**CHAPTER-2**

**INTRODUCTION**

Similarity query over time series data, which aims to identify samples that are similar to the sample of interest, has a significant number of applications in various areas, such as signal processing, speech recognition, and disease diagnosis. For example, the similarity query over time series ECG data can be used to detect Premature Ventricular Contraction disease. Undoubtedly, the wide application has made the similarity query over time series data popular. Nevertheless, as the rapid development of the Internet of Things (IoT), tremendous volumes of data are generated with high velocity. As reported in, the total amount of data will reach 149 zettabytes by 2024. Such tremendous volumes of data will seriously affect the efficiency of the similarity query. To improve query efficiency, many data owners migrate their data to the computationally powerful cloud and delegate the cloud server to perform similarity queries. However, since the data are private assets of the data owners and the cloud server is not fully trusted, exposing the plaintext data to the cloud server may inflict severe economic loss to the data owners. For addressing the privacy issue, data owners usually leverage encryption techniques to encrypt the data before outsourcing them to the cloud, yet a consensus has emerged that encryption techniques will hinder the cloud server to perform the similarity queries over the outsourced time series data.

To tackle the dilemma of similarity queries over encrypted time series data, various schemes were proposed. Based on the similarity metric, existing schemes can be divided into three categories, i.e., Lp-norm based similarity query schemes, edit distance-based similarity query schemes, and dynamic time warping (DTW) based similarity query schemes. However, Lp based schemes, cannot support similarity queries over time series data with different lengths. Although edit distance and DTW distance-based schemes, can support similarity queries over time series data with different lengths, the schemes, can only return approximate query results and the schemes, suffer from the linear search efficiency. Hence, existing schemes still have issues in supporting similarity queries over time series data with different lengths, query accuracy and query efficiency

**LITERATURE SURVEY**

# Title: An Efficient and Privacy-Preserving Range Query over Encrypted Cloud Data

### Author: [Wentao Wang](https://ieeexplore.ieee.org/author/37089491921); [Yuxuan Jin](https://ieeexplore.ieee.org/author/37089490776); [Bin Cao](https://ieeexplore.ieee.org/author/37533853600)

# Abstract: The growing power of cloud computing prompts data owners to outsource their databases to the cloud. In order to meet the demand of multi-dimensional data processing in big data era, multi-dimensional range queries, especially over cloud platform, have received extensive attention in recent years. However, since the third-party clouds are not fully trusted, it is popular for the data owners to encrypt sensitive data before outsourcing. It promotes the research of encrypted data retrieval. Nevertheless, most existing works suffer from single-dimensional privacy leakage which would severely put the data at risk. Up to now, although a few existing solutions have been proposed to handle the problem of single-dimensional privacy, they are unsuitable in some practical scenarios due to inefficiency, inaccuracy, and lack of support for diverse data. Aiming at these issues, this paper mainly focuses on the secure range query over encrypted data. We first propose an efficient and private range query scheme for encrypted data based on homomorphic encryption, which can effectively protect data privacy. By using the dual-server model as the framework of the system, we not only achieve multi-dimensional privacy-preserving range query but also innovatively realize similarity search based on MinHash over ciphertext domains. Then we perform formal security analysis and evaluate our scheme on real datasets. The result shows that our proposed scheme is efficient and privacy-preserving. Moreover, we apply our scheme to a shopping website. The low latency demonstrates that our proposed scheme is practical.

# Title: Achieving Efficient and Privacy-Preserving Exact Set Similarity Search over Encrypted Data

# Author: [Yandong Zheng](https://ieeexplore.ieee.org/author/37086847433); [Rongxing Lu](https://ieeexplore.ieee.org/author/37309741500); [Yunguo Guan](https://ieeexplore.ieee.org/author/37086474789); [Jun Shao](https://ieeexplore.ieee.org/author/37712761400); [Hui Zhu](https://ieeexplore.ieee.org/author/37076349100)

# Abstract: Set similarity search, aiming to search the similar sets to a query set, has wide application in today's recommendation services. Meanwhile, the rapid advance in cloud technique has promoted the boom of data outsourcing. However, since the cloud is not fully trustable and the data may be sensitive, data should be encrypted before outsourced to the cloud. Undoubtedly, data encryption will hinder some basic functionalities, e.g., set similarity search. For achieving set similarity search over encrypted data, many solutions were proposed, yet they either only satisfy weak security requirements, or only achieve approximate similarity, or have low efficiency or under the model of two cloud servers. Therefore, in this article, we propose a new efficient and privacy-preserving exact set similarity search scheme under a single cloud server. Specifically, we first design a symmetric-key predicate encryption (SPE-Sim) scheme, which can support similarity search over binary vectors. Then, we represent the set records to be binary vectors and employ the B+ tree to build an index for them. After that, based on SPE-Sim and the B+ tree-based index, we propose our scheme and it can achieve efficient set similarity search while preserving the privacy of set records and query contents. Finally, security analysis and performance evaluation indicate that our scheme is privacy-preserving and efficient.

# Title: Efficient and Privacy-Preserving Spatial Keyword Similarity Query Over Encrypted Data

### Author: Songnian Zhang; Suprio Ray; Rongxing Lu; Yunguo Guan; Yandong Zheng;

# Abstract: As a popular and practical query type in location-based services, the spatial keyword query has been extensively studied in both academia and industry. Meanwhile, with the growing demand for data privacy, many privacy-preserving spatial keyword query schemes have been proposed to deal with queries over encrypted data. However, none of the existing schemes preserve access pattern privacy, and the recent research illustrates that leaking such privacy may incur inference attacks and thus disclose sensitive information. In addition, most existing schemes only consider the boolean keyword search, which is not quite practical and flexible in real-world applications. To address the above issues, in this paper, we propose two privacy-preserving spatial keyword similarity query schemes that can preserve full and partial access pattern privacy, respectively. First, we present a basic privacy-preserving spatial keyword similarity query scheme (PPSKS) by integrating a secure set membership test (SSMT) technique with secure circuits. After that, to improve performance, we propose a tree-based scheme (PPSKS+) by employing a new index called FR-tree together with a predicate encryption technique that can encrypt FR-tree. Formal security analysis shows that: i) our proposed schemes can protect outsourced data, query requests, and query results; ii) our PPSKS scheme can hide full access patterns, while the PPSKS+ scheme preserves m -access pattern privacy. Extensive experiments are also conducted, and the results indicate that our tree-based PPSKS+ scheme is much more efficient, almost two orders of magnitude better than our linear search PPSKS scheme in performing queries.

### Title: Privacy-Preserving Keyword Similarity Search Over Encrypted Spatial Data in Cloud Computing

# Author: [Fuyuan Song](https://ieeexplore.ieee.org/author/37086884107); [Zheng Qin](https://ieeexplore.ieee.org/author/37594217000); [Liang Xue](https://ieeexplore.ieee.org/author/37087228384); [Jixin Zhang](https://ieeexplore.ieee.org/author/37085860762); [Xiaodong Lin](https://ieeexplore.ieee.org/author/37289264200); [Xuemin Shen](https://ieeexplore.ieee.org/author/37278603300)

# Abstract: With the proliferation of cloud computing, data owners can outsource the spatial data from the Internet of Things devices to a cloud server to enjoy the pay-as-you-go storage resources and location-based services. However, the outsourced services may raise privacy concerns, since the cloud server may not be fully trusted for both data owners and search users. If the data owners and search users conventionally encrypt the spatial data and query requests, the efficiency and functionality of query processing are weakened. Most of the existing works only focus on spatial data search or keyword search and do not consider spatial keyword search over encrypted data. In this article, we first design a geometric range query (GRQ) scheme, which can generate an arbitrary geometric range to fit the search user’s desired spatial data while protecting location privacy. Furthermore, based on GRQ, we propose a multidimensional spatial keyword similarity search scheme with access control (MSSAC) by integrating the polynomial function and matrix transformation. Specifically, an access control strategy is defined by a role-based polynomial function, which is embedded in the vectors of indices and trapdoors to achieve efficient and lightweight access control. Moreover, MSSAC enables the cloud server to execute compute-then-compare operations for spatial keyword search in a privacy-preserving manner by leveraging techniques of randomizable permutation and matrix multiplication. The formal security analyses and extensive experiments demonstrate that GRQ and MSSAC preserve the privacy of data owners and search users while achieving efficient spatial keyword search.

# Title: Privacy-preserving multi-keyword fuzzy search over encrypted data in the cloud

# Author: [Bing Wang](https://ieeexplore.ieee.org/author/37085454558); [Shucheng Yu](https://ieeexplore.ieee.org/author/37310268900); [Wenjing Lou](https://ieeexplore.ieee.org/author/37275345500); [Y. Thomas Hou](https://ieeexplore.ieee.org/author/37280000900)

# Abstract: Enabling keyword search directly over encrypted data is a desirable technique for effective utilization of encrypted data outsourced to the cloud. Existing solutions provide multi-keywsord exact search that does not tolerate keyword spelling error, or single keyword fuzzy search that tolerates typos to certain extent. The current fuzzy search schemes rely on building an expanded index that covers possible keyword misspelling, which lead to significantly larger index file size and higher search complexity. In this paper, we propose a novel multi-keyword fuzzy search scheme by exploiting the locality-sensitive hashing technique. Our proposed scheme achieves fuzzy matching through algorithmic design rather than expanding the index file. It also eliminates the need of a predefined dictionary and effectively supports multiple keyword fuzzy search without increasing the index or search complexity. Extensive analysis and experiments on real-world data show that our proposed scheme is secure, efficient and accurate. To the best of our knowledge, this is the first work that achieves multi-keyword fuzzy search over encrypted cloud data.