大阪大学
OSAKA UNIVERSITY

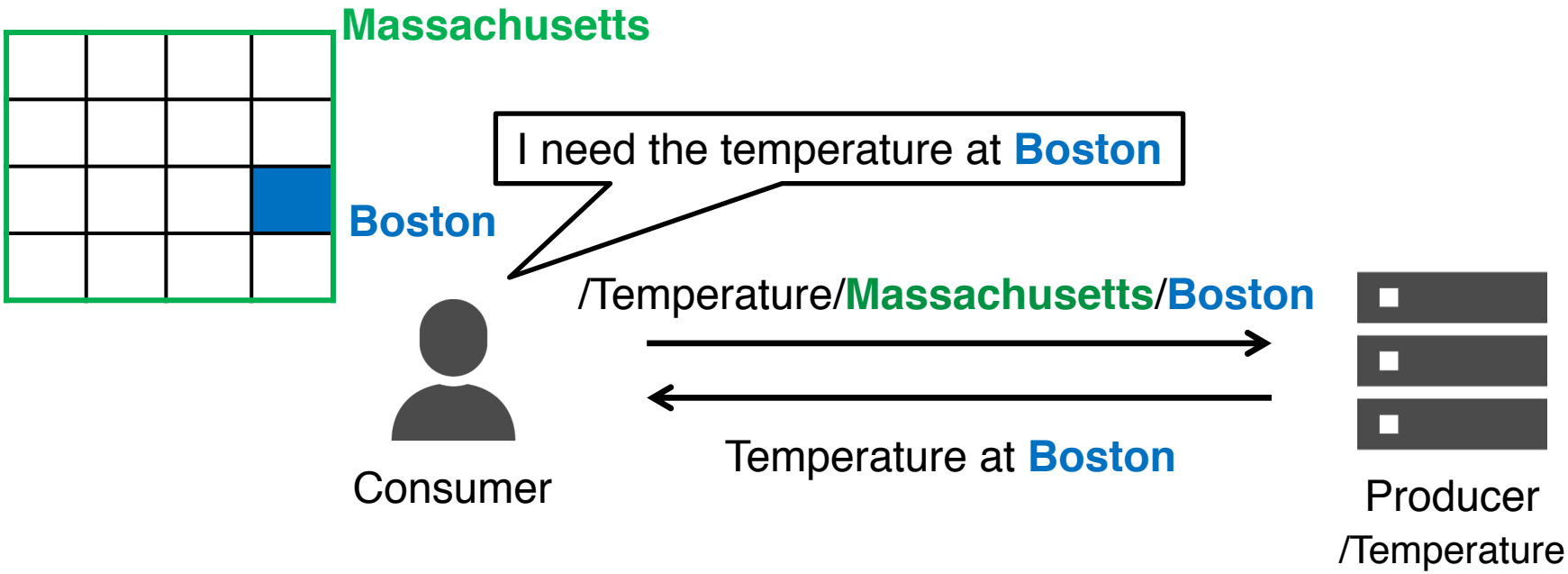
Location Privacy Protection with a Semi-honest Anonymizer in Information Centric Networking

Kentaro Kita, Yoshiki Kurihara,
Yuki Koizumi and Toru Hasegawa

Graduate School of Information and Technology,
Osaka University, Japan

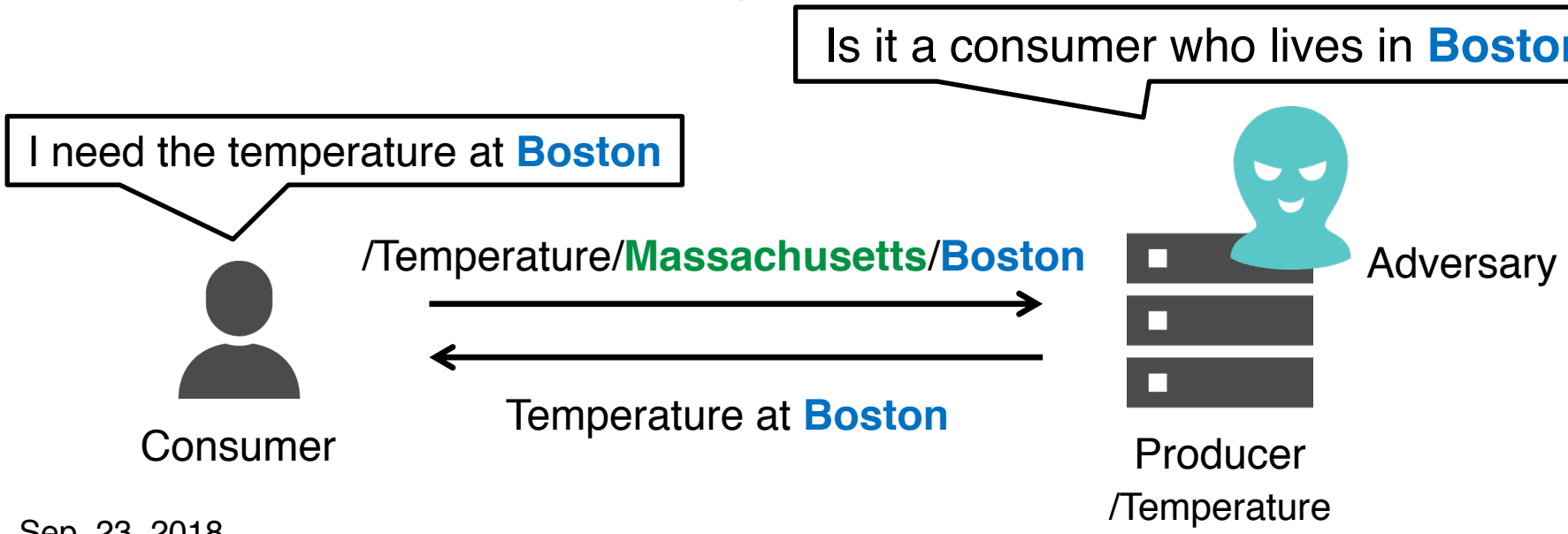
Location-Based Services

- System model of LBSs
  - Consumers choose locations of their interests (target locations) from a set of locations where producers offers its services (service area) and send names of the locations to the producers
  - Producers return data based on the locations



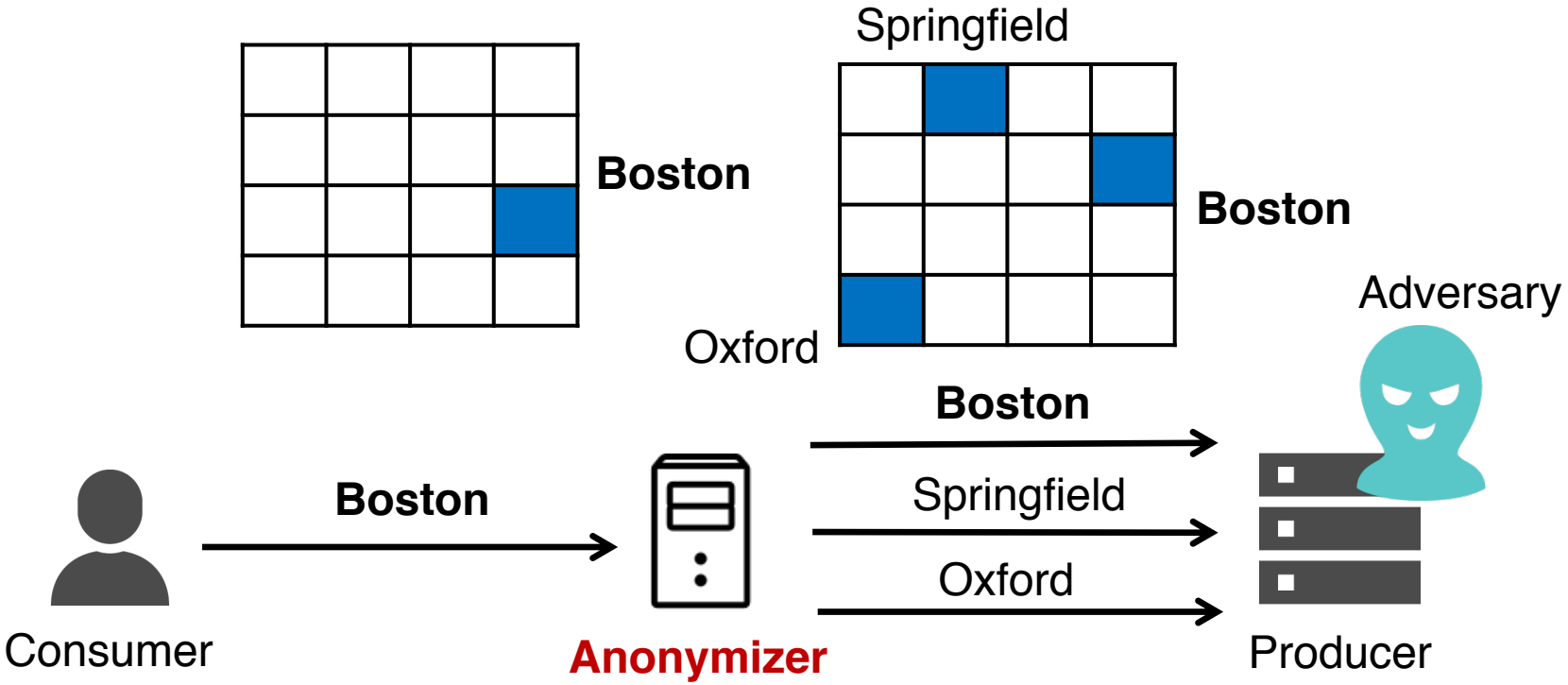
Privacy in LBSs

- Goal : **Location Privacy**
  - Hiding consumers' target locations from adversaries (including producers) in LBSs
- Privacy Problem in LBSs
  - Consumers' target locations can easily be linked to their sensitive information
    - home locations, life styles



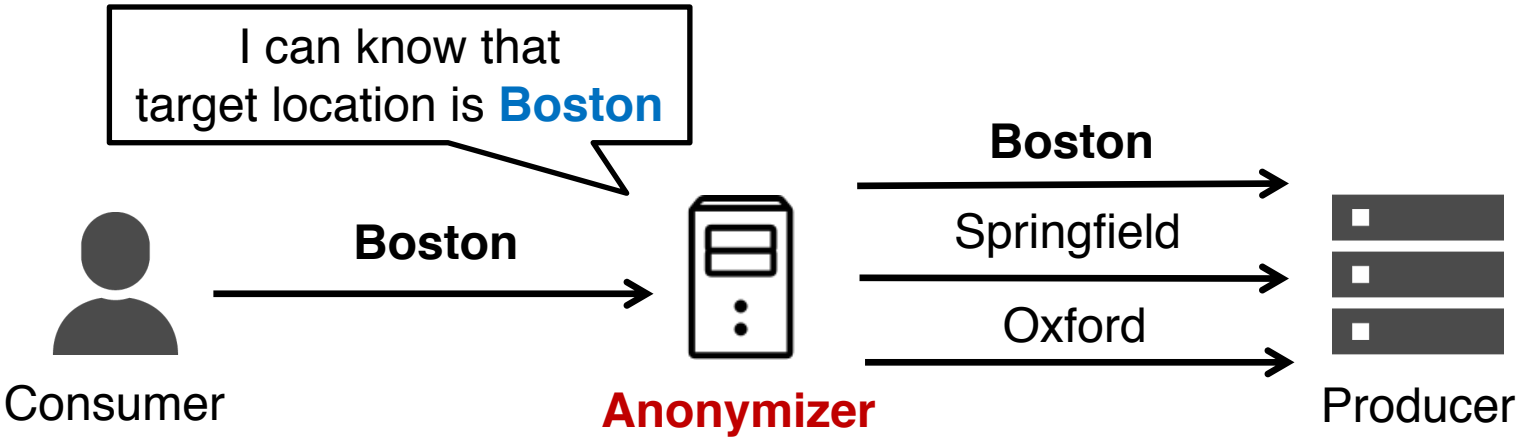
Existing Approaches

- Honest anonymizer to achieve location anonymity
  - Hiding each consumer's target location into other $k-1$ dummy locations to achieve k-anonymity of locations
    - Anonymous location set : a set of k locations which includes a consumer's target location
  - An anonymizer generates anonymous location sets from consumers' requests about their target locations



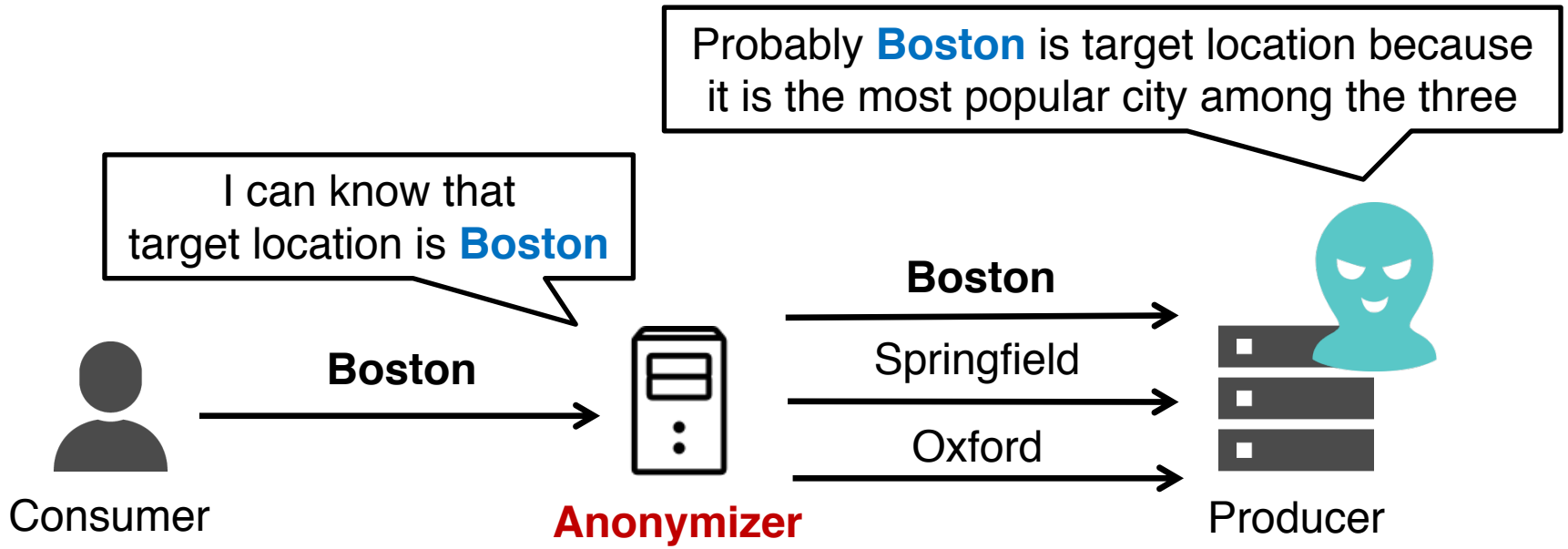
Problems #1 in Existing Approaches

1. The anonymizer can identify consumers' target locations
   - Hence, the anonymizer must be honest (trusted third party)



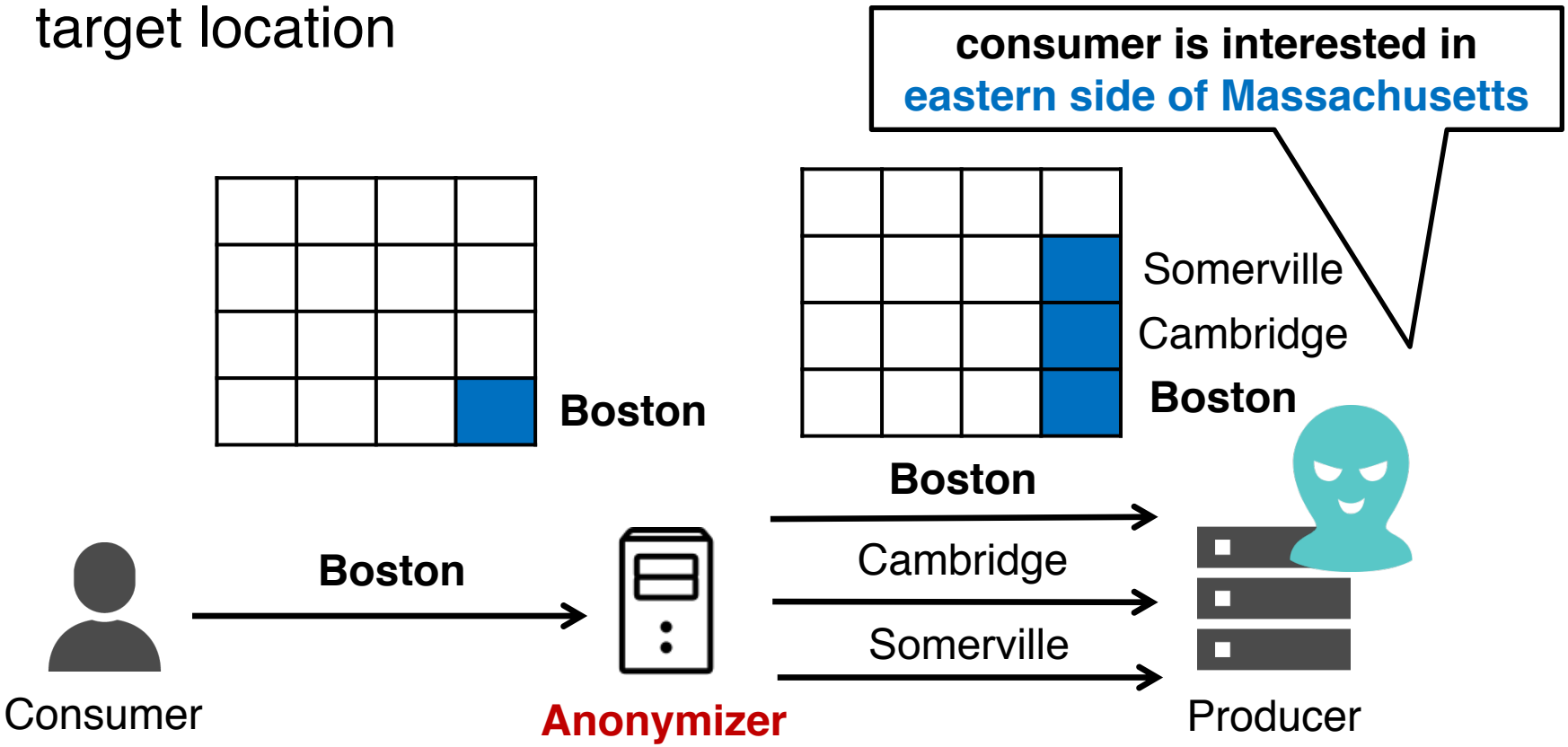
Problems #2 in Existing Approaches

1. The anonymizer can identify consumers' target locations
   - Hence, the anonymizer must be honest (trusted third party)
2. Adversaries can infer target locations from anonymous location sets by leveraging popularities of locations



Problems #2 in Existing Approaches

1. The anonymizer can identify consumers' target locations
   - Hence, the anonymizer must be honest (trusted third party)
2. Adversaries can narrow target location to a region with a certain degree of accuracy even if they cannot infer target location



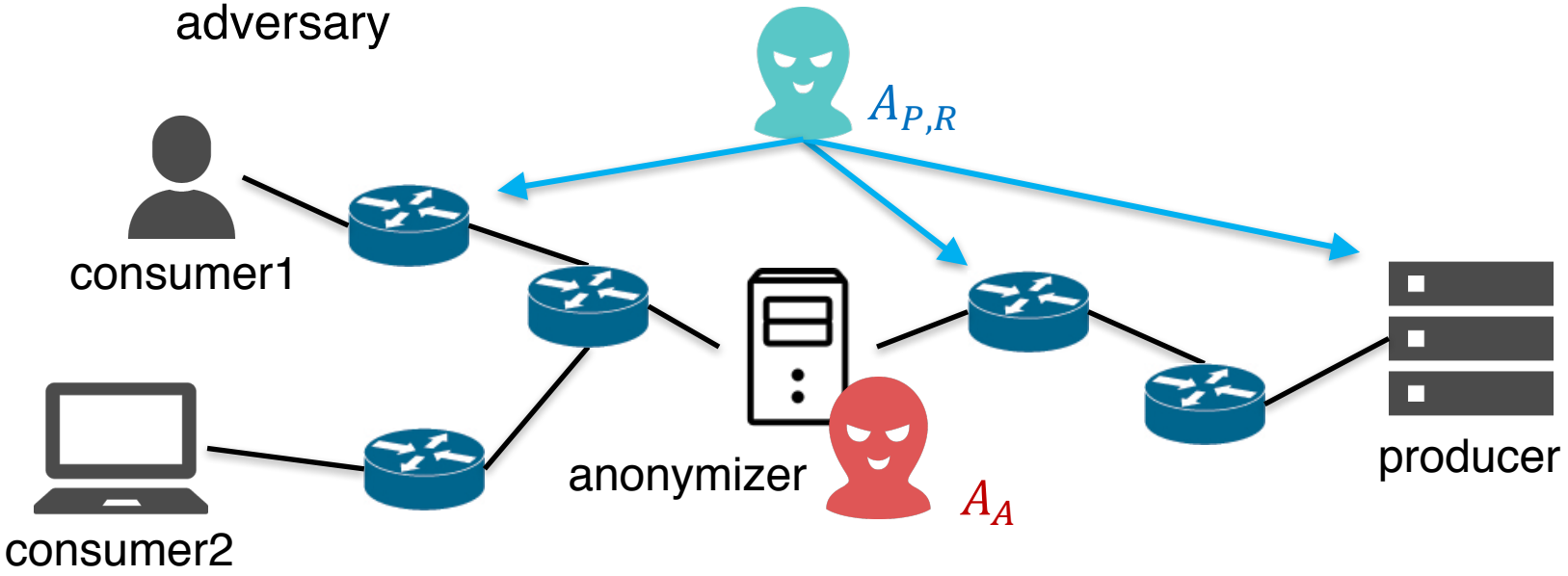
Challenges

1. Semi-honest anonymizer
   - Designing a semi-honest anonymizer in NDN
     - An semi-honest entity **follows prescribed protocols** but attempts to gain more information than allowed from the protocols, and **does not collude** with others to launch attacks

2. Dummy locations selection
   - Rigorously defining location anonymity satisfying the following two requirements
     1. Preventing adversaries from **probabilistically inferring** target locations
     2. Minimizing **geographical information** of target locations leaked to adversaries

Adversarial Model

- Two semi-honest adversaries who attempt to infer target locations from received/eavesdropped packets
  - $A_{P,R}$: An adversary on some producers and networks
    - An adversary on producers as well as on routers should be considered
  - A_A : An adversary on the anonymizer
    - Unlike existing studies, we assume that the anonymizer is also an adversary



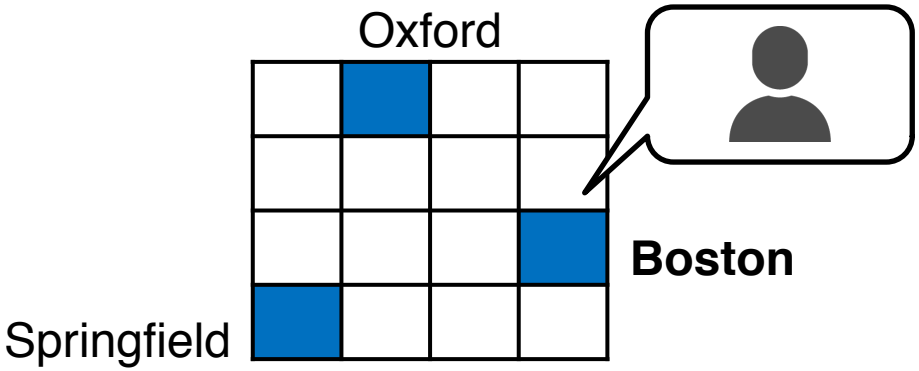
Location Privacy

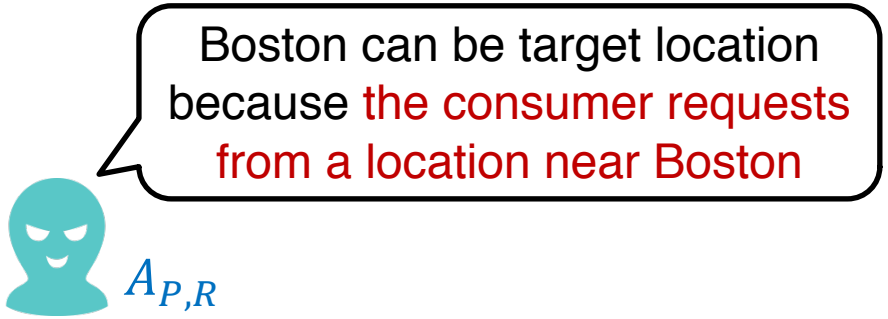
- Is it sufficient to achieve location anonymity to protect location privacy as in existing approaches?

- Location Privacy
  = location anonymity + session anonymity
  - Session anonymity ensures indistinguishability of consumers
    - Adversaries cannot gain information about consumers
      - Who is the consumers
      - Whether two requests are from the same consumer or not

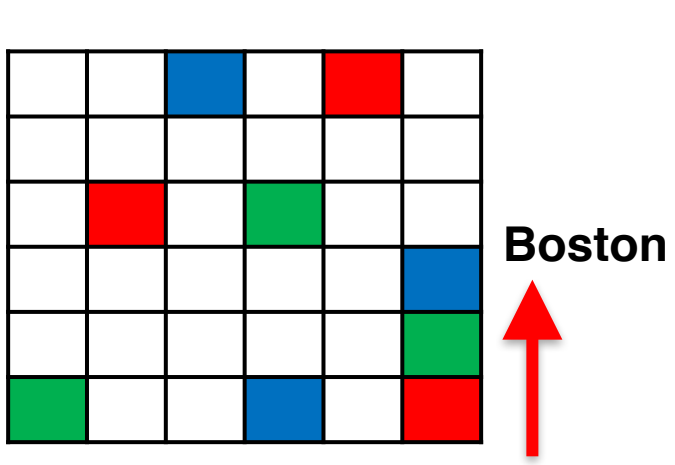
Necessity to Achieve Session Anonymity

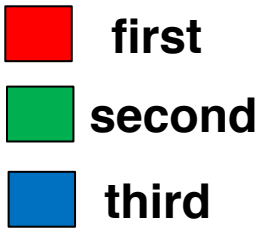
- Auxiliary information about consumers breaks k-anonymity
  - Adversaries can Infer target location based on the possibility that the consumer chooses each location as target location

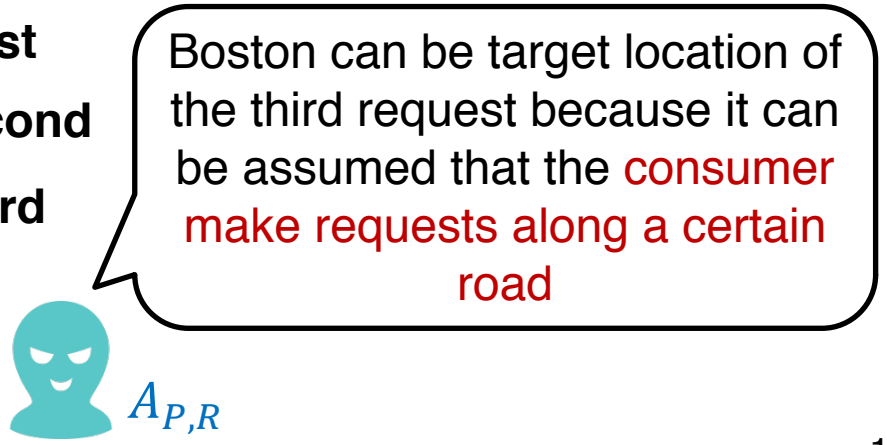




  - Adversaries can Infer target location based on the past anonymous location sets of the consumer







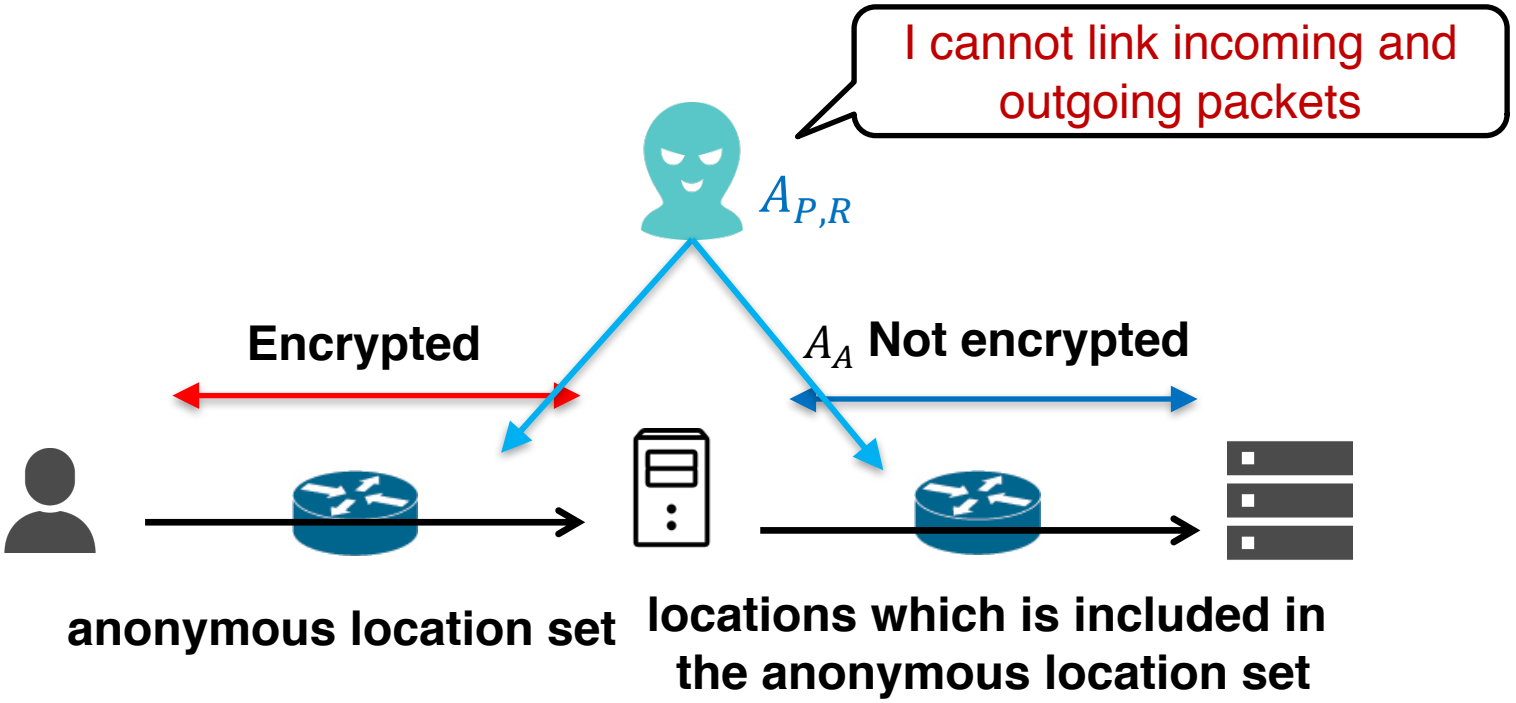
Design Rationale of Architecture

- Solution to achieve location anonymity
  - Each consumer makes request specifying an anonymous location set to the anonymizer instead of target location
    - The anonymizer generates a map of anonymous location sets for all the locations and distribute it to consumers

- Solution to achieve session anonymity
  - We leverage lack of source/destination addresses on packets in NDN (against A_A)
    - Interest and Data packets do not convey any information about consumer.
  - The anonymizer also works as a mix-router (against $A_{P,R}$)

Anonymizer as a mix-router

- Session anonymity against $A_{P,R}$
  - The Anonymizer acts as a Chum's mix router to prevent $A_{P,R}$ from link incoming and outgoing packets at the anonymizer
    - Encryption/decryption at the anonymizer
    - Batching N incoming packets
    - Sometimes make dummy requests



Requirements to Location Anonymity

1. Preventing adversaries from **probabilistically inferring** target locations
   - Location k-anonymity
     - Adversaries cannot infer a consumer's target location l_T from her/his anonymous location set $\mathcal{L}$

       $$P[l_T = l_i \mid \mathcal{L}] = P\left[l_T = l_j \mid \mathcal{L}\right] (\forall l_i, l_j \in \mathcal{L})$$

2. Minimizing **geographical information** of target locations leaked to adversaries
   - Location t-closeness
     - Each anonymous location set $\mathcal{L}$ is scattered uniformly throughout the service area S

       $$D[\mathcal{L}, S] \leq t$$

       ,where $D[\cdot,\cdot]$ is the difference between two geographical distributions

Requirement #1 to Location Anonymity

- Location k-anonymity
  - Adversaries cannot infer a consumer's target location l_T from her/his anonymous location set $\mathcal{L}$

$$P[l_T = l_i \mid \mathcal{L}] = P\left[l_T = l_j \mid \mathcal{L}\right] (\forall l_i, l_j \in \mathcal{L})$$

  - $P[l \mid \mathcal{L}] = P[\mathcal{L} \mid l]P[l]/P[\mathcal{L}]$ (Bayes' theorem)

The probability that $\mathcal{L}$ is used under the condition that target location is l

the probability that l is selected as a target location (popularity)

  - We should take these two factors into account to generate anonymous location sets

Solution #1 to Location Anonymity

- Making disjoint anonymous location set
  - If we divide the service area into disjoint anonymous location sets, the anonymous location set for each target location is deterministically determined
    - $\forall l \in \mathcal{L},\ P[\mathcal{L}|l] = 1$ and $\forall l \notin \mathcal{L},\ P[\mathcal{L}|l] = 0$

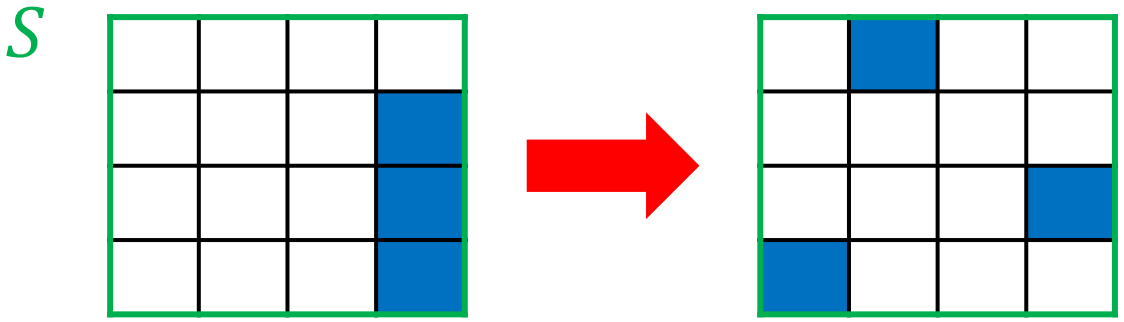
  $\rightarrow P[l \mid \mathcal{L}] = P[\mathcal{L} \mid l]P[l]/P[\mathcal{L}] = P[l]/P[\mathcal{L}]$
- Maximizing entropy of popularities of locations
  - $H_{\mathcal{L}} = -\sum_{l \in \mathcal{L}} p_{l,\mathcal{L}} * \log_2 p_{l,\mathcal{L}}$
    - where $p_{l,\mathcal{L}} = P[l]/\sum_{l_i \in \mathcal{L}} P[l_i]$ (normalized popularity)
  - Selecting k locations so that their popularities $P[l]$ are as close as possible
    - We evaluate later

Requirement #2 to Location Anonymity

- Location t-closeness
  - Each anonymous location set $\mathcal{L}$ is scattered uniformly throughout the service area S

    $$D[\mathcal{L}, S] \leq t$$

    where $D[\cdot,\cdot]$ is the difference between two distributions
  - Motivation to achieve location t-closeness
    - If all the locations in an anonymous location set is close, adversaries can narrow target location to a region with a certain degree of accuracy even if they cannot infer target location



  - Solution
    - Combining sufficiently scattered locations to generate anonymous location sets

Anonymous Location Sets Generation

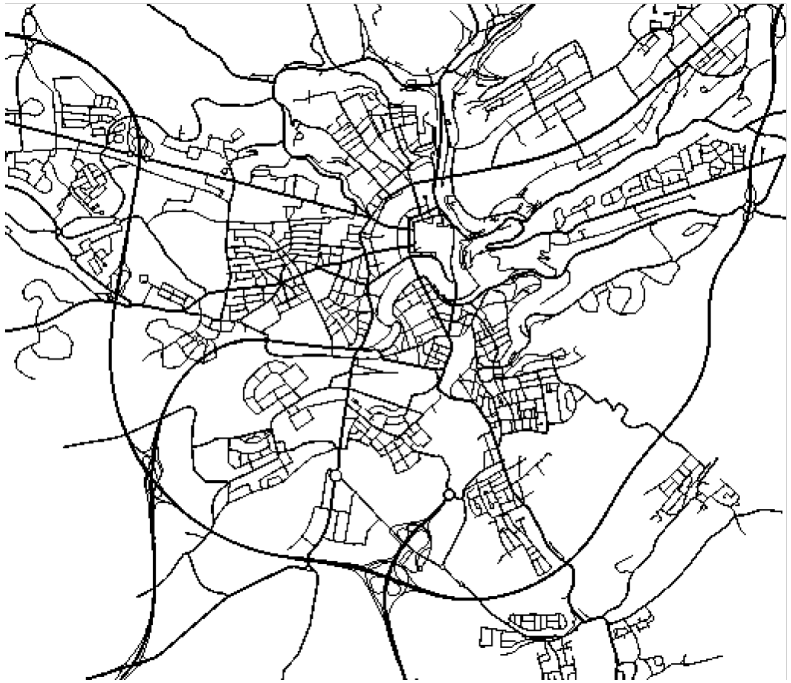
- Overview of our algorithm to generate anonymous location sets
  1. Dividing the service area into k segments
     - k is degree of k-anonymity
     - Each segment consists of neighboring locations
  2. Selecting a location from each segment according to the popularities and combine those k locations
     - Locations with similar popularities that are located far enough can be combined.
     - Anonymous location sets become disjoint

Evaluation of Anonymous Location Sets

- Measurements
  - Entropy of popularities of locations in each anonymous location set
  - Ratio of size of the range covered by each anonymous location set with respect to that of a service area
- Conditions
  - An LBS which collect speed of vehicles in each location
  - Use SUMO simulator to obtain vehicle movements
  - The service area is approximately 60 km^2 and is divided into 1024 locations
  - anonymity degree $k = 15$

Generated Anonymous Location Sets

- Examples of anonymous location sets
  - A set of locations painted with the same color is one anonymous location set.



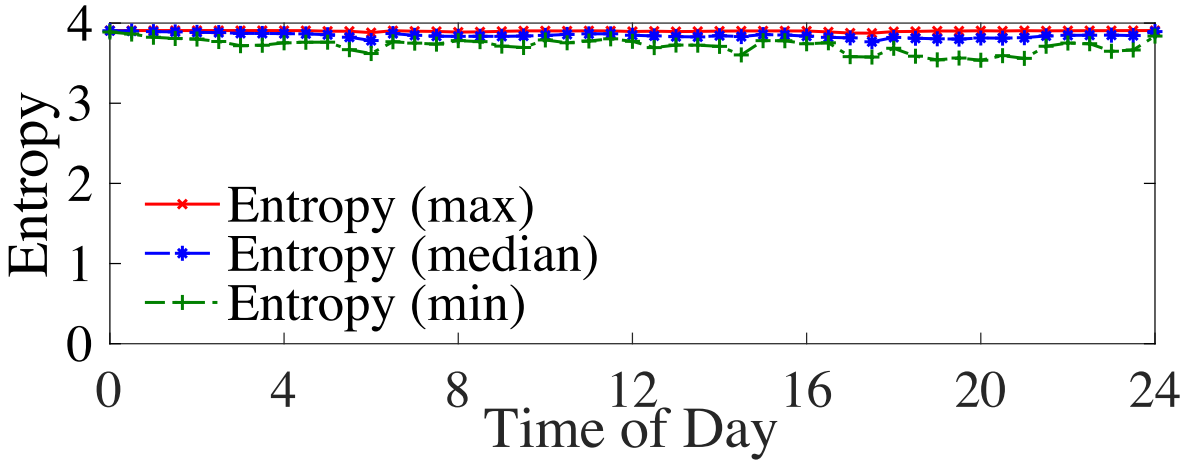
Service are

19	40	-1	-1	-1	32	46	-1	-1	-1	-1	-1	17	51	46	50	-1	51	33	19	17	46	-1	-1	-1	49	21	16	29	20	42	6
16	48	-1	-1	-1	-1	35	37	-1	-1	-1	-1	2	-1	42	-1	-1	-1	23	16	32	49	47	43	-1	15	36	41	10	37	46	4
3	7	9	-1	-1	-1	-1	41	28	-1	-1	47	21	1	4	10	50	24	8	28	48	44	41	26	9	48	22	12	26	39	40	7
20	49	8	5	-1	-1	-1	-1	42	23	15	16	32	5	35	12	14	4	13	29	22	30	20	10	51	25	24	30	50	45	51	14
30	38	50	45	6	-1	-1	44	40	47	3	29	48	44	41	14	45	3	18	38	32	17	33	43	50	21	23	45	26	49	16	24
12	39	27	34	13	11	51	43	36	37	49	7	28	34	24	11	39	1	21	42	14	47	35	44	37	32	20	33	47	34	12	39
18	25	26	29	21	4	24	33	26	45	25	6	43	20	18	30	25	3	2	28	34	38	19	8	22	35	25	46	36	31	13	17
10	31	17	14	2	1	22	36	23	13	38	19	15	33	9	8	13	11	1	5	18	23	27	31	30	44	41	27	29	43	10	28
39	27	44	35	28	4	21	13	22	2	34	42	48	27	40	23	11	15	1	-1	48	19	43	44	50	38	32	42	41	48	1	20
22	31	27	29	40	8	51	38	39	32	49	29	33	44	15	18	18	8	3	40	36	21	47	40	30	33	17	23	25	10	3	38
25	23	31	20	24	5	33	37	45	24	37	21	46	47	12	3	9	42	4	2	35	18	22	26	27	22	29	51	16	12	11	-1
22	16	10	36	26	7	39	-1	-1	16	38	17	36	25	7	28	6	38	5	7	15	39	31	41	31	-1	24	14	28	5	-1	-1
47	32	12	19	9	43	46	45	-1	26	31	14	9	6	31	36	3	24	5	12	6	19	25	46	15	13	18	37	26	6	-1	-1
-1	48	30	11	2	18	42	34	41	6	20	1	12	27	15	37	48	1	7	8	13	37	49	23	19	36	33	-1	7	43	-1	-1
49	50	41	15	1	3	13	4	9	30	19	10	11	7	34	35	28	2	4	10	20	34	27	45	40	35	44	-1	8	30	-1	-1
-1	14	17	-1	6	51	50	43	5	8	11	35	-1	5	2	40	11	29	16	9	17	14	-1	51	-1	-1	-1	-1	9	34	-1	-1
21	32	45	50	4	-1	45	35	21	19	11	17	16	7	1	3	19	21	23	15	7	24	43	48	-1	-1	-1	-1	20	-1	-1	-1
42	39	47	-1	2	51	49	33	29	-1	47	37	41	34	10	39	38	10	12	11	3	24	14	26	-1	-1	-1	-1	13	-1	-1	-1
49	46	-1	-1	8	-1	-1	28	-1	-1	40	38	22	20	14	24	29	20	18	6	12	33	40	42	16	18	-1	34	25	-1	-1	-1
-1	-1	-1	-1	9	31	30	32	43	-1	36	23	26	42	13	25	14	51	-1	17	22	31	37	41	-1	30	27	10	-1	-1	-1	-1
-1	-1	-1	-1	44	5	18	-1	46	45	41	18	39	37	15	16	15	45	-1	39	4	32	44	-1	-1	-1	17	21	23	50	-1	-1
-1	-1	50	46	48	6	12	-1	38	44	34	27	25	17	10	20	36	19	29	28	6	46	50	-1	-1	38	23	47	22	49	-1	-1
-1	-1	-1	-1	27	2	-1	42	47	23	31	28	36	32	-1	49	-1	11	3	1	7	51	-1	-1	15	51	38	19	29	42	-1	-1
-1	-1	-1	-1	-1	-1	4	15	48	19	24	35	-1	43	26	40	-1	-1	9	8	5	2	35	-1	40	14	46	30	27	36	32	45
-1	-1	-1	-1	-1	-1	3	1	5	30	-1	-1	49	27	46	32	35	18	37	31	12	34	1	-1	12	-1	51	19	-1	25	45	-1
-1	-1	-1	-1	-1	-1	6	51	4	2	22	-1	48	26	36	39	11	16	33	41	43	39	4	10	14	51	33	35	34	42	-1	-1
-1	-1	-1	-1	-1	-1	8	49	14	-1	13	8	4	5	2	1	2	26	17	-1	48	47	24	3	-1	49	22	44	32	-1	-1	-1
-1	-1	-1	-1	-1	-1	12	50	31	-1	40	28	47	35	43	9	8	7	25	20	28	13	21	9	-1	37	24	28	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	9	-1	45	16	44	-1	-1	34	-1	22	5	44	11	9	7	38	39	3	4	8	21	50	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	13	-1	25	41	-1	-1	-1	30	-1	-1	6	-1	-1	-1	26	5	15	13	29	10	18	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	11	21	42	-1	-1	-1	37	50	-1	-1	1	-1	-1	-1	48	43	31	23	-1	16	17	40	46	51	51	-1
-1	-1	-1	-1	-1	29	7	33	-1	-1	-1	-1	33	-1	-1	-1	2	6	-1	-1	-1	-1	36	20	47	27	30	41	-1	-1	-1	-1

Anonymous location sets

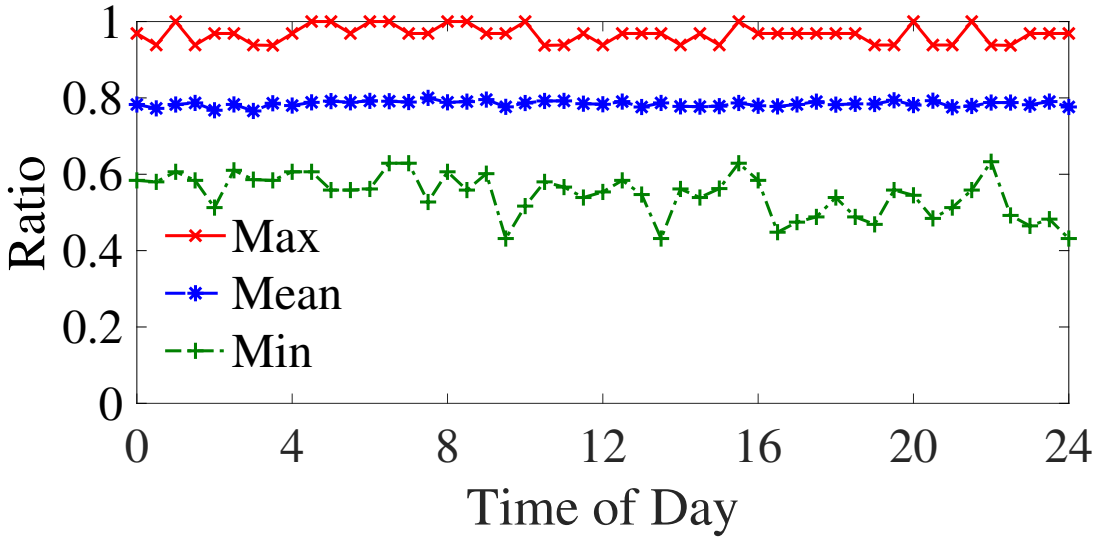
Result #1

- Location k-anonymity : Location Entropy of the popularities of locations
  - The larger the entropy, the smaller the differences in popularities of locations and adversaries cannot infer target locations.
    - optimal value is $H_{\mathcal{L}} = -\log_2\left(\frac{1}{k}\right)$ =3.91
  - Observation
    - Our algorithm generates good anonymous location sets because the entropy is sufficiently close to the optimal value



Result #2

- Location t-closeness : Ratio of size of the range covered by each anonymous location set with respect to that of the service area
  - The greater the ratio is, the more adversaries cannot gain geographical information of target locations
  - Observation
    - Even the worst anonymous location set covers a sufficiently large area of the service area.



Conclusions

- Conclusions
  - We define location privacy as a combination of location anonymity and session anonymity
  - We propose an architecture to achieve session anonymity under the adversarial model that none of the anonymizer, producers, and networks are honest
  - We propose an anonymous location sets generation algorithm to achieve location anonymity which is defined using k-anonymity and t-closeness