A Mathematical Theory of Location k-Anonymity (2025)

Ole Kristian Aamot (2025)

Abstract

We develop a rigorous mathematical framework for **location k-anonymity**, extending the classical notion of k-anonymity to continuous spatial domains. Unlike traditional formulations, which rely on dataset-based indistinguishability, we define anonymity in terms of **metric spaces, probability measures, and adversarial inference**. We show that location k-anonymity can be expressed as a constraint on posterior identification probability and characterize the trade-off between privacy and utility. Limitations of deterministic anonymity are also formally derived.

---

1. Introduction

Location-based services rely on precise spatial data, yet such data is inherently identifying. Classical k-anonymity ensures that each individual is indistinguishable among at least $k$records , but this definition is insufficient in continuous spatial settings where:

- location is not discrete,
- adversaries possess auxiliary knowledge,
- and trajectories reveal identity over time.

This paper introduces a **mathematical theory** to address these limitations.

---

2. Preliminaries

2.1 Spatial model

Let:

- (X, d)be a metric space representing geographic locations,
- Ube a finite set of users,
- $\ell: U \to X$map users to locations.

2.2 Classical k-anonymity

A dataset satisfies k-anonymity if each record is indistinguishable from at least $k-1$others .

We reinterpret this in spatial terms.

---

3. Location k-Anonymity

Definition 1 (Spatial k-anonymity)

A mechanism satisfies location k-anonymity if for every reported region $R \subseteq X$:

$$|\{u \in U : \ell(u) \in R\}| \geq k$$

This corresponds to **spatial cloaking**, where precise coordinates are replaced by regions.

---

4. Probabilistic Formulation

Deterministic definitions are insufficient under adversarial knowledge.

Definition 2 (Probabilistic k-anonymity)

Let Abe an adversary with prior distribution $P(u)$. After observing output R, define posterior:

$$P(u \mid R)$$

The system satisfies probabilistic k-anonymity if:

$$\max_u P(u \mid R) \leq \frac{1}{k}$$

This reframes anonymity as a **bound on re-identification probability**.

---

5. Utility–Privacy Trade-off

Let:

- $\text{diam}(R)$denote region size,
- ϵdenote service error.

We define:

$$\epsilon \propto \text{diam}(R)$$

Thus:

- Increasing k→ larger R→ higher privacy
- But also → reduced utility

This formalizes the trade-off observed in anonymization systems.

---

6. Adversarial Models

We consider three adversary classes:

6.1 Snapshot adversary

Observes a single query.

6.2 Background adversary

Has auxiliary data (e.g., home/work locations).

6.3 Trajectory adversary

Observes sequences:

$$(\ell_t(u))_{t=1}^{T}$$

We show that trajectory uniqueness drastically reduces anonymity, even when each point individually satisfies k-anonymity.

---

7. Impossibility Results

Theorem 1

No deterministic mechanism can guarantee k-anonymity under arbitrary auxiliary information.

Sketch:
If an adversary knows one user lies in region R, then:

$$P(u \mid R) = 1$$

Thus anonymity collapses.

Theorem 2 (Dimensionality curse)

As dimensionality (e.g., time + space) increases, the minimum region size required for k-anonymity grows exponentially.

---

8. Extensions

8.1 Continuous k-anonymity

Define density function $f(x)$:

$$\int_R f(x)\, dx \geq k$$

8.2 Hybrid privacy models

We combine:

- k-anonymity
- differential privacy

This follows trends in hybrid models that strengthen guarantees beyond classical approaches .

---

9. Discussion

Strengths

- Simple, interpretable privacy guarantee
- Compatible with spatial cloaking systems

Weaknesses

- Vulnerable to background knowledge
- Poor performance in high-dimensional data
- Deterministic nature enables inference attacks

---

10. Conclusion

We have presented a unified mathematical theory of location k-anonymity that:

- embeds anonymity in metric and probabilistic frameworks
- formalizes adversarial inference
- proves fundamental limitations

Future work should focus on:

- trajectory-aware privacy
- integration with stochastic privacy mechanisms
- real-world deployment constraints