A Survey on Spatial Keyword Search over Encrypted Data

Zenebe Yetneberk(✉)

School of Cyber Engineering, Xidian University, Xi'an, China

Abstract. Many real-time applications use spatial keyword queries, which provide location information and text descriptions of Points Of Interest (POI). Since a promising research subject can pose such a procedure, performing spatial keyword queries over encrypted data is difficult. Several schemes for addressing secure spatial and keyword queries have been proposed, but previous surveys have only summarised and tested secure query algorithms. They still lack a privacy-preserving overall review of the spatial keyword query. This paper outlines a secure spatial keyword query's three main components: secure spatial query, secure textual query, and secure spatial-textual query. Some evaluation criteria have been carefully selected to aid in the evaluation of existing spatial keywords. Following that, we expand on and evaluate recent related research's benefits and disadvantages using the suggested criteria. Finally, we see some unanswered problems that researchers can use to perform more analyses and studies.

Keywords: Privacy preservation · Secure queries · Spatial keywords query · Privacy-preserving query

1 Introduction

Many researchers have been paying attention to spatial keyword queries in recent years due to the widespread usage of location-based services. By outsourcing their spatial textual data, including indexes, to a cloud service provider (CSP), data owners can competently support the online spatial keyword query process. Users who want to run queries on their data can submit requests to CSP, which will handle them quickly.

However, those outsourced services may incur some privacy leakage problems since the spatial-textual data is sensitive, and the cloud server may not be completely trustworthy. Besides, collecting spatial-textual objects consumes both human and financial resources, considered business secrets for competitors and any unauthorized parties. Moreover, in any case, if data users' spatial keyword queries are illegally obtained, potential attackers could use that data to eavesdrop on the privacy information [17].

J. Xiong et al. (Eds.): MobiMedia 2021, LNICST 394, pp. 413–431, 2021.
https://doi.org/10.1007/978-3-030-89814-4_30

There are already some surveys on spatial keyword queries. Lisi *et al.* [8] proposed a benchmark comparing the performance of spatial keyword query algorithms. Eldawy and Mokbel [15] classified the present work in this field into three distinct aspects, namely, approach, design, and components. Chen *et al.* [7] explained the existing studies on the query of multi-modal road network location-based data and categorized the prevalent work. Qi *et al.* [27] introduced the key ideas for underlying safe region techniques within which no query update in a very given region ranged a few times and illustrated how they were applied in several continuous spatial query algorithms to provide various query types. However, none of them has considered secure and privacy-preserving problems of spatial keyword queries, as we compare them in the Table 1. Thus, we encourage to conduct a systematic survey that summarizes recent state-of-the-art spatial, text, and spatial keyword search solutions for protection and privacy.

Table 1. Comparison of our survey with other survey

Covered topics	[7]	[27]	[22]	[30]	Our survey
Give a comprehensive review	N	N	N	N	Y
Proposed a set of evaluation criteria	Y	N	N	Y	Y
Summarize the pros and cons	N	N	N	N	Y
Analyze the performance	N	N	N	Y	Y
Propose some open issues and future direction	Y	Y	Y	N	Y

This survey will classify and thoroughly review privacy-preserving spatial keyword queries by grouping them into three main parts: spatial, textual, and spatial keyword queries. We study each query type one by one for analyzing their pros and cons. To instruct our analysis on current works' success, we propose a set of evaluation criteria that assist our judgment on potential investigation patterns.

According to our survey, some open research issues and important research directions that merit additional research efforts have been identified. The following are some of the most important contributions to our work:

- An extensive review of privacy problems on secure spatial, textual, and spatial keyword query technologies.
- Recommending a set of assessment criteria, and seriously overview existing work as well as analyze the strengths and weaknesses of these work based on our proposed criteria.
- Based on a detailed analysis, we propose some open issues and forecast future research trends.

2 Background

This section introduces basic concepts related to the spatial query, text query, and spatial-textual query. Spatial-keyword queries have three main types [13].

- Boolean Range Query (BRQ), $Q(R, doc)$: This query returns spatial objects containing all the keywords in the set *doc* and the range R, where a spatial range is R, and a set of keywords is *doc*.
- Boolean kNN Query (BkQ), $Q(loc, doc, k)$: It returned a set of k objects, each containing all the *doc* keywords ranked according to its spatial proximity to the location of query *loc*.
- Top-k Query (TkQ), Q(loc; doc; k): Based on their combined textual and spatial significance scores, this results in a collection of ranked k objects, where the *doc* is a set of keys where the *doc* is a set of keywords in the query, *loc* is the location of query, and k is the parameter in the top-k query that represents the number of return objects.

Indexing Techniques: An efficient indexing structure is vital to managing big data in the cloud server, specifically in Location-Based Services (LBS) providers. These auxiliary structures accelerate the database transactions and query process by storing the database's indexed columns separately for fast look-ups. An efficient index needs to be customized to the database. Here, we are going to discuss the popular index structures, which can be found in different types of databases.

2.1 Textual Index

This index is used for databases containing only the most widely used indexes, such as bitmaps and inverted files.

- Bitmap: It stores data as an array of bits. Each bit indicates whether the object contains the keyword. Bitmaps use logical operations on the bits storing the data to answer a query. A bitmap is fast and suitable for attributes whose values frequently repeat [38].
- Inverted Files: This data structure stores the mapping from some content to its location. The inverted file includes mapping keywords to files containing the keyword. Index files are popular in large-scale search engines as they allow for a speedy full-text search [29].

2.2 Spatial Indexing

The spatial index scheme can be classified into three categories: R-tree, grid, and space-filling curve indices.

- R-tree: R-tree [19] or its variant (*e.g.* the R*-tree [1]) is included in this index. Most geo-textual indices belonging to this category use the inverted files for text indexing, as illustrated in Fig. 1. The R-tree-based indices loosely combine the R-tree and the inverted files. It makes it easy to arrange spatial, and text data [41] independently. On the other hand, most existing indices strongly combine the R-tree with a text index.

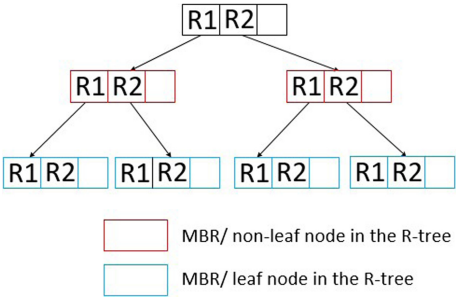

Fig. 1. R-tree

- Grid: This grid-based index entails allocation of relevant objectives to their position in the grid. It combines the index of a grid with a text index (*e.g.* the inverted file). Indices for text and grid can be prearranged either be organized separately [33] or combined tightly [21].
- Space-filling Curve: It is a curve whose range includes 2-dimensional space in its entirety. The index combines the inverted files with Z-curved based index [12] as illustrated in Fig. 2. A Hilbert curve-based index [9] is also included under this indexing category.

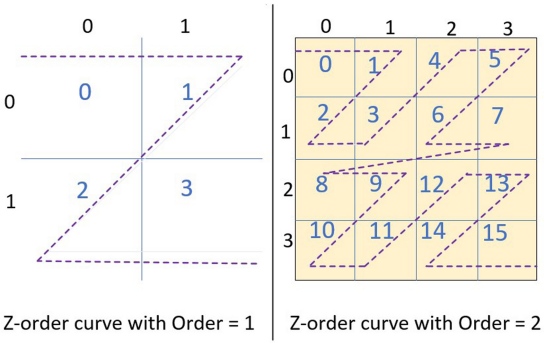

Fig. 2. Z-order curve with order 1 and order 2

2.3 Combination Indexing Schemes

Spatial and text indexing is associated with these geo-textual indices. Based on how the spatial and text indices are combined, the index can be categorized. It includes a text-first loose and tight combination.

A text-first index loses its combination index by using the text-first index. For the top-level index, the inverted file and organized posting are contained in a spatial structure in each inverted list. It can be R-tree, a space-filling curve, or a grid. On the other hand, the top-level spatial-first index is a spatial structure whose leaf nodes contain inverted files or bitmaps for the objects' text in the nodes.

Conversely, both the text and spatial index are tightly correlated with the tight combination index. A scheme to concurrently prune the search space with query processing. There are indexes of two kinds. In each inverted list, one incorporates the spatial information (*e.g.*, [12]) and the other integrates a text description into each spatial index node (*e.g.*, [13]).

Geo-Textual indexing: These are the 12 most common types of geo-textual indices for a set of spatial web objects [8]. The geo-textual indices are grouped into three features: the text index, the spatial index, and a hybrid of the two indices.

3 Evaluation and Categorizations Criteria for Secure Queries

In this unit, to categorize and compare the current system's benefits and drawbacks, we suggest a set of categorization requirements and evaluation criteria.

3.1 Categorization Criteria

We classify some related literature into three basic categories based on the query method; such as privacy-preserving spatial query, privacy-preserving textual query, and privacy-preserving spatial keyword query.

3.2 Evaluation Criteria

- Threat Model (TM): The principal research works commonly consider two different models. Some researchers contemplate the CSP as "Curious but Honest (CH)." The cloud service provider will monitor protocols that have been developed. Nevertheless, it will expose both data consumers' and owners' sensitive data. On the other hand, the CSP is considered by some researchers to be "Malicious (M)" *i.e.*, in which the CSP perhaps will return incorrect results for retrieval. Therefore, subsequent research studies seek to show recovery effects.
- Plain Text Attack (PTA)
  - *Chosen-Plaintext Attack (CPA)*: Let the cloud service provider in the dataset derive cyphertext for selected objects' plaintext and try to recover the cyphertext.
  - *Known-Plaintext Attack (KPA)*: Considering CSP, pairs of objects in the dataset can achieve plaintext-cyphertext. It seeks to overcome the secret key by using the data of these pairs to decode additional cyphertext.

3.3 Security Requirements (SR)

- Authenticity (Au): Authenticated query processing helps clients check if the query results returned by the CSP are accurate.
- Confidentiality (Cn): Since security information is essential, the data needed can only be accessed by legitimate data users. Thus, data information should not be revealed in the transmission process, and its confidentiality should be assured.
- Integrity (*In*): This refers to the user's query content in the entire process, where the cloud server is unable to get the actual query content or infer the query through the cyphertext.

3.4 Privacy Requirements (PR)

Privacy is the capability of a person or group of people to hide information about themselves. We categorize the privacy type into the following features:

- Query Privacy (QP): In the whole process, this applies to the user's query content. The cloud server is unable to get the actual query content or infer the query over the cyphertext.
- Data Privacy (DP): It refers to the data stored in the dataset. Under the process, the cloud server cannot get the actual data sets over cyphertext.
- Result Privacy (RP): This word applies to the content of the cloud server. We should consider the cloud server's resulting privacy related to the query's exact results that should not be revealed to it. From the user's point of view, they can obtain real data that satisfies the query's condition.
- Path Patterns Privacy (PPP): When executing the query, it refers to the index's traversal path. The CSP cannot reveal the actual traversal route.
- Access Pattern Privacy (APP): It refers to a relationship that defines matching points in a particular query containing the points. The cloud server cannot access this relationship between the query and the received results.

3.5 Performance Requirements (PR)

- Correctness (Cr): Correctness of data analysis is an essential aspect of performance. The query results should be correct and match the data user requests since some data analytics attacks may lead to incorrect analytical results and serious consequences.
- Scalability (Sc): When the number of objects increases, it relates to how the device reacts and allows us to assess its capacity for a broad network.
- Mobility (Mo): Mobility refers to the users' status, whether at the static or continuous positions. Static refers to the user's motion (rest at a fixed point), and dynamic refers to the user moving from one given place to another within a short period.

– Efficiency (Ef): Secure data collection and analysis should be effective and efficient since the service provider's energy power consumption needs to be minimized. Linear and non-linear are the two levels of efficiency in computational complexity measurement.
– Other requirements: Time Stamp (TS), which refers to the exact time that a query executes, and Place Description (PD), which gives some additional information about searched for places.

4 Literature Review

We analyze the published work in this survey only by looking for papers from the ACM Digital Library, Springer, IEEE Xplore Digital Library. We use search keywords, including privacy-preserving keyword query, spatial query, spatial keyword query, secure keyword, spatial and spatial keyword query. We are classifying some of the related references into three basic categories based on the query methods. Including privacy-preserving spatial query, privacy-preserving textual, and privacy-preserving spatial keyword search.

4.1 Secure Textual Query

It deals with multi-keyword Search, which has better flexibility and efficiency than the single keyword search [24], a cryptographic primitive Hierarchical Predicate Encryption (HPE) uses attribute hierarchy for simple range queries.

Boolean Search

a) Privacy-Preserving keyword Search (PPKS). A multi-round protocol between CSP and data users is a single keyword search technique. The keyword index links an individual keyword with its associated files. The heuristic pseudo-random function's primary importance is to encrypt a dictionary-based keyword index for individual files [6]. The virtues of $PPKS$ systems operate in multiple file formats, including compressed files, multimedia files, Etc. The keyword index should first be developed with priority.

b) Secure Privacy-Preserving Keyword Search (SPKS). The system helps the CSP decrypt the details and return the keyword-containing file [25]. $SPKS$ enables users to reduce communication and computational overheads, providing information and query privacy for the users. To efficiently search over encrypted data, it implements six algorithms.

Table 2. Summarazation and comparison of secure textual query

Keyword Search Category	Keyword Search Technique	Merits	Demerits
Boolean Keyword Search	PPKS [6]	—Due to pseudorandom function, it has better security	—Not applicable for multiple keyword search
	SPKS [25]	—Provide query and data privacy for the users —Has less communication and computational cost	—The system is not resiliently chosen and is known to be a plane text attack
	FKS [10,23]	—Utilized the multiway tree to improve search efficiency	—Give unsorted (ranked) search result —Search semantics is not considered
	APKS [24]	—Has better flexibility and efficiency compared to single keyword search	—Do not prevent keyword attack
	APKS+ [24]	—Prevents dictionary keyword attack —Accomplishes index and query privacy	—Not all the attributes are hierarchical
	ABE [20]	—Afford the best quality search on encrypted data —Fast accessing method	—Non-efficiency —Non-existence of attribute revocation mechanism
Ranked Keyword Search	RSSE [35]	—Provide effective protocol —Has strong security promise as compared to SSE schemes —Efficient support of relevance score dynamics	—Has a slight relevance score information leakage in contrast to keyword privacy
	K-gram [40]	—Effective and secure method in encrypted environment	—Can not support a multiway tree structure —Relatively need high storage space

c) Fuzzy Keyword Search (FKS). This search technique enhances the system's usability when searching for inputs that exactly match. The editing distance is used to quantify keyword similarity and construct fuzzy keyword sets. The data user can check the correctness and completeness of the search results. There are two approaches: wildcard and straight forward for dealing with the edit distance [10,23].

d) Authorized Private Keyword Search (APKS). A multi-keyword search that has greater flexibility and efficiency compared to a single keyword search [24]. Hierarchical Predicate Encryption (HPE) is a primitive cryptographic attribute hierarchy used in a fine-grained authorization system for a simple range of queries. Each user obtains a Local Trusted Authority (LTA) search authorization. During the encryption and decryption processes, $APKS+$ provides a hidden key to conceal the information from the attackers. Thus, $APKS+$ prevents attacks by dictionary keywords, protects the index and privacy of queries.

e) Attribute-Based Encryption (ABE). A novel cryptography solution that implements access control policies [20]. It asserts that one upload includes several policies that $PEKS$ and HVE do not endorse. Attribute-Based Encryption uses an access policy when searching for Boolean expressions for encrypted information. ABE affords better search quality on encrypted data and fast accessing methods.

f) Predicate Privacy Preserving in Public Key Encryption (PPP_PEKS). PPP_PEKS uses a randomization scheme to search for keywords to the best of $PEKS$. The technique of randomization will randomize keywords. Thus, the trapdoors do not provide any meaningful information [42]. A pair of secret keys exchange with the data owner and the recipient, which is not logical for many users. Two mechanisms were specifically present to accept guessing attacks: $PEKSrand - BG$ for brute-force guessing and $PEKSrand - SG$ for statistical guessing, as shown in Fig. 3. Ranked Search The ranked keyword search has a better performance than the Boolean Search by minimizing the major drawback. There is also support for indirect mapping between keywords and trapdoors. So, total computation, communication, and storage overheads are sensible in $PEKS$. Note: -Boolean searches on searchable Encryption have two significant drawbacks.

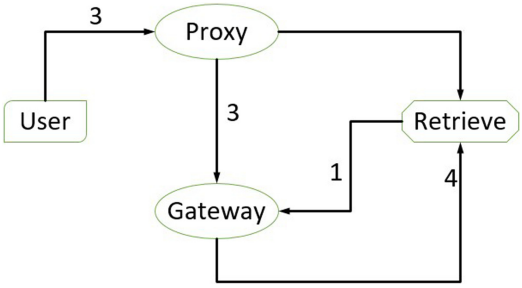

Fig. 3. Working principle of PEKSrand

- Retrieving all files leads to network traffic.
- The user wants to decrypt every file that contains the keyword queried.

a) Keyword Search-Based on Ranking over Encrypted Data. The Ranked Searchable Symmetric Encryption (RSSE) framework backing rank search is a cryptographic primitive build on the SSE. It comprises four algorithms, specifically $KeyGen$, $BuildIndex$, $TrapdoorGen$, $SearchIndex$. Accordingly, a well-organized $RSSE$ framework adopts the Order Preserving Symmetric Encryption (OPSE) scheme and supports deterministic properties [35].

b) *K*-gram based fuzzy keyword Ranked Search (*K*-gam) Searchable encryption methods allow data users to query encrypted data based on keywords securely. These approaches only support precise keyword search and fail to perform spelling mistakes/morphological variants of words [40].

4.2 Secure Spatial Query

Nowadays, there are many searchable encryptions proposed based on a spatial query. Here we classify some papers based on the searchable encryption method

supporting either kNN query or range query, and we will discuss their merits and demerits. By doing this, we can carefully select the technique for practical implementations that keep encrypting information retrieve.

Table 3. Summarazation and comparison of secure spatial query

Spatial query category	Spatial query schemes	Merits	Demerits
Secure kNN query processing methods	VD-kNN [3]	—Process encrypted Voronoi diagrams and return an exact result —It is more secure and accurate	—For k >1, this may place a heavy load on the data owner
	B-kNN [3]	—To minimise overheads, it uses the notion of query squares	—It is inefficient to query the process by itself —It has a high computational cost
	T-kNN [3]	—It can support any k range value —Decrease the data owner load —This method is less expensive compared to the VD-kNN method with higher accuracy	—It may not return true kNN at all times
Secure range query processing methods	hRQ [34]	—hRQ approach is secure and efficient —Protect ordering information	—It may introduce false positives
k-Nearest Neighbor Classification (Relational data)	PPkNN [36]	—It protects data confidentiality, the input query of the user, and hides the pattern of data access —To provide an efficient solution for classification problem	—It does not solve the DMED (Data Mining encrypted data) problem —It is not used to encrypted highly sensitive information
Secure k-Nearest Neighbor Query (Outsourced environments)	SkNN [4]	—It protects the queries the database	—Encrypted data are not secure

Secure kNN Query Processing Methods

a) Voronoi Diagram kNN Method for Secure NN queries (VD-kNN). A system works by handing out encrypted Voronoi diagrams and returning reliable results to protect the closest neighborhood queries [11].

b) Basic *k* Nearest Neighbor (B-*k*NN). This system summarises an individual pair of encrypted data points, and the CSP must govern according to Secure Distance Comparison Methods (SDCM). This scheme sums up an individual pair of encrypted data points, and the CSP must govern according to Secure Distance Comparison Methods (SDCM). CSP should determine the data point That is closer to the encrypted query point [11].

c) Triangulation-Based *k*NN (T-*k*NN). This method deals with any value of k by processing encrypted Delaunay triangulation [11] and reducing the data owner's load. The technique is an approximation for k>1 and achieves high precision in practice.

d) Privacy-Preserving *k*-Nearest Neighbor (PP*k*NN). PPkNN is a safe k-NN classifier based on encrypted, semantically secure data. A scheme where Alice does not engage in any calculations until the encrypted data outsource to the CSP [11]. The protocols used to generate a PPkNN scheme are Secure Multiplication (SM), Secure Squared Euclidean Distance (SSED), Secure Bit Decomposition (SBD), Secure Minimum, and Secure Bit-OR (SBOR). The PPkNN protocols protect the confidentiality of records, input requests, and obscure access to data.

e) Secure Processing of k-Nearest Neighbor Query over Encrypted Data (S*k*NN). The SkNN method is implemented to securely understand k-nearest data tuples to Q using the encrypted T database within CSP. Nothing is known to CSP about the actual contents of the t database and the query record Q [16]. The effective SkNN protocol satisfies the following conditions.

- Secure the patterns of CSP data access.
- Computing the k-Nearest Neighbors query Q correctly.
- Incurring low overhead computation of the end-user.

Yousef *et al.* [32] proposed A secure kNN system that preserves information confidentiality, user input queries, and data access patterns.

Secure Range Query Processing Method

(a) Mutable Order-Preserving Encryption (*m*OPE). This method tolerates safe range query evaluation and is the only known Order-Preserving Encoding System (mOPE) verifiably secure to date [26]. In [3,4], the mOPE is distinct from the previous techniques. In a client-server environment, the mOPE scheme works. The client has a hidden key to the asymmetric cryptography method, such as AES, and the data setoff (cyphertexts) can be stored on the CSP in increasing order of plaintexts.

b) Half-Space Range Query (*h*RQ). This scheme can accomplish polyhedral queries on encrypted data. $\hat{R}$-trees are essential for encrypting and outsourcing the index to the database [36]. The $\hat{R}$-trees index is preferable because there is a lower leakage of information than bucketization schemes such as a k-means, BNL-tree, and kD-tree [38].

4.3 Secure Spatial Keyword Query

Privacy-Preserving Top-K Spatial Keyword Queries (PkSKQ). In terms of spatial proximity and textual significance, This scheme considers k objects ideal for the query. The privacy-preserving top-k query, where an improved version of $ASPE$ encrypts spatial and textual data, is facilitated by a secure index. An enhanced variant of $ASPE$ is $ASPEN$ with Noise [32].

Table 4. Summarazation and comparison of secure spatial keyword query

Spatial keyword query category	Spatial keyword Query Schemes	Merits	Demerits
Privacy-preserving Top-k spatial Keyword queries	PkSKQ [32]	—Enable search over encrypted tree index —Secure pruning techniques based on keywords —Improved spatial-textual data query output on broad scales —Valid and secure scheme	—If CSP is colluding with the data users, the data users can reveal the secure structure —The use of pattern hiding techniques such as Private Information Retrieval (PIR) [39] and Oblivious RAM [18,31] is ineffective
Privacy-preserving boolean spatial keyword queries	PPBSKQ [14]	—Efficient and secure query —Improved Index (dimension expansion and space reduction)	—Transmission channel insecurity —Challenge to get reduced space

Keyword-Based Secure Pruning. The primary goal is to prune the nodes that do not contain any of the queried keywords. A powerful and storage-saving technique for storing information about an object's existence in a dataset is a Bloom filter [2].

Privacy-Preserving Boolean Spatial Keyword Queries (PPBSKQ). This scheme is a standard query for spatial keywords that takes both the spatial range and the keyword into account. All the query keywords must be understood by each resulting candidate and located with a definite query range [14]. The first

Table 5. Comparison of existing geo-textual indices

RF	Algorithm	Threat model		Plain-Text Attack		Security			Privacy					Performance				Other	
		CH	M	CPA	KPA	Au	Cn	In	QP	DP	RP	PPP	APP	Cr	Sc	Mo	Ef	TS	PD
[14]	PPBSKQ	✓		✓	✓	Δ	✓	✗	✓	✓	✓	✓	✗	✗	✓	static	linear	✗	✗
[25]	SPKS	Δ	Δ	✗	✗	Δ	✓	✓	✗	✓	✓	✓	✓	✓	Δ	static	linear	✗	✗
[10]	FKS	✓		✓	✓	Δ	✓	✗	✓	✓	✓	✗	✗	✗	Δ	static	nonlinear	✗	✗
[23]	FKS	✓		✓	✓	Δ	✓	✗	✓	✓	✓	✗	✗	✗	Δ	static	nonlinear	✗	✗
[24]	APKS	✓		✗	✗	Δ	✓	✗	✗	✓	✗	✗	✗	✗	✓	static	nonlinear	✗	✗
	APKS+	✓			✓	Δ	✓	✓	✓	✓	✓	✓	✓	✓	✓	static	Δ	✗	✗
[20]	ABE	Δ		Δ	Δ	Δ	✓	✓	✓	✓	✓	✓	✓	✓	Δ	Δ	Δ	✗	✓
[35]	RSSE	✓		✓	✓	Δ	✓	✓	✓	✓	✓	✓	✓	✓	✓	both	nonlinear	✗	✗
[40]	K-gram	✗	✗	Δ	✓	Δ	✓	✗	✓	✓	✓	✗	✓	✗	✓	static	Δ	✗	✗
[11]	VD-kNN	✓		✓	✓	Δ	✓	✓	✓	✓	✓	✓	✓	✓	✓	static	linear	✗	✗
	B-kNN	✓		✓	✓	Δ	✓	✓	✓	✓	✓	✓	✓	✓	✓	static	linear	✗	✗
	T-kNN	✓		✓	✓	Δ	✓	✓	✓	✓	✓	✓	✓	✓	✓	static	linear	✗	✗
[28]	SkNN	✓		✓	✓	Δ	✓	✓	✓	✓	✓	✓	✗	✓	Δ	static	nonlinear	✗	✗
[16]	PPkNN	✓		✓	✓	Δ	✓	✓	✓	✓	✓	✓	✓	✓	Δ	static	nonlinear	✗	✗
[36]	hRQ	✓		✗	✗	Δ	✓	✓	✓	✓	✗	✓	✓	✓	✗	static	nonlinear	✗	✗
[32]	PkSKQ	✓		✓	✓	Δ	✓	✗	✓	✓	✓	✓	✗	✗	✓	static	linear	✗	✗

Notes. ✓: Denotes that the method satisfies the criterion;
✗: Denotes that the method does not satisfies the criterion;
Δ: Denotes that the method does not care about the criterion or, (the method meets the criterion but not clarified)

difficulty is the creation of consolidated keyword data. In the meantime, under the widely accepted known background thread model [5,34], it can support query processing. A new spatial-textual Bloom filter encoding approach transforms spatial and text information into vectors to find this issue out. Based on ASPE [37], the mapped information in the Bloom filter encoder can be encrypted.

5 Comparison, Analysis, and Summary

Privacy-preserving refers to both query and data privacy. The queried keywords are one way secure and will not be revealed to the CSP since each keyword is encrypted with a secure key. Location privacy is also achieved by using improved encryption methods known to the data owner and end-user only. Keywords are secure, similar to query keywords. The document location is still not hidden from the CSP as the secure index is built on the encrypted documents based on their locations. Therefore, the system ensures the privacy preservation of both the query and the data, but the CSP can know the querys' accurate location.

In this survey, we summarize three basic privacy-preserving queries and evaluate them. The performance evaluation result of all reviewed papers under a group of secure keyword query (see Table 2), secure spatial query (see Table 3), and secure spatial keyword query in Table 4. We summarize the comparison of different methods against the evaluation parameters in Table 5.

Our survey shows that rank-based data retrieval has better information security, fast search access, and does not outflow data to untrusted authorities, and is sufficient for encrypted data searching. The protected kNN query protocol over encrypted data preserves data confidentiality, user requests privacy, and hides patterns of access to information. A privacy-preserving kNN classification is proposed to provide a novel solution to the PPkNN classifier issue over encrypted data.

In addition to privacy promises, a privacy-preserving range query scheme effectively concerns overhead storage, the processing time of queries, and overhead communication. The cloud's overhead storage should be low since most of the information stored in the cloud is typically huge. As many applications need real-time queries, the database processing time desired is negligible. Other than encrypted data objects, the communication overhead denotes the information transmitted between the data owner and the CSP. The data transferred between the data users and the CSP, rather than the query's exact results, is also referenced. Because of bandwidth restrictions and additional time requirements, the overhead is low for uploading and downloading.

Access pattern privacy and path pattern privacy, hiding data access patterns from the CSP, are not considered adequately in most privacy-preserving spatial and textual queries. Secure privacy-preserving keywords and authorized private keyword searches are also do not consider in a detailed manner. Another important parameter is to support scalability since most of the methods give good results for a few data. However, in reality, there are a vast amount of data in could server to process. Thus, these methods are not applicable for real-time applications.

Threat Model. The server follows the mandatory protocol for communication and correctly implements the algorithms needed. It may, however, attempt to obtain information about the data and the content of the queries with the help of domain knowledge. Most of the methods are built based on this curious but honest model, except for SPKS, ABE, and k-gram techniques. The critical issue occurs if it is not trusted by either the data user or the data owner.

Authorization and Access Control. During our survey, not almost all methods consider this privacy parameter, which controls every request to a server and determines the granted or denied request based on specific rules.

Computation and Communication Costs. In our survey, we have found that some of the methods are linear. Which includes SPKS, VD-kNN, B-kNN, T-kNN, PkSKQ, and PPBSKQ, and some others are groups under the category of nonlinear (FKS, APKS, RSSE, hRQ, PPkNN, SkNN. The rest of the methods are unknown whether they include or not.

Other Attributes. It includes attributes other than text and geographic locations, including timestamp (actual time location), place description, and some comments to find places. These attributes are essential and common in recent social media usage (Twitter) has high significance. Except for ABE and hRQ, all others have not these attributes.

6 Open Research Issues and Future Research Directions

Based on the proposed evaluation criteria, we have come across some open issues with spatial keyword search over encrypted data outlined in Section VIA. Furthermore, In Section VIB, we are trying to suggest a list of future directions for study.

6.1 Open Issues

Our serious survey has found some open research issues in privacy-preserving spatial, keyword, and spatial keyword queries, which need future exploration.

First, all the existing research work in our survey on secure and privacy-preserving spatial queries are based on the static objects in which users who initiate a query are at a fixed position. Besides, existing work assumes that the data to be searched is not updated within some time interval. However, in our everyday lives, we move around outside to use our smartphones and other mobile devices to receive information anywhere and check for locations. At this point, we need a secure and privacy-preserving dynamic spatial keyword, which is still an open issue.

Second, one of the vital privacy parameters is authorization or access control. All existing work research on the spatial keyword queries is based on a curious but honest server. They have not considered a complete dishonest threat model.

Third, there still a lack of sound solutions for having additional attributes like timestamp and place description in the existing secure queries. These attributes

seem very simple but have a high significance in a practical situation as used on Twitter. Nevertheless, in our survey, except for ABE and hRQ methods, other methods do not include these attributes.

6.2 Future Direction

We propose some future research directions to implement a usable and secure authentication system with privacy preservation queries.

- Study on a secure and privacy-preserving keyword, spatial and spatial keyword queries systems are in our day-to-day activities. In the face of such unsafe and dynamic cyberspace, some open issues must be resolve as soon as possible. Present-day, the commonly used mobile application for find location has become common as it improves the querying time and system performance by guaranteeing its security and user privacy. How to shield the user's data is a central research topic. Specifically, when the user-sensitive information like medical records stored in a CSP or daily routing recorded in the CSP is not fully trusted.
- Secure and privacy-preserving spatial keyword queries on dynamic spatial keyword objects are not well exploited. A series of factors could affect security and system performance. Designing usable, fast, secure, and privacy-preserving systems is an important topic, particularly when security and privacy become hot issues. Besides, for the system to operate efficiently and accurately for customers, a suitable searching algorithm plays a key role. Advanced algorithms should further examine to support efficiency, security, and privacy. At the same time help to forecast some open issues and imminent research tips. We strongly suggest that improving the security and privacy of existing secure search schemes should be highlighted in future research.
- The cost of authentication is a source-restricted mobile device that should be considered. Most mobile devices (for instance, smart bracelets and mobile phones) have limited electricity, computation capability, and storage space. It makes them applicable in a native search service, an online business directory, a GPS navigation system, and other applications due to an increasing overview of geo-positioning technologies and geolocation services. Thus, it is essential to consider secure and privacy-preserving methods and algorithms implemented without quality degradation in all these limiting parameters.
- The Continuous k Nearest Neighbors (CkNN) and Continuous Range (CR) queries are specific types of Continuous Spatial Queries (CSQs). In both cases, a protected region-based query processing techniques can be managed using secure regions to find more kinds of queries. In the past few years, privacy-preserving spatial keyword queries have attracted substantial interest. Therefore, CSQ has implemented some sympathetic extension, which has implications for privacy-preserving continuous top-k spatial keyword queries and privacy-preserving continuous range spatial keyword queries.
- Most current research effort considering the curious but honest threat model of the cloud server. The server will obey the schemes planned but will attempt

to reveal the data users' and owners' sensitive data. A performance assessment of the curious but honest or dishonest threat model is not addressed well. Future research should improve the quality of model performance evaluation. To prove the effectiveness of privacy-preserving spatial keyword search method, more holistic model evaluation should be conducted.
– When the data owner sends the data, index, and search algorithm to the CSP, data, query, and location privacy should be maintained. The cloud service provider will process the query that the data user sends. In reality, for spatial keyword queries, a lightweight and powerful privacy-preserving scheme is highly predictable. A concrete application demands these severe issues are answered. In this situation, cryptography-based schemes may not be valid in the spatial keyword search. A trivial and well-organized solution is highly expected. Therefore, future research work should give proper attention to the effect of scalability in the system.

Therefore, future research work should give proper attention to the effect of scalability in the system.

7 Conclusion

We have reviewed the recent developments in privacy-preserving spatial, keyword, and spatial keyword queries. This paper pointed out the basic secure privacy-preserving queries and further proposed a sequence of evaluation criteria for assessing current works' performance. We presented a comparative evaluation of the topical outcomes by dividing the existing privacy-preserving searching methods into three categories based on the point of interest they are designed. We found that some of the designs of the current systems are based on honest but curious models. We build some open problems and predict forthcoming research advice based on our survey. We have confidence in taming the security and privacy of existing secure search schemes getting proper attention in the future study.

This paper marks the beginning of a new age of protected privacy research. Many previous surveys, for example, concentrated on secure spatial and secure keyword queries separately, rather than offering a detailed overview of privacy-preserving spatial keyword queries. The study of a stable, privacy-preserving spatial keyword query in both static and dynamic situations is still in its early stages, and there are many important research questions to be answered, such as:

– Our day-to-day activities, the widely used mobile application for finding a location, have become popular and capable of boosting querying time and device efficiency by maintaining user privacy and protection.
– Secure and privacy-preserving spatial keyword queries on dynamic spatial keyword objects are not well exploited.
– The cost of authentication in a source-restricted mobile device should be taken into account.

– The consistency of model performance assessment should be improved in future research.

We expect the importance of questions like these to grow with increasing commercial interest in a secure and privacy-preserving keyword, spatial and spatial keyword queries system.

Acknowledgements. This work was supported by the National Natural Science Foundation of China (No. 62072361), the Fundamental Research Funds for the Central Universities (No. JB211505)

References

1. Beckmann, N., Kriegel, H.P., Schneider, R., Seeger, B.: The r*-tree: an efficient and robust access method for points and rectangles. In: Proceedings of the 1990 ACM SIGMOD International Conference on Management of Data, pp. 322–331, SIGMOD 1990. Association for Computing Machinery, New York, NY, USA (1990). https://doi.org/10.1145/93597.98741
2. Bloom, B.H.: Space/time trade-offs in hash coding with allowable errors. Commun. ACM **13**(7), 422–426 (1970). https://doi.org/10.1145/362686.362692
3. Boldyreva, A., Chenette, N., Lee, Y., O'Neill, A.: Order-preserving symmetric encryption. In: Joux, A. (ed.) EUROCRYPT 2009. LNCS, vol. 5479, pp. 224–241. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-01001-9_13
4. Boldyreva, A., Chenette, N., O'Neill, A.: Order-preserving encryption revisited: improved security analysis and alternative solutions. In: Rogaway, P. (ed.) CRYPTO 2011. LNCS, vol. 6841, pp. 578–595. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-22792-9_33
5. Cao, N., Wang, C., Li, M., Ren, K., Lou, W.: Privacy-preserving multi-keyword ranked search over encrypted cloud data. IEEE Trans. Parallel Distrib. Syst. **25**(1), 222–233 (2014). https://doi.org/10.1109/TPDS.2013.45
6. Chang, Y.-C., Mitzenmacher, M.: Privacy preserving keyword searches on remote encrypted data. In: Ioannidis, J., Keromytis, A., Yung, M. (eds.) ACNS 2005. LNCS, vol. 3531, pp. 442–455. Springer, Heidelberg (2005). https://doi.org/10.1007/11496137_30
7. Chen, L., Shang, S., Yang, C., Li, J.: Spatial keyword search: a survey. GeoInformatica **24**(1), 85–106 (2019). https://doi.org/10.1007/s10707-019-00373-y
8. Chen, L., Cong, G., Jensen, C.S., Wu, D.: Spatial keyword query processing: An experimental evaluation. Proc. VLDB Endow. **6**(3), 217–228 (2013). https://doi.org/10.14778/2535569.2448955
9. Chen, Y.Y., Suel, T., Markowetz, A.: Efficient query processing in geographic web search engines. In: Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data, pp. 277–288, SIGMOD 2006, Association for Computing Machinery, New York, NY, USA (2006). https://doi.org/10.1145/1142473.1142505
10. Cheng, L., Jin, Z., Wen, O., Zhang, H.: A novel privacy preserving keyword searching for cloud storage. In: 2013 Eleventh Annual Conference on Privacy, Security and Trust, pp. 77–81 (2013). https://doi.org/10.1109/PST.2013.6596039
11. Choi, S., Ghinita, G., Lim, H., Bertino, E.: Secure KNN query processing in untrusted cloud environments. IEEE Trans. Knowl. Data Eng. **26**(11), 2818–2831 (2014). https://doi.org/10.1109/TKDE.2014.2302434

12. Christoforaki, M., He, J., Dimopoulos, C., Markowetz, A., Suel, T.: Text vs. space: efficient geo-search query processing. In: Proceedings of the 20th ACM International Conference on Information and Knowledge Management, pp. 423–432, CIKM 2011. Association for Computing Machinery, New York, NY, USA (2011). https://doi.org/10.1145/2063576.2063641
13. Cong, G., Jensen, C.S., Wu, D.: Efficient retrieval of the top-k most relevant spatial web objects. Proc. VLDB Endow. **2**(1), 337–348 (2009). https://doi.org/10.14778/1687627.1687666
14. Cui, N., Li, J., Yang, X., Wang, B., Reynolds, M., Xiang, Y.: When geo-text meets security: privacy-preserving Boolean spatial keyword queries. In: 2019 IEEE 35th International Conference on Data Engineering (ICDE), pp. 1046–1057 (2019). https://doi.org/10.1109/ICDE.2019.00097
15. Eldawy, A., Mokbel, M.F.: The era of big spatial data. Proc. VLDB Endow. **10**(12), 1992–1995 (2017). https://doi.org/10.14778/3137765.3137828
16. Elmehdwi, Y., Samanthula, B.K., Jiang, W.: Secure k-nearest neighbor query over encrypted data in outsourced environments. In: 2014 IEEE 30th International Conference on Data Engineering, pp. 664–675 (2014). https://doi.org/10.1109/ICDE.2014.6816690
17. Ghinita, G.: Privacy for location-based services. Synthesis Lect. Inf. Secur. Privacy Trust **4**(1), 1–85 (2013)
18. Goldreich, O., Ostrovsky, R.: Software protection and simulation on oblivious rams. J. ACM **43**(3), 431–473 (1996). https://doi.org/10.1145/233551.233553
19. Guttman, A.: R-trees: A dynamic index structure for spatial searching. In: Proceedings of the 1984 ACM SIGMOD International Conference on Management of Data, pp. 47–57, SIGMOD 1984. Association for Computing Machinery, New York, NY, USA (1984). https://doi.org/10.1145/602259.602266
20. Hohenberger, S., Waters, B.: Attribute-based encryption with fast decryption. In: Kurosawa, K., Hanaoka, G. (eds.) Public-Key Cryptography - PKC 2013, pp. 162–179. Springer, Heidelberg (2013)
21. Khodaei, A., Shahabi, C., Li, C.: Hybrid indexing and seamless ranking of spatial and textual features of web documents. In: Bringas, P.G., Hameurlain, A., Quirchmayr, G. (eds.) DEXA 2010. LNCS, vol. 6261, pp. 450–466. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15364-8_37
22. Krumm, J.: A survey of computational location privacy. Pers. Ubiquit. Comput. **13**(6), 391–399 (2009)
23. Li, J., Wang, Q., Wang, C., Cao, N., Ren, K., Lou, W.: Fuzzy keyword search over encrypted data in cloud computing. In: 2010 Proceedings IEEE INFOCOM, pp. 1–5 (2010). https://doi.org/10.1109/INFCOM.2010.5462196
24. Li, M., Yu, S., Cao, N., Lou, W.: Authorized private keyword search over encrypted data in cloud computing. In: Proceedings of the 2011 31st International Conference on Distributed Computing Systems, pp. 383–392, ICDCS 2011. IEEE Computer Society, USA (2011). https://doi.org/10.1109/ICDCS.2011.55
25. Liu, Q., Wang, G., Wu, J.: Secure and privacy preserving keyword searching for cloud storage services. J. Netw. Comput. Appl. **35**(3), 927–933 (2012)
26. Popa, R.A., Li, F.H., Zeldovich, N.: An ideal-security protocol for order-preserving encoding. In: 2013 IEEE Symposium on Security and Privacy, pp. 463–477 (2013). https://doi.org/10.1109/SP.2013.38
27. Qi, J., Zhang, R., Jensen, C.S., Ramamohanarao, K., HE, J.: Continuous spatial query processing: a survey of safe region based techniques. ACM Comput. Surv. **51**(3), 1–39 (2018). https://doi.org/10.1145/3193835

28. Samanthula, B.K., Elmehdwi, Y., Jiang, W.: K-nearest neighbor classification over semantically secure encrypted relational data. IEEE Trans. Knowl. Data Eng. **27**(5), 1261–1273 (2015). https://doi.org/10.1109/TKDE.2014.2364027
29. Schreter, I., Gottipati, C., Legler, T.: Inverted indexing, 15 May 2018. US Patent 9,971,770
30. Shen, J., Liu, D., Shen, J., Tan, H., He, D.: Privacy preserving search schemes over encrypted cloud data: a comparative survey. In: 2015 First International Conference on Computational Intelligence Theory, Systems and Applications (CCITSA), pp. 197–202 (2015). https://doi.org/10.1109/CCITSA.2015.46
31. Stefanov, E., et al.: Path ORAM: an extremely simple oblivious RAM protocol. In: Proceedings of the 2013 ACM SIGSAC Conference on Computer and Communications Security, pp. 299–310, CCS 2013. Association for Computing Machinery, New York, NY, USA (2013). https://doi.org/10.1145/2508859.2516660
32. Su, S., Teng, Y., Cheng, X., Xiao, K., Li, G., Chen, J.: Privacy-preserving top-k spatial keyword queries in untrusted cloud environments. IEEE Trans. Serv. Comput. **11**(5), 796–809 (2018). https://doi.org/10.1109/TSC.2015.2481900
33. Vaid, S., Jones, C.B., Joho, H., Sanderson, M.: Spatio-textual indexing for geographical search on the web. In: Bauzer Medeiros, C., Egenhofer, M.J., Bertino, E. (eds.) SSTD 2005. LNCS, vol. 3633, pp. 218–235. Springer, Heidelberg (2005). https://doi.org/10.1007/11535331_13
34. Wang, B., Yu, S., Lou, W., Hou, Y.T.: Privacy-preserving multi-keyword fuzzy search over encrypted data in the cloud. In: IEEE INFOCOM 2014 - IEEE Conference on Computer Communications, pp. 2112–2120 (2014). https://doi.org/10.1109/INFOCOM.2014.6848153
35. Wang, C., Cao, N., Ren, K., Lou, W.: Enabling secure and efficient ranked keyword search over outsourced cloud data. IEEE Trans. Parallel Distrib. Syst. **23**(8), 1467–1479 (2012). https://doi.org/10.1109/TPDS.2011.282
36. Wang, P., Ravishankar, C.V.: Secure and efficient range queries on outsourced databases using rp-trees. In: 2013 IEEE 29th International Conference on Data Engineering (ICDE), pp. 314–325 (2013). https://doi.org/10.1109/ICDE.2013.6544835
37. Wong, W.K., Cheung, D.W.l., Kao, B., Mamoulis, N.: Secure KNN computation on encrypted databases. In: Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data, pp. 139–152, SIGMOD 2009, Association for Computing Machinery, New York, NY, USA (2009). https://doi.org/10.1145/1559845.1559862
38. Wu, M.C., Buchmann, A.P.: Encoded bitmap indexing for data warehouses. In: Proceedings 14th International Conference on Data Engineering, pp. 220–230. IEEE (1998)
39. Yi, X., Paulet, R., Bertino, E.: Private Information Retrieval, 1st edn. Morgan & Claypool Publishers, New York (2013)
40. Zhou, W., Liu, L., Jing, H., Zhang, C., Wang, S.Y.S.: K-gram based fuzzy keyword search over encrypted cloud computing. J. Softw. Eng. Appl. **06**(01), 29–32 (2013)
41. Zhou, Y., Xie, X., Wang, C., Gong, Y., Ma, W.Y.: Hybrid index structures for location-based web search. In: Proceedings of the 14th ACM International Conference on Information and Knowledge Management, pp. 155–162 (2005)
42. Zhu, B., Zhu, B., Ren, K.: PEKSRand: providing predicate privacy in public-key encryption with keyword search. In: 2011 IEEE International Conference on Communications (ICC), pp. 1–6 (2011). https://doi.org/10.1109/icc.2011.5962452