Proceedings of the 8th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/2026.35334

Privacy Preservation in Data Streams: An Efficient Query Algorithm Based on CGP

Shiwen Xu

School of Computer Science and Information Engineering, Chuzhou University, Chuzhou, China
2118910038@qq.com

Abstract. There are serious problems of privacy protection for continuous spatial data streams in the edge computing scenario. Traditional Concentrated Geo-Privacy (CGP) mechanisms and the existing adaptive methods both have a utility-efficiency bottleneck. Due to high concurrency and resource limitations in the environment, the current model has a large computational cost, and local spatial pruning methods often overlook the risk of cardinality leakage to deduce the true density. In response to the aforementioned weaknesses, this paper puts forward a lightweight collaborative geographic privacy preservation mechanism for dynamic data streams called Pruning-CGP. First, a physically separated query architecture will separate data ingestion from real-time queries to reduce computation. Second, in the zero-concentrated Differential Privacy (zCDP) setting, a discrete Gaussian mechanism is used for cardinality-preserving noise addition. In conjunction with the public prior, an adaptive distribution of the privacy budget is realised. Extensive physical stress tests with actual data have shown that the engineering feasibility of the proposed architecture has been realised. Under the above-mentioned privacy-equivalent constraints, the system has blocked side-channel cardinality leakage. Although it has a relatively low recall rate of about 31.5% and keeps the distance error at a practical hundred-meter level (e.g., 62 meters), it also achieves an extremely fast sub-millisecond response speed of 0.16 ms. Based on the above results, the Pruning-CGP mechanism can be applied to address the demand for a high-performance, low-resource algorithm in the context of edge computing.

Keywords: Zero-Concentrated Differential Privacy, Matrix Vectorization, Dynamic Data Streams, Edge Computing

1. Introduction

In recent years, with the spread of the Internet of Things (IoT) and edge computing architecture, lightweight data processing protocols have received much attention [1]. A large number of mobile phones generate large volumes of continuous spatial trajectory data streams. Although the above high-frequency trajectory data can be used for decision-making by the intelligent transportation system, it is also very susceptible to geographical privacy leakage [2]. To prevent background knowledge inference attacks, Differential Privacy (DP) and its derivative model, Concentrated Geo-

DOI: 10.54254/2755-2721/2026.35334

Privacy (CGP), have recently become popular technologies because of their strict mathematical theoretical lower bounds [3].

However, traditional privacy protection methods have significant deficiencies in the context of resource-constrained, high-dynamism edge data-stream processing. Unbounded data streams, in the event of continuous queries, may quickly exhaust the privacy budget [4]. Although the CGP framework has been introduced to reduce the budget overrun of long-term composition, most of the existing models are in a serial architecture, first adding global noise and then performing spatial queries [3]. The static global strategy has serious deficiencies: it will be computationally expensive at a large scale and easily exceed the computing capacity of edge nodes. Some adaptive privacy protection mechanisms aim to optimise the local perturbation of stream data [5]. However, although the above strategies reduce the amount of candidate sets that need to be processed, they do not strictly bound the cardinality of these sets. Attackers can infer the actual data density of the corresponding grid area based on the size of the pruned candidate set, thus creating a serious side-channel cardinality leakage problem.

In light of the above utility-efficiency bottlenecks and theoretical deficiencies, this paper introduces a high-efficiency collaborative query architecture called Pruning-CGP for data stream privacy protection. The basic idea is to address the efficiency limitation of traditional serial mechanisms by reducing the enormous space of global perturbations to a local finite neighbourhood through lightweight spatial index pruning, and to prevent cardinality leakage. The first few are shown in the following list.

•A physically decoupled $\mathcal{O}(1)$ query architecture: Overcoming the global traversal bottleneck of traditional static mechanisms, the construction of the spatial inverted index is physically decoupled from real-time privacy queries. Combined with underlying full-matrix vectorization technology, the actual retrieval response time is successfully compressed to an ultra-fast sub-millisecond level (0.16 ms).

A cardinality security masking mechanism based on z-CDP is proposed to address the hidden danger of side-channel leakage caused by local pruning, and an innovative discrete Gaussian mechanism is introduced within the unified framework of zero-Concentrated Differential Privacy (z-CDP). It can effectively prevent the density-based cardinality leakage vulnerability and has rigorous guarantees in theory and practice.

•A Practical Privacy-Utility Trade-off Analysis: Rather than using the ideal and impractical assumptions of previous studies, this paper carries out all-weather physical stress tests. The results show that, under strict zCDP equivalent constraints, a practical hundred-meter level distance error and an acceptable core recall rate (about 31.5%) can be achieved; at the same time, high security and excellent computational availability in resource-constrained edge devices are maintained.

2. Related work

2.1. Traditional spatial privacy and local differential privacy mechanisms

Early location privacy protection mainly relied on heuristic methods such as spatial cloaking. To overcome their theoretical defects, differential privacy was widely introduced. Efficient methods based on Local Differential Privacy (LDP) have been proposed for spatial range queries [6]. For stream data collection, enhanced LDP collection mechanisms for streaming data were proposed to improve statistical utility [7]. Targeting dynamic evolution features, publication methods based on dynamic sliding windows were designed [8]. However, most of the above works focus on the

DOI: 10.54254/2755-2721/2026.35334

statistical feature distribution of data and do not deeply optimize the computing power bottlenecks for spatial Top- k continuous retrieval tasks.

2.2. Concentrated Geo-privacy and spatial perturbation mechanisms

To address the problem of budget explosion in multi-combination queries for high-frequency stream data, the foundation for Concentrated Differential Privacy (CDP) and zCDP has been laid [9, 10]. Based on the above, the CGP framework was put forward [3]. Recently, approximation algorithms for online evolving data streams [4] and data-adaptive privacy-preserving mechanisms [5] have been studied. However, to enhance efficiency, most of the existing local perturbation mechanisms universally adopt a spatial partitioning and candidate set extraction strategy but fail to formally mask the cardinality of the extracted sets. There is a serious side-channel vulnerability that allows attackers to reverse-engineer the regional densities. The Pruning-CGP mechanism introduced in this paper effectively addresses the aforementioned problem by adding a discrete Gaussian mechanism under the zCDP framework and giving explicit priority to cardinality security over idealised utility.

3. Preliminaries and system model

3.1. Zero-concentrated differential privacy and mapping

Handling multiple compositions of high-frequency continuous stream data requires tight privacy bounds [11, 12]. This paper adopts zero-Concentrated Differential Privacy (z-CDP). Given a randomized mechanism $\mathscr{M}$ and any adjacent datasets $D, D^{'}$, $\mathscr{M}$ satisfies ρ -zCDP if for all $\alpha \in (1, \infty)$, its α -Rényi divergence satisfies $D_{\alpha}\left(\mathscr{M}(D) \middle|\middle| \mathscr{M}\left(D^{'}\right)\right) \leq \rho\alpha$. Based on the linear composition theorem, sequential mechanisms satisfying ρ_1 -zCDP and ρ_2 -zCDP satisfy $(\rho_1 + \rho_2)$ -zCDP. A mechanism satisfying ρ -zCDP can be losslessly mapped to the traditional geographic indistinguishability lower bound [14] via $\epsilon = \rho + 2\sqrt{\rho ln\,(1/\delta)}$.

3.2. Discrete and continuous Gaussian mechanisms

For continuous spatial coordinate perturbation, injecting Gaussian noise $\mathcal{N}\left(0, \sigma^2\right)$ with variance $\sigma^2 = \left(\Delta f\right)^2 / \left(2\rho\right)$ satisfies ρ -zCDP. However, regarding the number of trajectory points within a grid (i.e., cardinality), traditional continuous noise addition with rounding breaks strict privacy proofs. To explicitly defend against side-channel cardinality leakage in dynamic publications [13], this mechanism introduces the discrete Gaussian mechanism $\mathcal{N}_{\mathbb{Z}}\left(0, \sigma^2\right)$ defined over the integer domain. For any integer $x \in \mathbb{Z}$, its output probability is proportional to exp $\left(-x^2 / 2\sigma^2\right)$.

3.3. Dynamic data streams model

Continuously record trajectory information at the edge nodes as an unbounded data stream. Most query algorithms are stored in edge memory to meet the real-time requirements of a very dynamic environment.

DOI: 10.54254/2755-2721/2026.35334

4. Core privacy query mechanism design

To address the problem of low traversal efficiency and spatial drift in the entire graph, a high-performance collaborative query architecture (Pruning-CGP) is presented here, as shown in Figure 1.

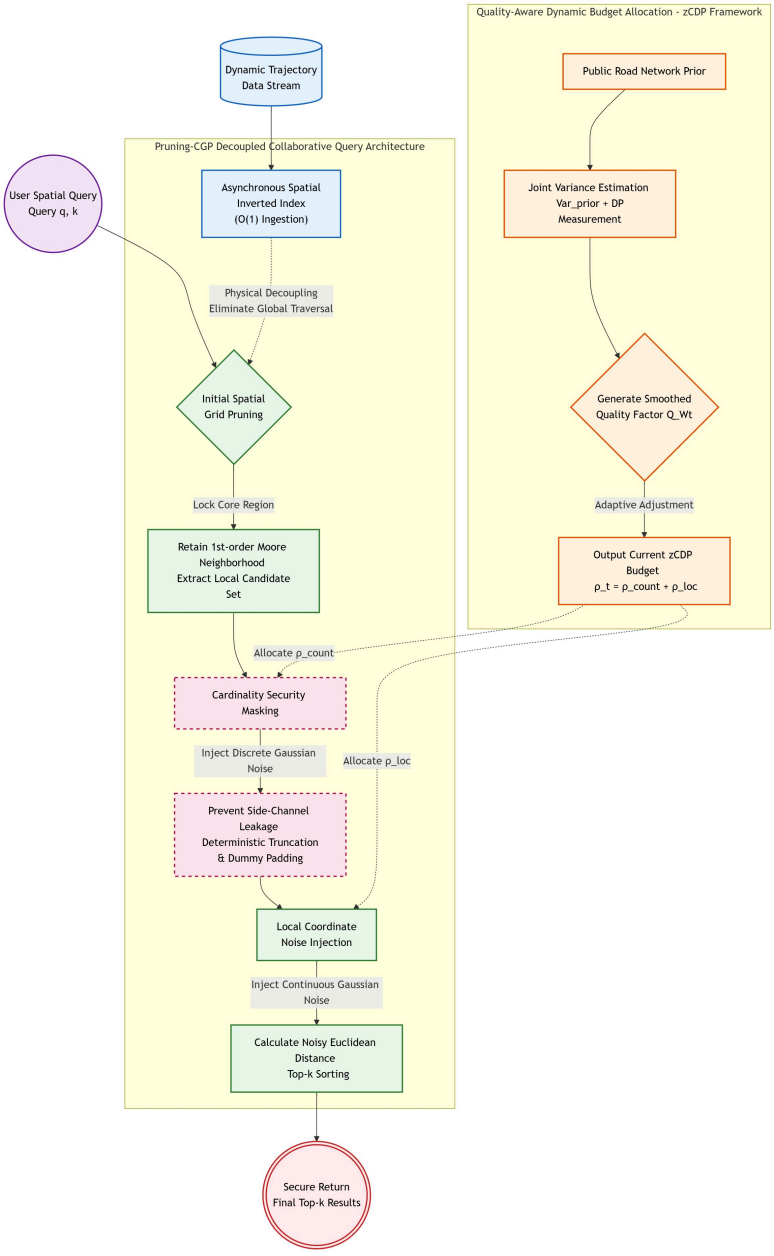

Figure 1. General architecture of Pruning-CGP

4.1. Data ingestion and O(1) spatial index construction

A public absolute geographic coordinate system is divided into grids with side length l. During the data ingestion phase, when a trajectory point p_i enters the active window, its pointer is appended to a hash table index map. This process strictly bounds the single-point insertion complexity to

DOI: 10.54254/2755-2721/2026.35334

amortized $\mathcal{O}(1)$. This architectural design physically decouples index construction from real-time privacy queries, overcoming the severe computational overhead of static global traversal paradigms.

4.2. Secure pruning and noise sink collaborative query

Upon receiving a query point q , the algorithm utilizes the public grid coordinate system to extract a localized candidate set within a 1-hop Moore neighborhood. To enforce cardinality security, the total privacy budget ρ_t is split into ρ_{count} and ρ_{loc} .

First, the algorithm forces integer masking on the candidate set cardinality using the discrete Gaussian mechanism satisfying ρ_{count} -zCDP. If the noisy count N_{noisy} is less than the true count, the set is deterministically truncated; otherwise, dummy coordinates are padded to match N_{noisy} . This step rigorously prevents attackers from inferring true local density via the size of the returned set.

Subsequently, the continuous Gaussian mechanism satisfying ρ_{loc} -zCDP is applied exclusively to the coordinates of the masked candidate set. Distances are calculated using full-matrix vectorization to retrieve the approximate Top- k set. According to the sequential composition property, the total privacy consumption is strictly bounded by $\rho_t = \rho_{count} + \rho_{loc}$, achieving rigorous privacy preservation while successfully restricting the worst-case query complexity to the local finite neighborhood.

5. Experimental evaluation and analysis

5.1. Experimental environment and default parameters

To measure the performance of boundary queries in a high-concurrency environment reliably, a matrix-vectorization engine was used as the base and the real-world GeoLife trajectory dataset was selected for evaluation. The ingestion and maintenance of the spatial inverted index are carried out in the background asynchronously and do not block the link for real-time queries. The reported query latency is specifically defined as the time taken for the in-memory algorithmic execution, excluding external network I/O and operating system scheduling overhead. The default core parameters of the experiment are as follows: Table 1

Table 1. Settings of the main experiment

Parameter	Description	Default / Range
N	Active Data Volume	20000 / [1000, 100000]
ϵ	Privacy Budget (Mapped from z-CDP ρ_t)	1.0 / [0.2, 2.0]
k	Nearest Neighbor Query Count	50
l	Grid Discretization Granularity (km)	0.1 / [0.1, 10.0]

5.2. Core hyperparameter sensitivity and privacy-utility trade-off

The core hyperparameter of this architecture is the grid discretization granularity l . Unlike the idealized utility assumptions in static models, this paper conducts extensive physical stress tests. To eliminate the variance caused by random noise injection and accurately capture the asymptotic theoretical trends, all reported utility metrics are averaged over 100 independent Monte Carlo

DOI: 10.54254/2755-2721/2026.35334

simulations under identical parameter settings. Figure 2 demonstrates the comprehensive impact of l on system performance, revealing a strict privacy-utility-efficiency monotonic trade-off strictly governed by zCDP theory.

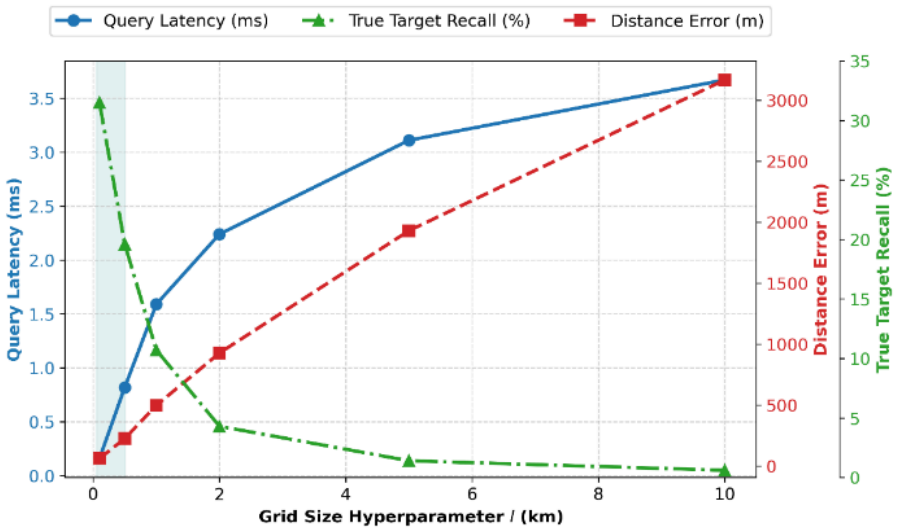

Figure 2. Core hyperparameter sensitivity analysis: monotonic trade-off among latency, distance error and recall rate

As illustrated, when the grid size expands (e.g., l increases from 0.1 km to 10.0 km), the indistinguishability region enlarges. According to the global sensitivity theorem of DP, the system is mathematically forced to inject exponentially larger Gaussian noise for masking. Consequently, the average distance error surges monotonically from 62.23 m to over 3100 m, while the true target recall rate decays from 31.5% to near 0%. Conversely, query latency climbs from 0.16 ms to 3.67 ms due to the increased candidate scanning scope.

5.3. Optimal zone analysis and edge deployment viability

Based on the real-world data distribution, this paper identifies the optimal parameter zone at $l = 0.1$ km. Within this tight local neighborhood, the Pruning-CGP architecture successfully suppresses the noise variance required for continuous spatial masking. The system tightly controls the average distance error at a practical 62 meters and maintains a robust core recall rate of 31.5% under strict privacy equivalence constraints.

By utilising the strengths of physical decoupling and vectorization acceleration, the query latency of the single-node system has been reduced to a new low of 0.16 ms. It has a fast response speed and can thus avoid the computational bottlenecks of current mechanisms in resource-constrained environments; it is suitable for a highly dynamic edge-computing deployment.

6. Conclusion

Addressing privacy leakage and computing power bottlenecks of highly dynamic spatial data streams in edge computing environments, this paper proposes a lightweight collaborative query architecture named Pruning-CGP. By physically decoupling data ingestion from real-time queries and leveraging full-matrix vectorization, the architecture achieves a theoretical $\mathcal{O}(1)$ lower bound. Under the unified theoretical framework of zero-concentrated differential privacy (zCDP), this

DOI: 10.54254/2755-2721/2026.35334

research introduces discrete Gaussian cardinality masking to rigorously block side-channel density leakage vulnerabilities brought by local spatial pruning. Real-world physical stress tests verify that, compared to existing paradigms that overlook cardinality risks or global traversal overheads, Pruning-CGP offers a highly scalable and secure solution. By setting the optimal grid parameter ($l = 0.1$ km), the system achieves an ultra-low sub-millisecond query latency of 0.16 ms while maintaining a practical utility baseline (62 m error, 31.5% recall) under strict privacy constraints, providing a robust and efficient privacy-preserving paradigm for future edge intelligence applications.

References

[1] Lu, W., Ren, Z., Xu, J., et al. (2021). Edge blockchain assisted lightweight privacy-preserving data aggregation for smart grid. *IEEE Transactions on Network and Service Management, 18*(2), 1246–1259.

[2] Sun, X., Zhang, W., He, H., et al. (2025). Survey of differentially private methods for trajectory data. *Chinese Journal of Network and Information Security, 11*(3), 1–18.

[3] Liang, Y., & Yi, K. (2023). Concentrated geo-privacy. In *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security*, 1934–1948.

[4] Patil, R. A., & Patil, P. D. (2024). Efficient approximation and privacy preservation algorithms for real time online evolving data streams. *World Wide Web, 27*(1), 5.

[5] Wang, T., Yang, X. Y., Ren, X. B., et al. (2021). Data-adaptive privacy-preserving mechanism for data stream publishing in real-time (in Chinese). *Sci Sin Inform, 51*, 1199–1216.

[6] Zhang, X., Fu, N., & Meng, X. (2020). Towards spatial range queries under local differential privacy. *Journal of Computer Research and Development, 57*(4), 847-858.

[7] Bao, E., Yang, Y., Xiao, X., & Yin, W. (2021). CGM: An enhanced mechanism for streaming data collection with local differential privacy. *Proceedings of the VLDB Endowment, 14*(11), 2258-2270.

[8] Chen, Q., Ni, Z., Zhu, X., et al. (2023). Differential privacy histogram publishing method based on dynamic sliding window. *Frontiers of Computer Science, 17*(4), 174809.

[9] Dwork, C., & Rothblum, G. N. (2016). Concentrated differential privacy. *arXiv preprint arXiv: 1603.01887.*

[10] Bun, M., & Steinke, T. (2016). Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *Theory of Cryptography Conference*, 635-658. Springer.

[11] Chen, Q. (2023). *Research on differential privacy protection methods for dynamic high-dimensional data* [Doctoral dissertation]. Hefei University of Technology.

[12] Du, L. (2024). *Research on spatial temporal data distribution estimation and application under differential privacy* [Master's thesis]. East China Normal University.

[13] Wang, J. (2024). *Research on dynamic data publishing based on differential privacy* [Doctoral dissertation]. Nanjing University of Science & Technology.

[14] Andrés, M. E., Bordenabe, N. E., Chatzikokolakis, K., & Palamidessi, C. (2013). Geo-indistinguishability: Differential privacy for location-based systems. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security*, 901-914.