NTT

Importance of realistic adversary model for anonymity evaluation: A case study of trajectory data

Shogo Masaki
(NTT Corporation, Japan)
ref.: Masaki et al. in prep.
email: masaki.shogo_at_lab.ntt.co.jp

Plan

- Introduction
  - PWS Cup
  - Motivation
- Anonymity evaluation of trajectory data
  - preliminary
  - consideration of properties
  - data publication model
  - adversary background knowledge model
  - inference attack
  - experiments
- Summary

Introduction: PWS Cup

Innovative R&D by NTT

- PWS Cup 2015, 2016
  - the first attempts of anonymization competition initiated in Japan
  - It was fun!

Introduction: PWS Cup

- PW
  - n in
  -

Introduction: PWS Cup

- PW
  - n in
  -

Introduction: PWS Cup

- PWS Cup 2015, 2016
  - the first attempts of anonymization competition initiated in Japan
  - It was fun!
  - stimulating and gave us new ideas on anonymization techniques
  - going global is very welcomed for progresses in this field

Introduction: Motivation

- Expectation for the competition: **indication of *effective* anonymization methods**

- To do so, it is necessary to evaluate anonymity & utility in appropriate manner.

Introduction: Motivation

- Anonymity evaluation?
  - data publication model
  - adversary background knowledge model
  - inference attack model

- The realistic models should be constructed with consideration of the properties of the target data type.

Introduction: Motivation

- Anonymity evaluation?
  - data publication model
  - adversary background knowledge model
  - inference attack model

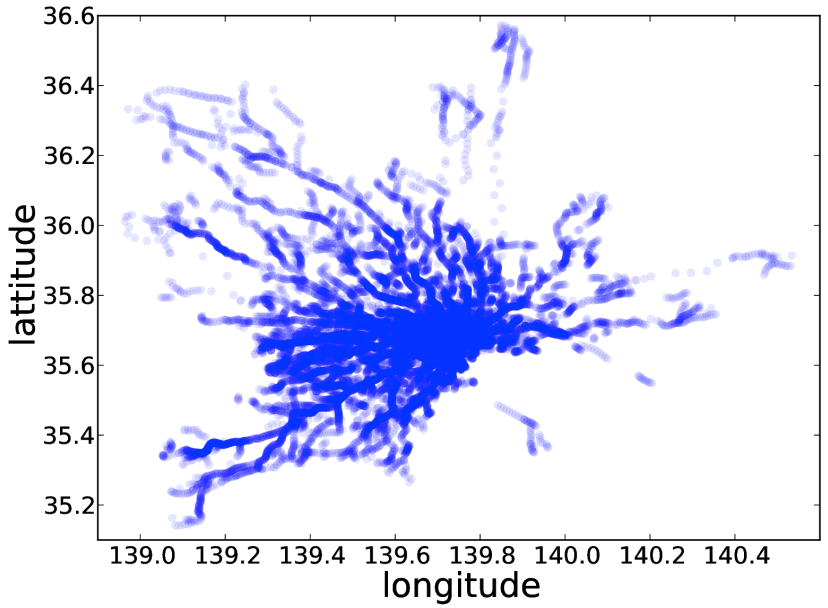
- The realistic models should be constructed with consideration of the properties of the target data type.

- **This talk:**
  - our on-going work on the anonymity evaluation of trajectory data
  - preliminary experimental results

Anonymity evaluation of trajectory data: preliminary

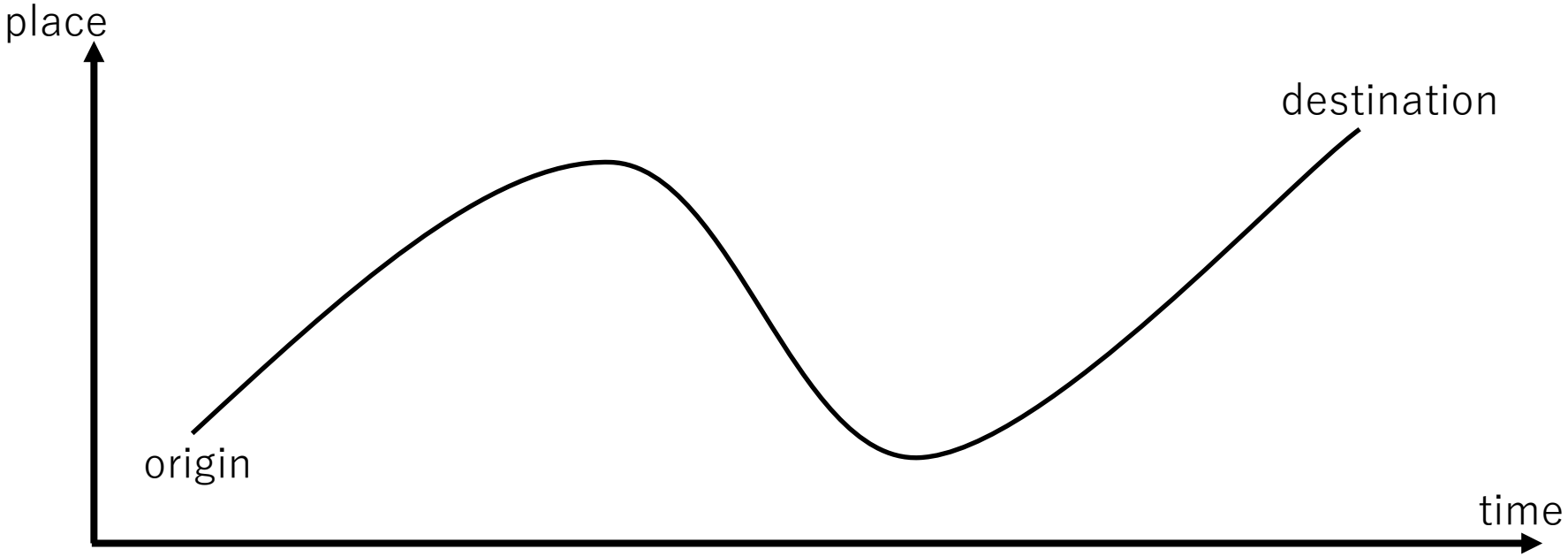
- trajectory data:
  - time-series of location data
  - high commercial and research potential
  - can be a scope of the future anonymization competition

ID	time	latitude	longitude
1	8:00	35.6836	139.4710
1	8:03	35.6830	139.4713
3	8:00	35.7207	139.5555
3	8:04	35.7214	139.5577
3	8:10	35.7216	139.5600
3	8:20	35.7222	139.5622
…	…	…	…



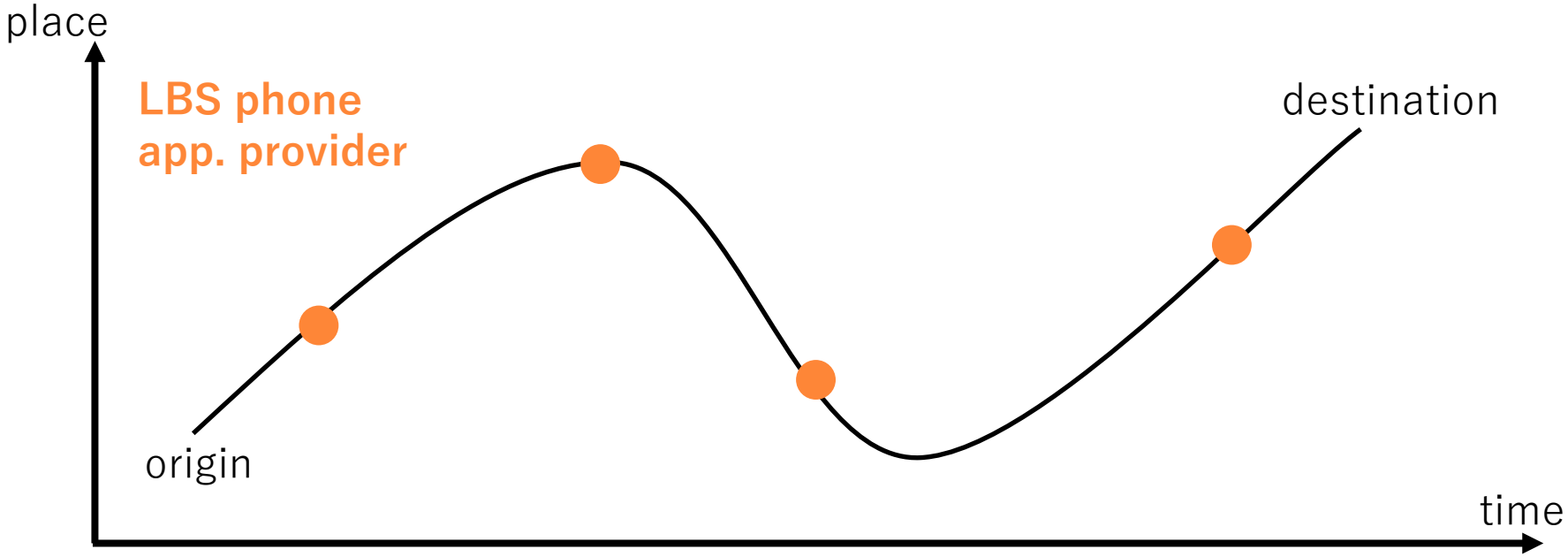
Anonymity evaluation of trajectory data: consideration of properties

- More than 2 entities can obtain the trajectory of one mobility but at different time-stamps.



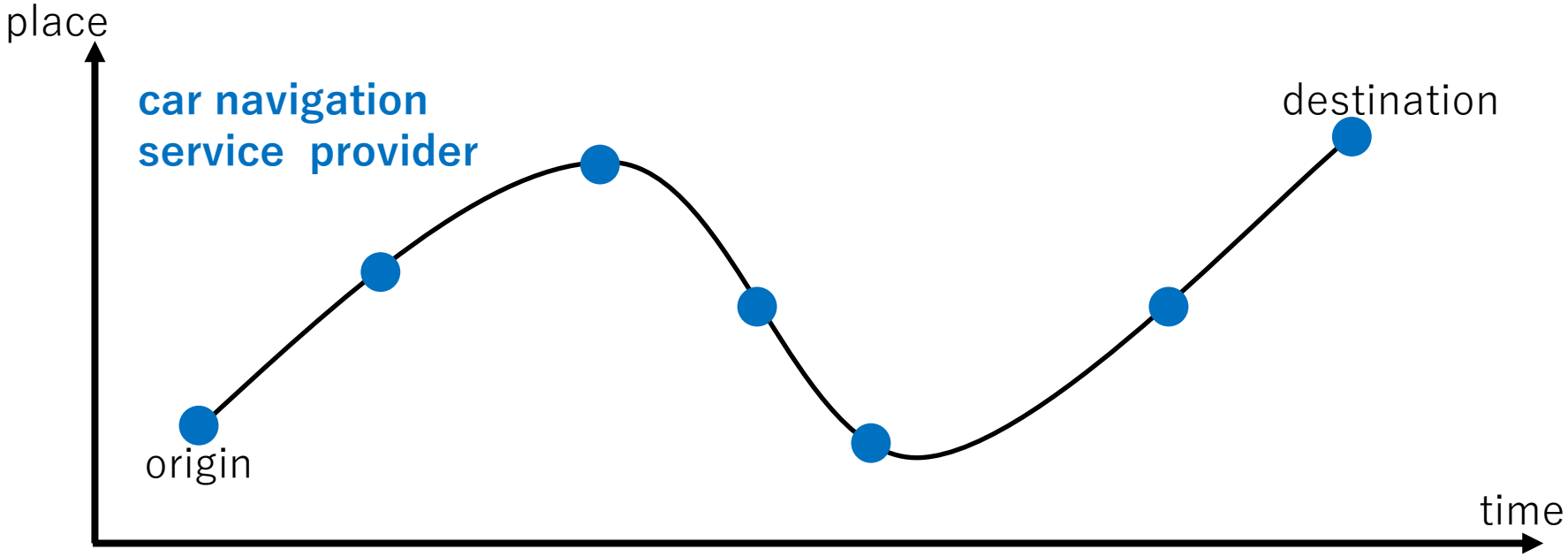
Anonymity evaluation of trajectory data: consideration of properties

- More than 2 entities can obtain the trajectory of one mobility but at different time-stamps.



Anonymity evaluation of trajectory data: consideration of properties

- More than 2 entities can obtain the trajectory of one mobility but at different time-stamps.



Anonymity evaluation of trajectory data: data publication model

- A trajectory data holder provide another trajectory data holder with the anonymized data.
- Why?
  - The data receiver may want to improve statistics in analyses.

- The data receiver may re-identify individuals in the anonymized data by finding the similar trajectories.

Anonymity evaluation of trajectory data: data publication model

- A trajectory data holder provide another trajectory data holder with the anonymized data.
- Why?

trajectory data holder can be a strong adversary

- The data receiver may re-identify individuals in the anonymized data by finding the similar trajectories.

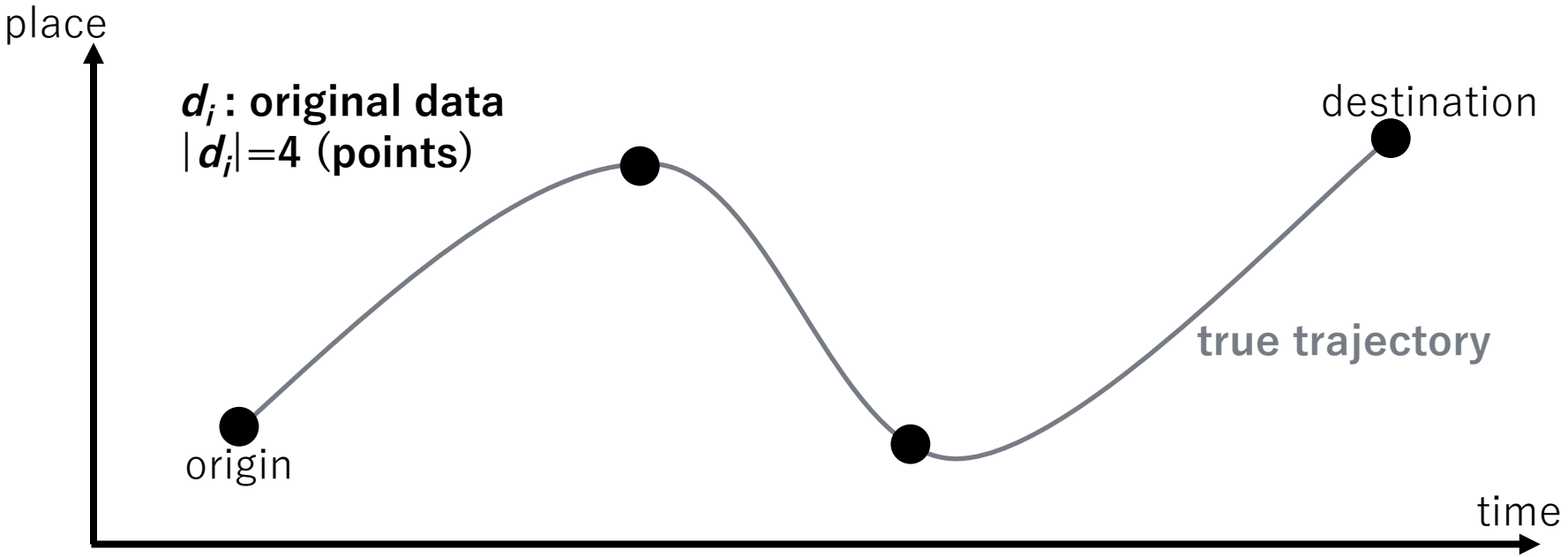
Anonymity evaluation of trajectory data: adversary background knowledge model

Innovative R&D by NTT

- The data sets of 2 trajectory data holders cannot be available in general.
- We synthesize the adversary background knowledge from the original data in a very simple way.

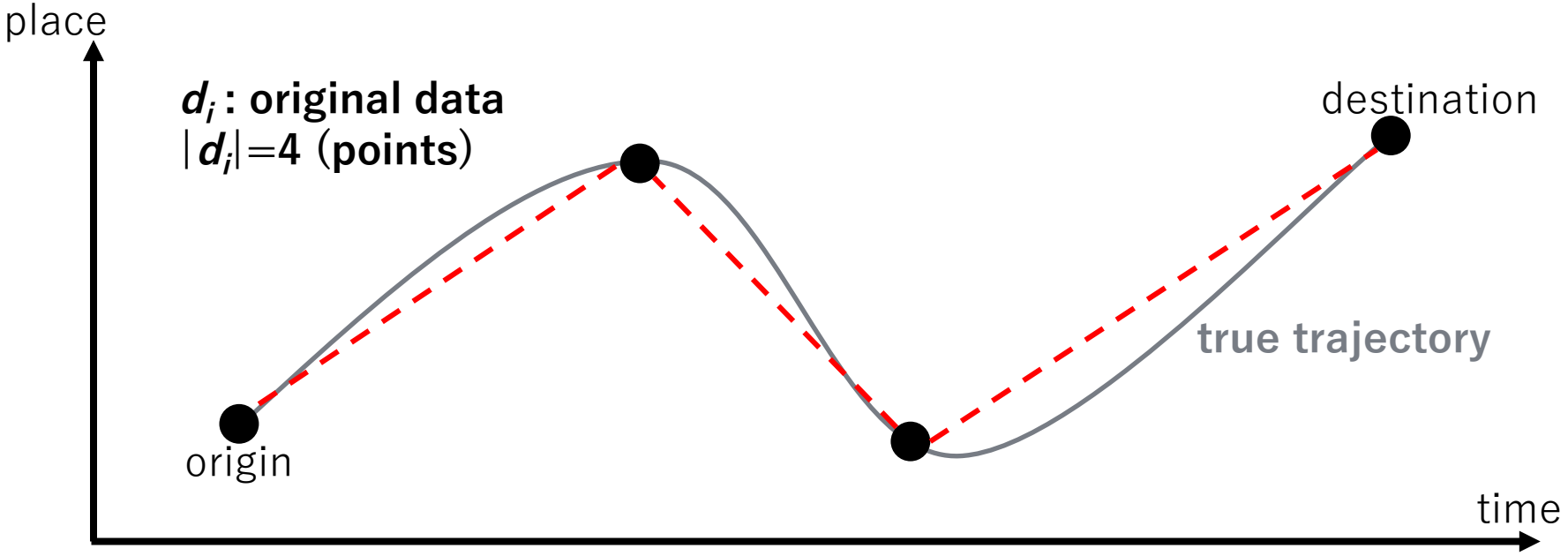
Anonymity evaluation of trajectory data: adversary background knowledge model

- The original data is given but not the true trajectory.



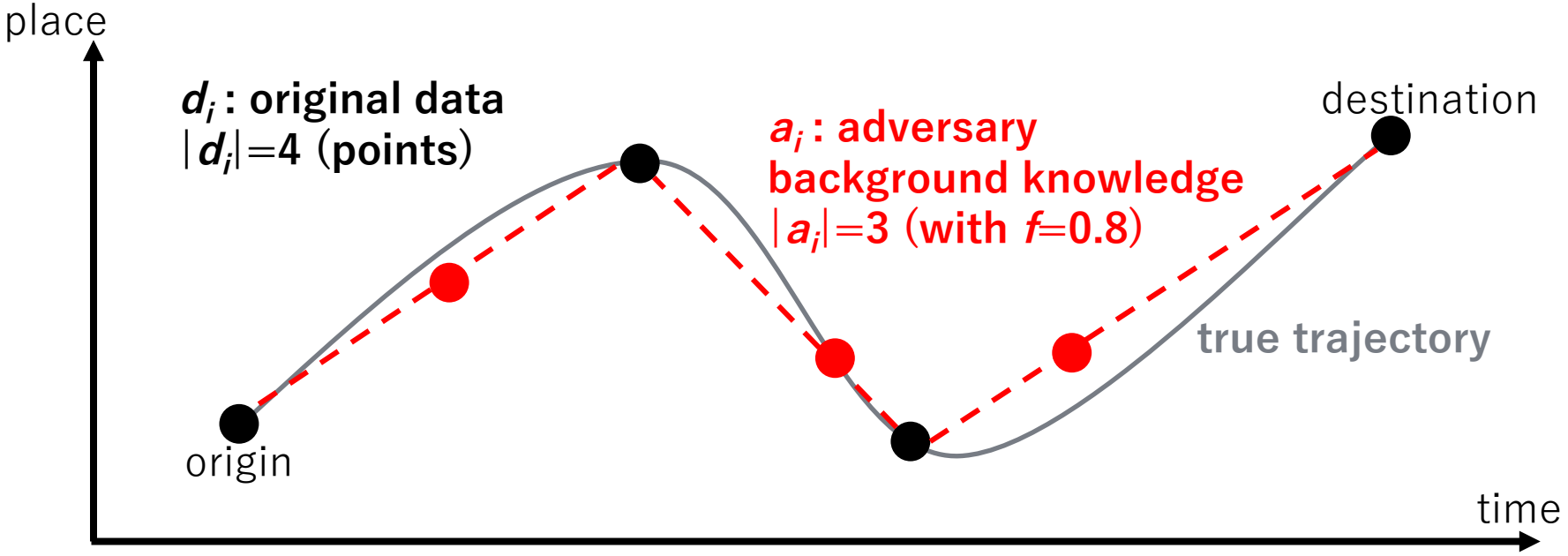
Anonymity evaluation of trajectory data: adversary background knowledge model

- We linearly interpolate lat. & lon. as a function of time in all time intervals.
  - i.e., $\mathrm{lat} = \alpha_{\mathrm{lat}} \times \mathrm{time} + \beta_{\mathrm{lat}}$
    $\mathrm{lon} = \alpha_{\mathrm{lon}} \times \mathrm{time} + \beta_{\mathrm{lon}}$



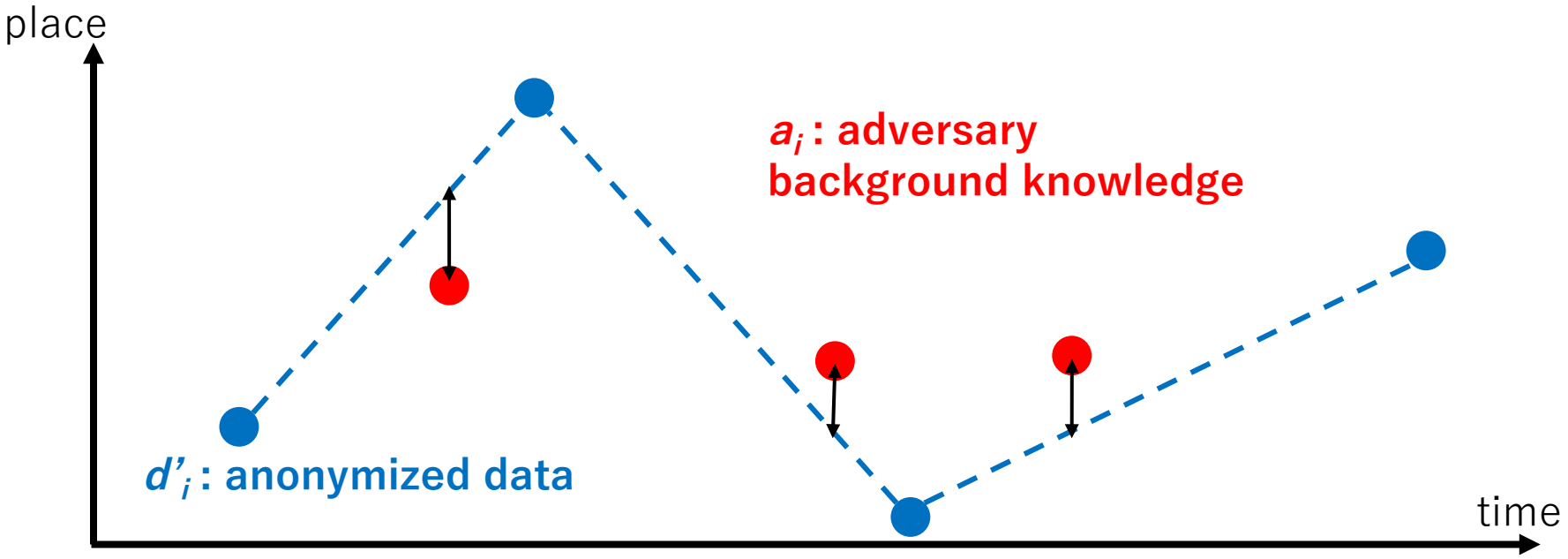
Anonymity evaluation of trajectory data: adversary background knowledge model

- We choose a time-stamp randomly between the origin and destination time.
- The total number of time-stamps is $f \times |d_i|$.
- We obtain a trajectory as adversary background knowledge.



Anonymity evaluation of trajectory data: inference attack

- We measure the geological distance between anonymized trajectory and adversary background knowledge using linear interpolation, again.
- We search the closest trajectory to re-identify an individual.



Anonymity evaluation of trajectory data: experiments

Innovative R&D by NTT

- Data set
  - cabspotting data (Piorkowski+'09)
  - 536 taxis trajectories in SF

- Pre-processing
  - use only 15 taxis
  - split by 4 hours
    → 1,333 trajectories (=virtual taxis)
    with 242,416points (~180 points/trajectory)

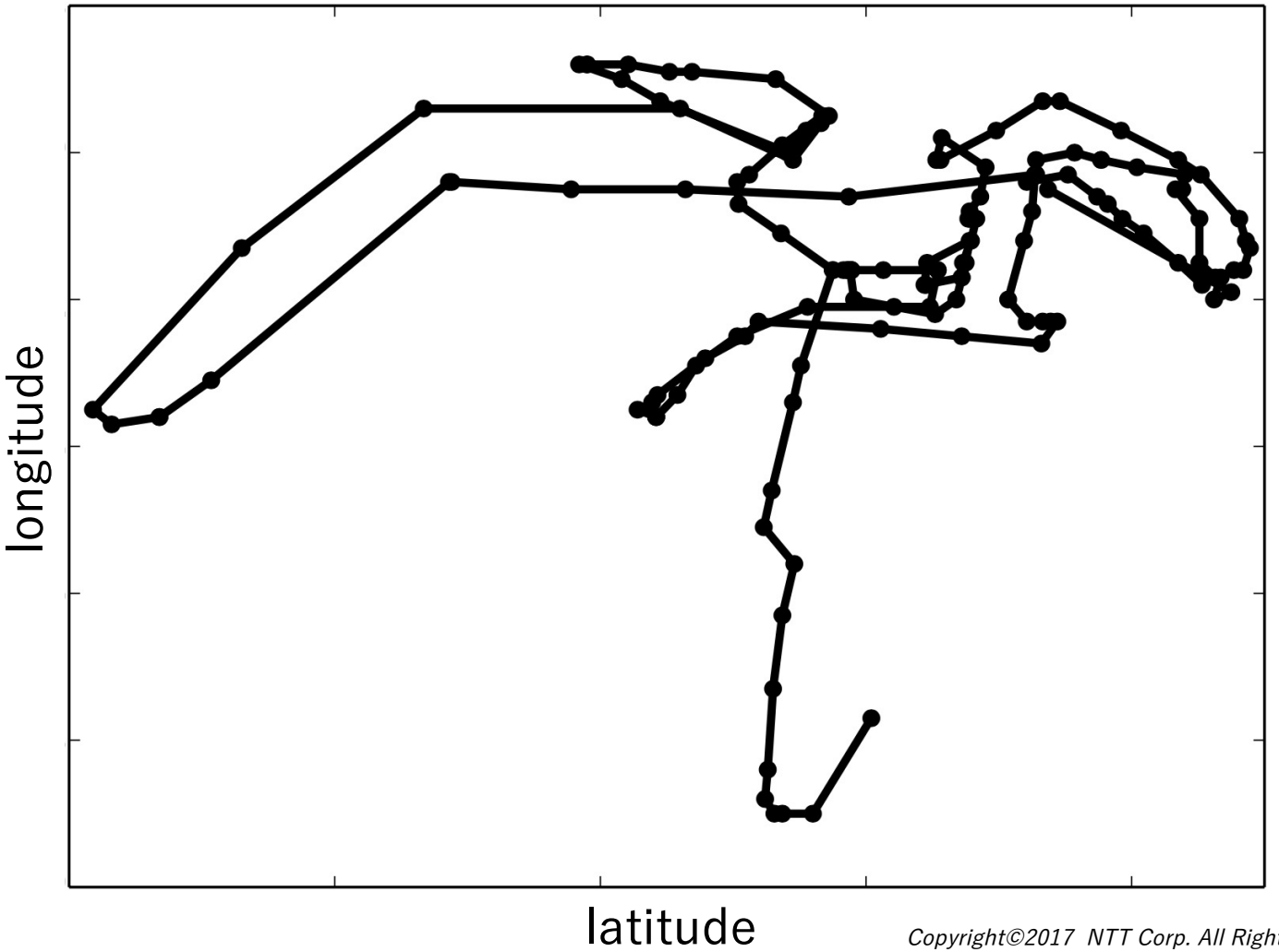
Anonymity evaluation of trajectory data: experiments

- Anonymization
  - add the Laplace noise on lat. & lon. (cf., Geo-indistinguishability; Andres+'13)
  - 2 different sizes of the noise
    - average spatial error: 110m (small), 2km (large)

- Anonymity evaluation
  - our method with $0.1 < f < 2.0$
  - POI extraction attack (drawn from Primault+'15)

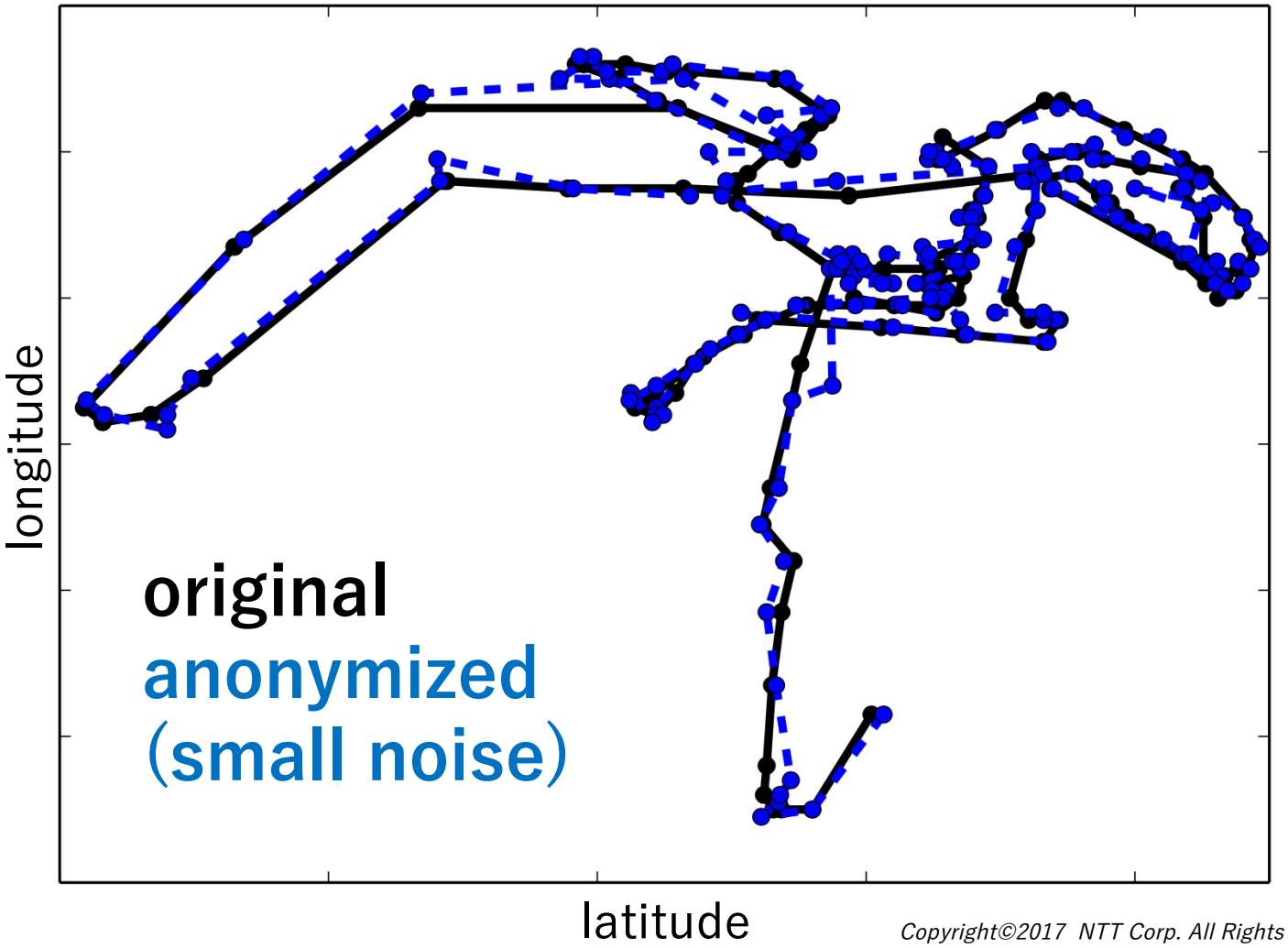
Anonymity evaluation of trajectory data: experiments

- an original trajectory



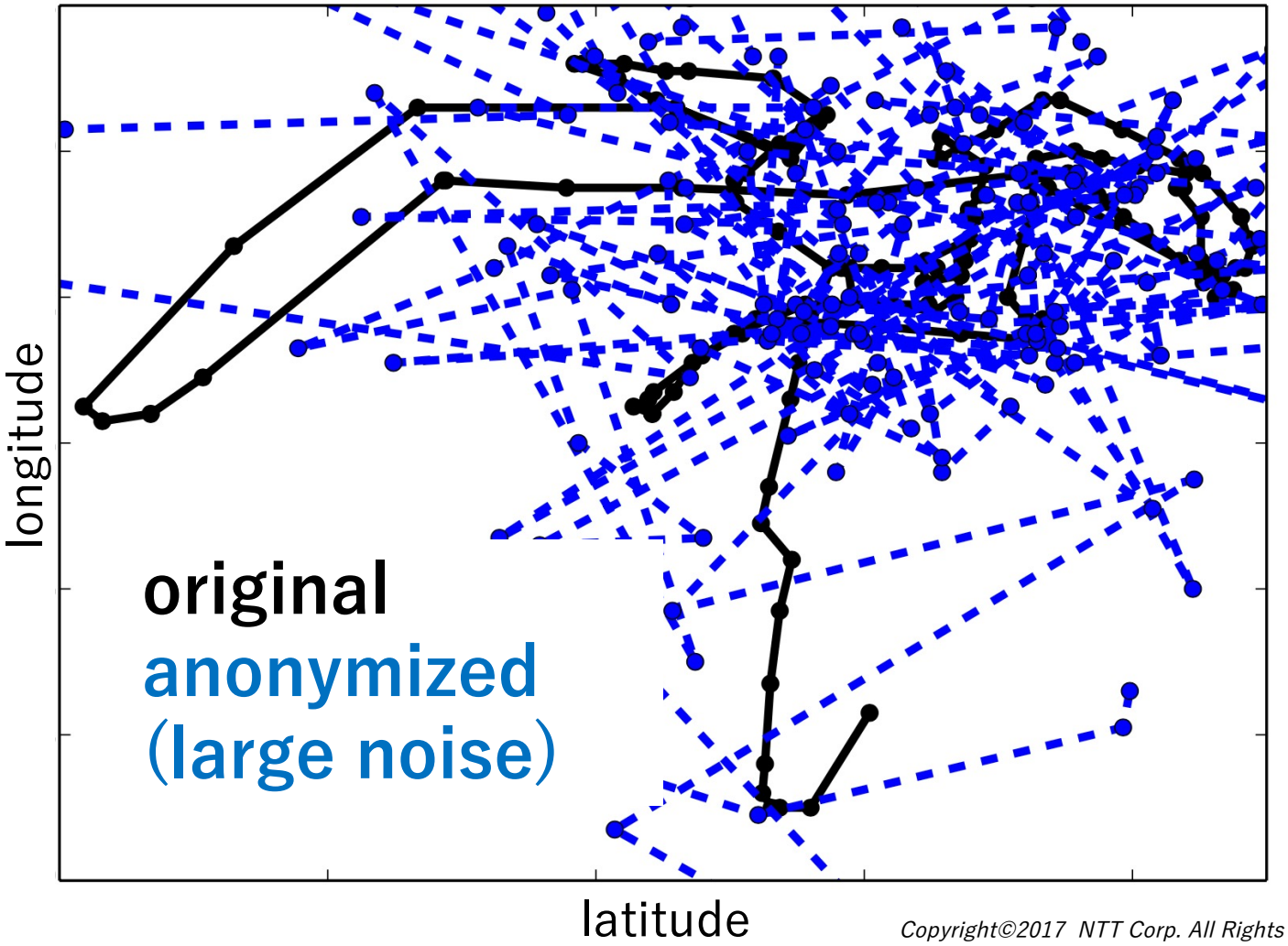
Anonymity evaluation of trajectory data: experiments

- with the anonymized trajectory



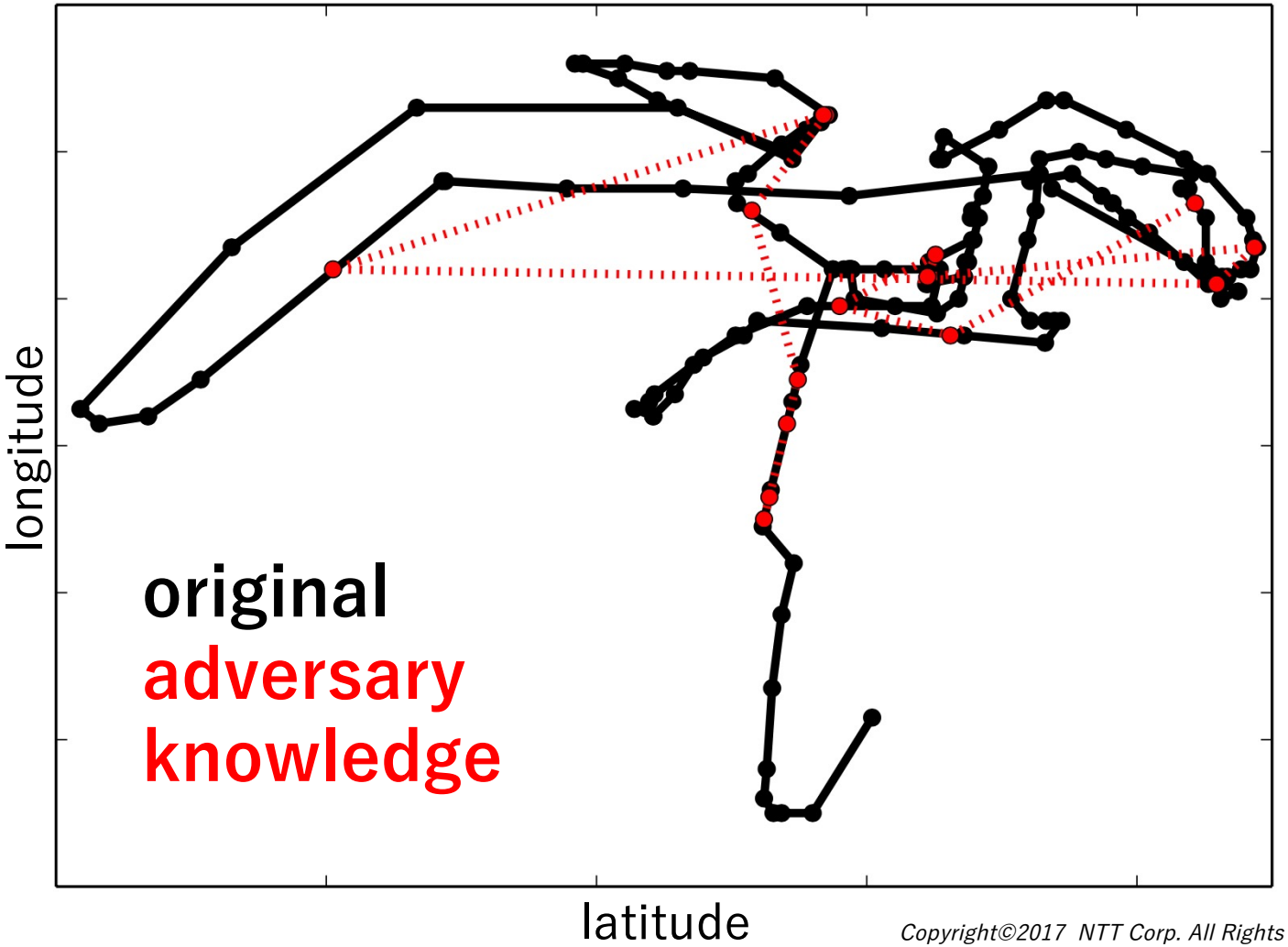
Anonymity evaluation of trajectory data: experiments

- with the anonymized trajectory



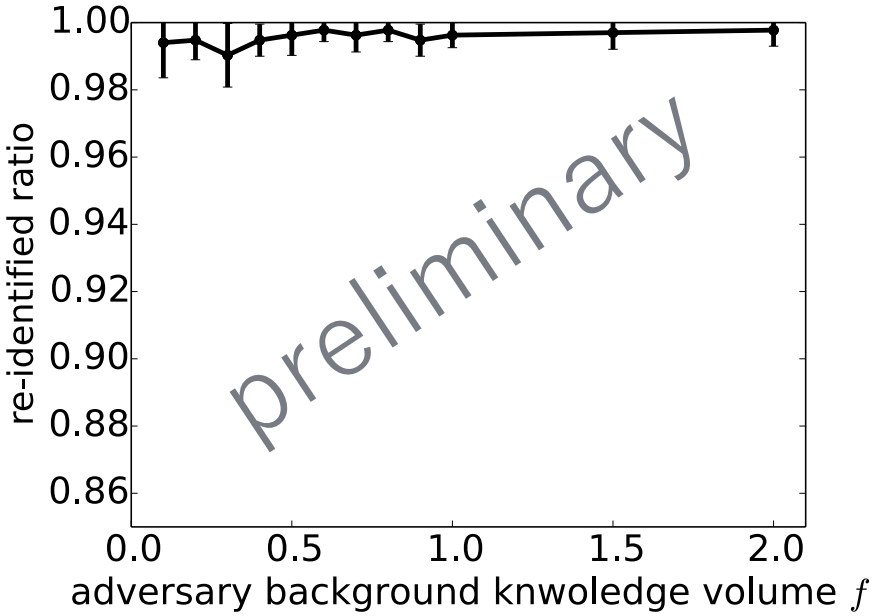
Anonymity evaluation of trajectory data: experiments

- with the correspondence in the adversary knowledge



Anonymity evaluation of trajectory data: experiments

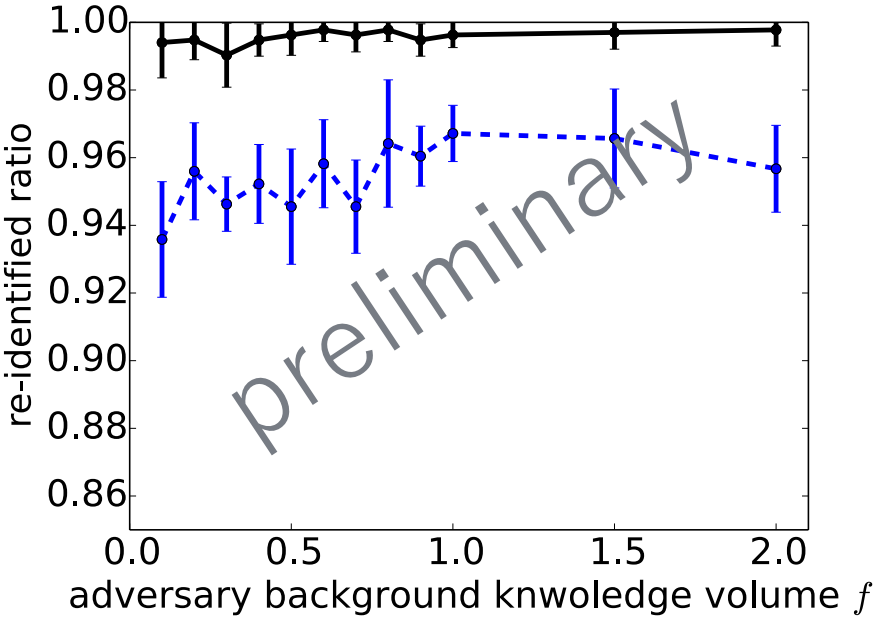
- Results for small noise
  - more than 98% for all f–values (adversary knowledge volume)



  - much higher than 1.5% from a POI extraction attack (Primault+'15)

Anonymity evaluation of trajectory data: experiments

- Results for large noise
  - still very high re-id ratio (>92%)



  - A trajectory data holder can be a strong adversary against to the anonymized trajectory data.

Summary

Innovative R&D by NTT

- The anonymization competition is fun!
- Anonymity evaluation is important.

- As a case study, we show our on-going work on anonymity evaluation of trajectory data.
- Our preliminary results mean that a trajectory data holder can be realistic and a very strong adversary.

- **Detailed & realistic model construction is needed for convincing anonymity evaluation.**