**Privacy-Preserving Skyline Query Processing Over Encrypted Cloud Databases**

Lipika Baniya, [lbaniya@siue.edu](mailto:lbaniya@siue.edu)

**Abstract**

With the rapid expansion in data volumes, many businesses and individuals are increasingly inclined to outsource their data to public cloud services and to support that skyline query has become an important issue in database research for extracting interesting objects from multi-dimensional datasets. Skyline Queries emerges as an important query for retrieving data from encrypted cloud databases. This paper explores the state-of-the-art research on skyline queries by reviewing key publications and analyzes the limitations and strengths of the existing queries and suggests future development directions. Through an in-depth review of foundational concepts, fast encrypted skyline query processing, user-defined skyline queries and location-based skyline queries, we explore the strengths and limitations of current skyline queries. Despite increasing popularity and expansion in the cloud databases including preserving privacy and encryption, challenges remain in retrieving data while preserving security. This paper provides summary of the current approaches and recommendations for future research to preserve privacy and security over encrypted cloud databases.

1. **Introduction**

The on-demand deployment of computing resources via the internet is known as cloud computing. With the recent rise of industry heavyweights like Google, Amazon, and Google at forefront providing and developing cutting-edge and affordable cloud computing it has become a widespread recognition worldwide. More and more organizations are using cloud due to its pay-as-you-go pricing model and many more benefits.

Cloud databases offer scalable and flexible data storage and management for organizations. However, cloud has always had security concerns especially with sensitive data such as financial records, health information, or intellectual property. Outsourcing sensitive data to third party cloud providers introduces substantial privacy and security concerns to any organizations. To address this, encrypting the data before outsourcing has been practiced.

Once the data is encrypted, simple queries do not work hence executing complex queries such as skyline queries range queries become non-trivial. This research paper focuses on the state-of-the-art methods developed to perform skyline queries over encrypted databases in cloud-native or multi-tenant environments, where privacy and efficiency must coexist. Skyline queries ask for a set of interesting points from a potentially large set of data points.[1] It is an important class of multi-criteria decision-making queries in databases, frequently used in applications such as recommendation system, cloud resource selection, and financial analytics. [3] With increasing reliance and popularity on cloud storage and services, securely executing complex queries over encrypted cloud databases has become essential. This research investigates recent advancements on privacy-preserving skyline query processing, focusing on approaches that support user-defined, dynamic, and verifiable skyline queries while protecting sensitive data and access patterns. The goal is to analyze current techniques, identify limitations in scalability, leakage resilience, and expressiveness, and propose a path forward for secure, efficient, cloud-native skyline query execution.

This paper provides in-depth analysis of the current state of research on privacy preserving skyline queries by summarizing the key contributions so far. This paper examines recent papers and states the key problems to outline broad themes and identify gaps. Finally, this paper suggests a roadmap for future research and development in this area.

1. **Problem Statement**

Motivatedby cloud security concerns, there is an increasing interest in database systems that can store and support queries over encrypted data.Even with strong encryption, the data access pattern from untrusted storage has the potential to reveal sensitive information. Techniques like homomorphic encryption and oblivious RAM offer strong privacy guarantees but often make query execution prohibitively slow for real-world use. The central problem is how to design and evaluate skyline query processing systems that preserve privacy while maintaining scalability and practical performance in encrypted cloud environments. [4]

Despite having advanced cryptographic tools, the existing systems still face major limitations including limited support for user-defined dynamic skyline queries, high and insufficient focus on cloud-native deployment, scalability and verifiability of results. There is a pressing need for more research on queries that preserve privacy while maintaining its performance. Addressing these challenges through ongoing research is essential to meet the demands of modern applications and leverage the full potential of cloud databases.

1. **Related Work**

This section discusses the contributions and significance of the six core papers. Each paper addresses different aspects of privacy-preserving skyline queries, from foundational concepts to recent advancements in encrypted cloud databases.

A skyline query of a set of points (tuples) is defined as those points that are not dominated by any other point. A point p dominates another point q if p is as good or better than q in all dimensions and strictly better in at least one dimension. [1] Skyline queries help users make the best choices by considering multiple factors, making them great for decision-making in various fields. It can involve more than two dimensions and depend on the current position of a user. For example, users could be interested in restaurants that are near, cheap, and have good food. The distance would be based on the current location of the user. The idea of skyline is to give the user a big picture of interesting options and then let the user decide. Once the user moves on, the Skyline is re-computed continuously to give the user a choice of interesting restaurants based on user’s new location. [4] Since skyline queries can be dynamic, they must often adapt to user context, such as changing locations, requiring continuous re-computation.

There are two progressive Skyline algorithms, the first algorithm is based on Bitmaps, and the second algorithm is based on an extension of B-trees. [1] Both algorithms faced limitations in interactive environments due to dependencies on data clustering and value distribution, respectively. For the Bitmap algorithm, the order in which Skyline points are returned depends on the clustering of the data; for the B-tree algorithm, the order in which points are returned depends on the value distribution of the data. As a result, both algorithms are not advantageous in an interactive environment. To address these issues, the Nearest Neighbor (NN) algorithm was introduced, which is more interactive but lacks a definitive advantage over its predecessors due to trade-offs in different use cases. There is no clear winner as all algorithms have their virtues. [1]

To support skyline queries over encrypted cloud databases, the SecSkyline framework was proposed, it focuses on efficiency, dominance protection, and secure skyline queries over outsourced databases. Previously, a large body of work has been presented for querying encrypted databases, which has been mostly focused on secure keyword search but a major challenge for it was ensuring both data confidentiality and query privacy challenge on encrypted cloud databases. [4] The main challenge was to prevent the cloud from learning data patterns that could lead to indirect leakage. SecSkyline addresses this problem by utilizing a lightweight cryptographic technique called “additive secret sharing” that enables efficient encrypted skyline computation. It introduces database mapping in an encrypted query, which allows the cloud to securely map the encrypted outsourced database to the new space as it facilitates the subsequent secure skyline tuples fetching.[4] Furthermore, SecSkyline incorporates an effective technique for secure skyline tuple fetching and oblivious filtering of skyline and dominated tuples, all while keeping the cloud unaware of actual data content or positions. In gist, SecSkyline combines cryptographic primitive and careful protocol design to enable the cloud to perform the complex comparisons and filtering required for skyline queries, all while operating solely on encrypted data and ensuring that no sensitive information about the data, the query, or the results.

On the similar note, an innovative framework named SCALE (Secure Cloud-based Approximate Location-based skyline Execution) is introduced to process dynamic skyline queries over encrypted data. It supports secure subtraction and comparison operations without frequent decryption; the framework facilitates processing of dynamic skyline queries on encrypted data by reducing the need for decryption and preserves data confidentiality. [5] Since, there is only one interaction between a user and the cloud, it minimizes the communication cost and corresponding threats that come along. SCALE also introduces distributed versions, DIST-SCALE-S and DIST-SCALE-E, to enhance scalability via parallel processing.[5]

Additionally, ObliuSky User-Defined Skyline Queries addresses flexibility and customization in skyline dimensions. Unlike traditional skyline queries, which uses fixed dimensions, ObliuSky accommodates the tailored needs of user’s real-world applications by customizing dimensions, preferences and constraints.[2] However, this flexibility raises additional privacy concerns over cloud databases. With respect to the problem stated above, ObliuSky framework provides confidential protection and makes the cloud oblivious to the data patterns which may indirectly cause data leakages.[2] It ensures that the cloud learns no true data patterns while still correctly identifying skyline results. The framework processes user-defined skyline queries obliviously, safeguarding both user preferences and data content.

Finally, the SVLSQ (Secure and Verifiable Location-based Skyline Query) framework introduces verifiability to skyline queries over encrypted location-based data. Recognizing the untrustworthiness of cloud environments, SVLSQ avoids compromising privacy of datasets, queries, results and access patterns while also authenticating the soundness and completeness of the skyline results.[6] It introduces the SR-tree (semi-blind R tree) and SVSR-tree, which are indexing structures designed to support secure skyline computation while protecting query unlinkability. The method allows verification of both the soundness (correctness) and completeness (no missing results) of skyline queries, making it a solution for auditable skyline services in cloud databases. [6]

1. **Analysis**

Skyline queries have emerged as a critical operation for multi-criteria decision-making across a range of applications. Given their utility, skyline queries have been widely adopted to operate over encrypted datasets in cloud environments. Each work introduces unique strategies to address the challenges privacy, efficiency, and verifiability under encryption. The analysis of these papers reveals that while there has been significant progress and development in the field of skyline queries over encrypted database there are still gaps in research. This paper examines how these methods relate to one another, assess the current maturity of the field, and identify unresolved challenges that present opportunities for future research. The selected papers present different approaches to addressing these challenges through encryption techniques, secure query protocols, and verifiability mechanisms.

The early work focused on algorithm design (e.g. bitmap-based and B-tree skyline algorithms), then new privacy-preserving methods were introduced that uses encryption and secure computation to preserve privacy over cloud databases. The main goal of all algorithms mentioned in this paper is to protect against data pattern leakage, reduce communication rounds between user and server, use secret sharing, homomorphic encryption, and oblivious data access techniques. This analysis has been categorized and evaluated across four main dimensions: Privacy Preservation, Query Efficiency, Verifiability and Trust, and Flexibility and User Customization.

* 1. **Privacy Preservation Techniques**

The central concern of secure skyline query processing is how to preserve user and data privacy. SecSkyline, SCALE, and ObliuSky all focus on protecting sensitive information, but by employing different cryptographic techniques and assumptions.

* SecSkyline supports secure database mapping in an encrypted query, by allowing the cloud to securely map the encrypted outsourced databased to the new space.[4] It introduces an effective technique to tackle the challenging operation of computing absolute value in the secret sharing domain which hides search access patterns. It uses additive secret sharing to ensure that query patterns, access patterns, and skyline results are protected from the cloud. It introduces skyFetch and skyFilt, to hide the identity of skyline tuples. However, SecSkyline assumes a semi-honest and non-colluding adversary model where each cloud server honestly follows their protocol but may try to infer information from the data when it processes, which reflects real-world cloud threats.
* Unlike traditional skyline models, SCALE adopts a different approach to handle dynamic skyline query, which require real-time evaluation of comparison and distance evaluation online simultaneously such as user’s current location. To achieve this, SCALE transforms the skyline domination condition into a series of comparisons, by eliminating the need for subtraction operations which is difficult to perform securely on encrypted data.[5] This reduces complexity under encryption, but like SecSkyline, it assumes a cooperative cloud and lacks support for robustness.
* ObliuSky focuses on oblivious user-defined skyline queries, where users can specify the dimensions and filtering conditions over encrypted data.[2] It addresses key privacy concerns, including data pattern leakage, by using techniques such as lightweight secret sharing and shuffling that restricts tuple access and data relationships. One of the key features is to enable the cloud to perform skyline query processing on a shuffled, encrypted dataset by creating confusion on what tuples are selected. This framework supports real-world customization while supporting privacy and preventing inference attacks. [2]

All three systems achieve different levels of privacy, but none of them fully addresses both strong adversaries and verifiability, hence suggesting need for further research.

* 1. **Query Efficiency**

While security is essential, efficiency is also important when it comes to querying data from the cloud databases. Having too high security can impact performance and efficiency, suggesting a need of balance between security and performance. Performance is a key barrier to deploying skyline queries over encrypted data in real-world systems.

* SecSkyline shows great improvement in performance through the efficient use of secret sharing and parallel operations over state-of-the-art prior works in query latency, with up to 8130x improvement over FSSP and up to 813x improvement over SMSQ.[4]
* SCALE empathizes low communication cost, by supporting single-round interaction. It also introduces two distributed versions of SCALE, DIST-SCALE-S and DIST-SCALE-E, which further optimize performance by facilitating parallel processing.[5]

However, both solutions are optimized only under specific conditions such as SecSkyline for static queries and SCALE for dynamic skylines. Hence, concluding that no single solution offers universal efficiency across query types.

* 1. **Verifiability and Trust**

Verifiability means ensuring the correctness and completeness of results which remains to be addressed in many of the approaches provided in the selected topics.

* SVLSQ (Secure and Verifiable Location-based Skyline Queries) is the only approach among the selected papers that addresses this challenge. It addresses this by using a novel semi-blind R-tree (SVSR-tree) index, it supports both soundness and completeness verification while preserving data privacy. [6]
* In contrast, SecSkyline, SCALE, and ObliuSky do not verify query correctness, this method assumes that the cloud follows the protocols faithfully. This creates a potential trust gap, particularly untrusted settings.

SVLSQ thus fills a very essential gap, but at the same time introduces overhead due to cryptographic indexing, which may not be ideal for large-scale or real-time applications.

* 1. **Flexibility and User Customization**

Often in real-world, there are rarely queries that are tailored to specific user preferences.

* Among the selected papers, ObliuSky is notable for supporting user-defined skyline queries, which allows users to specify dimensions and constraints based on individual needs.[2] This increases applicability in personalized systems, such as recommendation systems.
* Other systems such as SecSkyline and SCALE focuses on general skyline and lack built-in support for user-specific constraints or flexible query dimensions.

Thus, ObliuSky offers greater flexibility and personalization, though it adds complexity and slightly reduced performance.

1. **Future Directions**

Security remains a critical concern as cloud databases continue to grow exponentially. While the reviewed works represents significant progress in secure skyline query processing, there is no single solution that balances privacy, performance, verifiability, and flexibility. Future research in skyline queries over encrypted cloud databases should focus on developing queries that balances performance and privacy security. There are several areas for future research and improvement.

* Integration of Verifiability into Efficient Systems: As mentioned many of the reviewed algorithms (e.g. SecSkyline, SCALE) assumes honest or semi-honest server and lack mechanisms to verify result correctness. With respect to that, this paper suggests the future work to aim to combine SVLSQ or embed verifiability into these frameworks without significantly compromising performance.
* Support for Richer Query Semantics: Most of the frameworks mentioned in the reviewed papers support basic or dynamic skyline queries. This paper suggests future research to aim to support range-based skylines, group-by skyline queries, or continuous skyline monitoring over encrypted data.
* Handling Frequent Data Updates: Most of the frameworks are focused on working with static datasets. Future work should be focused on how to handle efficient skyline recomputation when encrypted data is frequently inserted, deleted, or modified.
* Combine Algorithms for Hybrid Query Execution: Since, none of the frameworks provide a unified solution future works should seek opportunity to explore hybrid methods that combine the advantages of different skyline algorithms.

1. **Conclusion**

Skyline queries are a powerful tool for multi-criteria decision-making systems in different domains. They work well over encrypted cloud databases by ensuring both data utility and privacy in outsourced environments. This paper studied and analyzed recent papers that worked on privacy-preserving skyline query processing, most of them focused on key challenges such as preserving confidentiality, ensuring efficiency, supporting verifiability, and enabling flexible query structures.

The reviewed frameworks for instance SecSkyline, SCALE, ObliuSky, and SVLSQ all offer meaningful and significant contributions to all the key challenges that exist in the Skyline queries, however none of them fully satisfies all four requirements. Each of them has their own focus and strengths, SecSkyline and SCALE excels in efficiency, SVLSQ addresses verifiability, and ObliuSky introduces user-driven flexibility. This paper suggests that integrating these frameworks and developing a hybrid solution will define the next generation of privacy-preserving skyline query systems.

This paper suggests, research must focus on combining these strengths and addressing issues like dynamic data, adversarial threat models, and scalable verification, all while maintaining practical performance. As cloud-based encrypted databased continue to evolve, the ability to securely and efficiently execute skyline queries will be a cornerstone of privacy0-aware decision support systems.

1. **Core Papers**

* Shooting stars in the sky: An online algorithm for skyline queries [1]
* ObliuSky: Oblivious User-Defined Skyline Query Processing in the Cloud [2]
* An Efficient Framework for Secure Dynamic Skyline Query Processing in the Cloud [3]
* SecSkyline: Fast Privacy-Preserving Skyline Queries over Encrypted Cloud Databases [4]
* Towards Efficient and Privacy-Preserving User-Defined Skyline Query Over Single Cloud [5]
* Efficient Secure and Verifiable Location-Based Skyline Queries over Encrypted Data [6]

**References**

[1] Kossmann, D., Ramsak, F., & Rost, S. (2002, January). Shooting stars in the sky: An online algorithm for skyline queries. In VLDB'02: Proceedings of the 28th International Conference on Very Large Databases (pp. 275-286). Morgan Kaufmann.

[2] Zheng, Y., Wang, W., Wang, S., Hua, Z., & Gao, Y. (2024). ObliuSky: Oblivious User-Defined Skyline Query Processing in the Cloud. IEEE Transactions on Services Computing.

[3] Chen, P., Xu, B., Li, H., Wang, W., Peng, Y., Bhowmick, S. S., ... & Cui, J. (2025). An efficient framework for secure dynamic skyline query processing in the cloud. Data Science and Engineering, 10(1), 54-74.

[4] Zheng, Y., Wang, W., Wang, S., Jia, X., Huang, H., & Wang, C. (2022). SecSkyline: Fast privacy-preserving skyline queries over encrypted cloud databases. IEEE Transactions on Knowledge and Data Engineering, 35(9), 8955-8967.

[5] Zhang, S., Ray, S., Lu, R., Zheng, Y., Guan, Y., & Shao, J. (2022). Towards efficient and privacy-preserving user-defined skyline query over single cloud. IEEE Transactions on Dependable and Secure Computing, 20(2), 1319-1334.

[6] Wang, Z., Ding, X., Jin, H., & Zhou, P. (2022). Efficient secure and verifiable location-based skyline queries over encrypted data. Proceedings of the VLDB Endowment, 15(9), 1822-1834.