2081694

Thank you!

ville.p.makinen@nls.fi

vipemaki